“Reduced representation sequencing in barley (*Hordeum vulgare* L.): domestication genomics and rapid gene identification”

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Artem Pankin

aus Moskau

Köln, Januar 2016

Die vorliegende Arbeit wurde am Max-Planck-Institut für Pflanzenzüchtungsforschung in Köln in der Abteilung für Entwicklungsbiologie der Pflanzen (Direktor: Prof. Dr. George Coupland) angefertigt.

Berichterstatter:	Prof. Dr. George Coupland FRS
	Prof. Dr. Juliette de Meaux
Prüfungsvorsitzender:	Jun.-Prof. Dr. Maria Albani
Tag der mündlichen Prüfung:	19. Januar 2016

Table of Contents

ABSTRACT

Recent advances in barley genomics made it feasible to explore genetic diversity in a large-genome cereal and its wild relative on an unprecedented scale. Reduction of the barley genome complexity by reduced representation sequencing is a cost-efficient approach to survey genetic diversity in multiple genotypes in the context of adaptation and domestication. Moreover, reduced representation sequencing generates genome-wide panels of polymorphisms, thus enabling the application of mapping-by-sequencing approaches in cereals, which has been gaining ground as a rapid method to fine-map candidate genes underlying agronomic traits.

In the first part of this study, I focused on understanding the genetic architecture of the barley domestication syndrome and underlying processes modulating diversity in wild and domesticated forms. To this end, I interrogated ~330,000 SNPs in a set of 433 diverse wild and cultivated barley genotypes using a custom designed reduced representation sequencing assay. The SNPs originated from ~ 12,800 loci, enriched for homologs of flowering time, of meristem and inflorescence development, and of domestication-related genes. The diversity analysis identified an unexpectedly high rate of admixture in both wild and domesticated forms, with the latter case predominantly restricted to the landrace genotypes. I found evidence of a severe domestication bottleneck resulting in loss of genetic diversity and maintenance of extended haplotype blocks in strong linkage disequilibrium.

Selection scans identified multiple targets of selection related to the crucial domestication syndrome traits. For examples, several tests identified a sweep that occurred around the genes involved in the non-brittle rachis phenotype. The signatures of selection were found in the homologs of genes implicated in the regulation of photoperiodic flowering, gibberellin synthesis and seed dormancy. This map of barley genetic variation will inform future evolutionary and genome-wide association studies and support the advancement of barley breeding.

In the second chapter, I employed another reduced representation sequencing approach, the whole-exome sequencing, in combination with the mapping-by-sequencing algorithm. Using this toolbox, I identified the red/far-red light photoreceptor *HvPHYTOCHROME C* (*HvPHYC*), carrying a mutation in a conserved region of the GAF domain, as a candidate underlying the *early maturity 5* (*eam5*) locus in barley. I fine-mapped the gene using the SHOREmap algorithm applied on the whole-exome capture data from bulked early flowering segregants derived from a backcross of the Bowman(*eam5*)

introgression line.

Phytochromes play an important role in light signaling and photoperiodic control of flowering time in plants. Here, I show that the *eam5* interacts with the photoperiod response gene *PHOTOPERIOD-H1* (*Ppd-H1*) to accelerate flowering under non-inductive short days. The results accompanying my study suggest that HvPHYC participates in the transmission of light signals to the circadian clock and thus modulates light-dependent processes such as photoperiodic regulation of flowering. The diversity analysis indicates enrichment of the *HvPHYC-eam5* allele in barley cultivars from Japan despite the strong effect of this mutation on the barley clock. This invites further research into comparing physiological effects and the overall significance of the circadian clock on plant adaptation.

ZUSAMMENFASSUNG

Die jüngsten Fortschritte in der Gerstengenomik erlauben nun die umfassende Analyse der grossen Genome der kultivierten Gerste und ihrer Wildform. Die Reduktion der Genomkomplexität durch sogenanntes „reduced representation sequencing“ (RRS) bietet eine kosteneffiziente Möglichkeit, die Veränderung der genetischen Diversität während der Anpassung und Domestizierung von Gerste nach zuvollziehen. Ausserdem generiert das RRS genomweite Diversitätsmuster, die die Anwendung der „mapping-by-sequencing“-Strategie als schnelle Methode zur Feinkartierung von Kandidatengenen für agronomische Merkmale auch in Getreidearten erlauben.

Im ersten Teil dieser Arbeit beschäftigte ich mich mit der genetischen Architektur des Gersten-Domestizierungssyndroms und mit den Prozessen, die zur Modulation der genetischen Diversität in Wild- und Kulturgersten beigetragen haben. Dazu untersuchte ich mit Hilfe eines selbst entwickelten RRS-Assays ~330,000 Einzelnukleotid-Polymorphismen (SNPs) in einem Diversitätsset von 433 Wild- und Kulturgerstengenotypen. Die so identifizierten SNPs stammten von ~12,800 genetischen Loci, die speziell für Homologe von Blühzeit-, Meristem- und Infloreszenzentwicklungs-, sowie Domestizierungsgenen angereichert waren. Die Diversitätsanalysen zeigten eine unerwartet hohe Rate an genetischer Durchmischung von Wild- und Kulturgersten, vor allem zwischen Wildgerste und kultivierten Landrassen. Die Analyse deutete auf einen starken genetischen Flaschenhals während der Domestikation hin, der zur Erosion genetischer Diversität und Konservierung großer Haplotypenblöcke mit starkem Kopplungsungleichgewicht führte.

Selektionstests deuteten vor allem auf eine gezielte Selektion genetischer Loci hin, die zur Regulation entscheidender Domestikationsmerkmale beitragen. So zeigten verschiedene Tests beispielsweise eine lokale, selektionsbedingte genetische Erosion („selektive sweeps“) in unmittelbarer Nähe von Genen, die die Ährenbrüchigkeit regulieren. Signaturen für Selektion wurden ausserdem in Homologen von Genen identifiziert, die die photoperiodeabhängigen Blüte, die Gibberellinsynthese oder Samendormanz regulieren. Diese genomweiten Diversitätsmuster können zukünftig in Evolutions- und genomweiter Assoziationsstudien eingesetzt werden und zum Fortschritt in der Gerstenzüchtung beitragen.

Im zweiten Kapitel der Arbeit, verfolgte ich eine Kombination aus RRS mittels vollständiger Exomsequenzierung und der mapping-by-sequencing Methode. Mit Hilfe dieser Strategie konnte ich den Rotlicht-Photoerzeptor *HvPHYTOCHROME C* (*HvPHYC*) mit einer Mutation in einer konservierten Region der GAF-Proteindomäne als Kandidatengen des *early*

maturity 5 (*eam5*) Lokus von Gerste identifizieren. Die Feinkartierung des Gens erfolge mit dem SHOREmap Algorithmus, der auf die Exomsequenzdaten gepoolter frühblühender Segreganten der rückgekreuzten Bowman(*eam5*)-Introgressionslinie angewandt wurde.

Phytochrome sind ein wichtiger Bestandteil der lichtabhängigen Signalwege in Pflanzen und regulieren den Blühzeitpunkt in Abhängigkeit von der Photoperiode. Meine Daten deuten darauf hin, dass *eam5* mit dem *PHOTOPERIODE-H1* (*Ppd-H1*) interagiert und zur Beschleunigung der Blüte unter nicht-induktiven Kurztagsbedingungen beiträgt. Des Weiteren implizieren meine Ergebnisse eine Beteiligung von HvPHYC an der Transmission von Lichtsignalen zur zirkadianen Uhr und somit an der Modulation lichtabhängiger Prozesse, wie zum Beispiel die Photoperiode abhängige Regulation des Blühzeitpunktes. Trotz des starken Effekts des *HvPHYC-eam5* Allels auf die zirkadiane Uhr, deutet die genetische Diversitätsanalyse auf eine Anreicherung dieser Mutation in japanischen Kulturgersten hin. Diese Ergebnisse legen eine weitere Erforschung der physiologischen Funktion der zirkadianen Uhr und ihrer Bedeutung für die Anpassung von Gerste nahe.

PREFACE

Domesticated barley (*H. vulgare* ssp. *vulgare*) is a diploid plant (2n = 14) cultivated worldwide on the area of ~500,000 km^2 (~1.4x size of Germany) for a variety of the end-use products, i.e. malt, food and fodder (USDA, 2015). It originated from its wild progenitor (*H. vulgare* ssp. *spontaneum*) in the Fertile Crescent approximately 10,000 B.C.; however, the circumstances of barley domestication have remained largely obscure and heatedly debated (Fuller et al., 2012; Heun et al., 2012; Lev-Tadun et al., 2000).

To shed the light on the history of domestication, patterns of genetic variation have been surveyed in wild and domesticated barley using isolated gene assays and the Barley1K chip, which provided the genome-wide assessment of the ascertained set of ~1000 SNPs (Dai et al., 2012; Kilian et al., 2006; Morrell and Clegg, 2007; Russel et al., 2011). In 2011, the International Barley Genome Sequencing Consortium (IBGSC) presented the physical map of barley cultivar Morex covering 98% of its 5.1-Gb genome, which compares to ~20x of the Arabidopsis genome size (IBGSC, 2011). The Morex assembly contained ~26,000 high confidence gene models exhibiting homology with genes from other plant species and ~86% of the genome consisted of the repetitive elements (IBGSC, 2011). Using the segregating population sequencing approach (POPSEQ), Mascher et al. (2013) anchored 80% of the high confidence barley genes on the genetic map, thereby significantly improving resolution of the barley map compared to the IBGSC resource. Recent report on sequencing of 15,622 bacterial artificial chromosomes (BACs) carrying Morex DNA represents the ongoing efforts to fill in the gaps in the IBGSC assembly (Muñoz-Amatriaín et al., 2015). These developments and availability of the barley genomic resources have paved the way to more efficient exploration of genetic diversity and mapping genes of interest in barley.

With the advent of the cost-efficient sequencing technologies, whole-genome re-sequencing of plant populations in the domestication and natural adaptation contexts has gained much popularity (Huang et al., 2012; Lin et al., 2014; Schmutz et al., 2014; Zhou et al., 2014). However, all of these examples represent plant species with relatively small genomes (<1.1 Gb), in which whole-genome sequencing of hundreds of genotypes has become feasible. Yet, in the large-genome plant species, the analysis of genome variation at the population level remains challenging. Only the maize HapMap2 project (genome size ~2.3 Gb) have accumulated considerable amount of the population re-sequencing data (Chia et al., 2012; Hufford et al., 2012). In barley, the shotgun genome sequencing datasets have been generated only for 5 cultivated and wild barley lines (IBGSC, 2011), and, recently, for 10

more Tibetan wild and domesticated genotypes (Zeng et al., 2015).

One of the strategies to circumvent these technical limitations is the reduced representation sequencing (Hirsch et al., 2014). In this approach, only a selection of genomic regions is sequenced and analyzed. This selection may include a whole set of the coding regions of an organism, which is then termed whole-exome sequencing, or genes related to the specific pathways or chromosome regions. In barley, the whole-exome sequencing assay has been developed to capture 61.6 Mbp of coding regions and validated on 13 barley cultivars and 7 genotypes of wild *Hordeum* species (Mascher et al., 2013).

In Chapter 1, I aimed to reveal genome-wide effects of the domestication on the structure of genetic diversity and to catalog candidate loci underlying crucial domestication traits (*sensu* Abbo et al., 2014). To this end, I further reduced the complexity of the barley genome by designing an assay to capture a 2.4-Mbp set of genes. The selection of genes was enriched for the homologs of flowering time, inflorescence and meristem development genes and putative domestication genes. I applied this assay on 345 wild and 87 domesticated barley lines.

In Chapter 2, I aimed to fine-map a candidate gene and identify a candidate mutation underlying the *early maturity 5* locus in barley using the whole-exome sequencing and mapping-by-sequencing algorithm (Druka et al., 2010; Mascher et al., 2013; Schneeberger et al., 2009). Next, I estimated a role of *eam5* in natural adaptation and breeding by analyzing the diversity of the candidate gene in wild genotypes and barley cultivars, originating from various breeding programs.

CHAPTER 1: Population genomics provides insights into the history of barley (*Hordeum vulgare* L.) domestication

1 Introduction

Hordeum L. species, which include domesticated *H. vulgare* ssp. *vulgare* and its wild progenitor subspecies *H. vulgare* ssp. *spontaneum*, have a long history of interaction with humans. As early as 40,000 – 50,000 years ago, late Neanderthals consumed wild *Hordeum* grains in a cooked form as suggested by the analysis of the dental calculus (Henry et al., 2014). Domesticated barley is among a group of the Neolithic founder crops, which facilitated the establishment of the early agricultural societies (Lev-Tadun et al., 2000). The first traces of barley cultivation were found in the Fertile Crescent archaeological sites dated back to ~ 10,000 B.C. The Fertile Crescent, as one of the Vavilov's centers of agricultural origin (Vavilov, 1926), comprises the most of the wild barley diversity. The modern range of its documented occurrence spans the whole region between Western Anatolia and the Tibetan Plateau (Harlan and Zohary, 1966) and largely overlaps with the areas of barley cultivation. Following domestication, cultivated barley spread from its place of origin and formed several distinct groups adapted to the new environmental conditions, agricultural practices and target end uses. However, compared divergence of maize and its wild ancestor teosinte, wild and cultivated barley seem to morphologically diverge to a lesser extent (Gottlieb, 1984).

In an attempt to unify the domestication glossary, Abbo et al. (2014) proposed a strict definition of a domestication trait and coined the term crucial domestication syndrome (DS) trait. The domestication syndrome is a complex of all characters distinguishing wild and cultivated subgroups (Hammer, 1984). According to Abbo et al. (2014), the crucial DS traits are those, “without which the adoption of a species for domestication would be impossible”. In other words, domesticated genotypes should exclusively carry derived (or domesticated) phenotype (or character) of a trait. Therefore, only the genes that underlie the crucial DS traits should be considered the genuine domestication genes. In barley, the brittleness of rachis character (hereafter, brittleness) is the only well-characterized crucial DS trait, since it exhibits clear dimorphism between the wild and domesticated subgroups (Purugganan and Fuller, 2011; Abbo et al., 2014). In domesticated barley, the rachis is nonbrittle and the seeds remain attached after maturation, which ensures efficient harvesting. On the contrary, in wild barley, the brittleness facilitates the efficient spread of the seeds immediately upon maturation. The other barley morphological and physiological traits frequently mentioned in

the domestication context are spike row type, hulled and naked caryopsis, responsiveness to photoperiod and vernalization (reviewed in Abbo et al., 2014; Mayer and Purugganan, 2011; Pourkheirandish and Komatsuda, 2007). However, both wild and derived characters of these traits segregate in domestication barley and thus represent postdomestication divergence of cultivated populations (Cockram et al., 2007; Hemming et al., 2009; Taketa et al., 2004; Tolbert et al., 1979; Turner et al., 2005). The divergence traits have been widely used in breeding programs, adapting barley cultivars to the new cultivation areas and end uses. Consequently, these traits have been extensively studied and thus their genetic regulation has been relatively well understood (Campoli et al., 2013; Dubcovsky et al., 2005; Faure et al., 2012; Hemming et al., 2009; Karsai et al., 2005; Komatsuda et al., 2007; Koppolu et al., 2013; Ramsay et al., 2011; Taketa et al., 2008; Turner et al., 2001; Yan et al., 2006). By contrast, the understanding of barley crucial DS traits is extremely limited. Besides the brittleness, other traits have been suggested to constitute the DS, e.g. seed dormancy, synchrony of flowering, number of side shoots also referred to as tillers, tiller angle, spike width and endosperm groove depth (Badr et al., 2000; Pourkheirandish, T Komatsuda, 2007; Salamini et al., 2002; Zohary, 2004). However, dimorphism of these traits between the wild and domesticated barley still lacks reliable experimental evidence.

When phenotypes are not clearly defined, the so-called 'bottom-up' approach has proven instrumental in reconstructing the genomic architecture of the DS in several crops (Shi and Lai, 2015). Following this approach, multiple loci are scanned for the signatures of selection using the population genetics toolbox and the selected features are further linked to the traits using the molecular genetics techniques (Ross-Ibarra et al., 2007). In common bean, maize, rice, soybean and tomato the selection scans using the genome-wide re-sequencing data yielded hundreds of domestication features and only a handful of them have been linked to the known crucial DS traits (Huang et al., 2012; Hufford et al., 2012; Lin et al., 2014; Schmutz et al., 2014; Zhou et al., 2014). The rest of the selected features apparently associate with yet unstudied phenotypes. These scans revealed domestication genes that were orthologs of the genes implicated in developmental pathways such as the photoperiod, hormone synthesis and nitrogen metabolism (Hufford et al., 2012; Schmutz et al., 2013). However, the function of many of the domestication genes could not be inferred.

In barley, Russel et al. (2011) performed the selection scan using ~ 1000 SNP markers, which were surveyed in a selection of 448 geographically-matched landrace and wild barley accessions. Several genomic regions affected by selection under domestication have been identified, however, low resolution of the assay precluded isolation of candidate

domestication genes. In several studies, domestication selection tests have been conducted for up to seven isolated genes; however, the tests did not identify reliable selection signatures (Kilian et al., 2006; Morrell et al., 2013).

In addition to detecting domestication targets, a genome-wide diversity scan based on the re-sequencing data from multiple genotypes is a powerful approach to detailed understanding of the domestication history (Ross-Ibarra et al., 2007). The diversity analysis helps identify the number and location of domestication events as well as effects of domestication on the structure of modern wild and domesticated populations. Disentangling the domestication history of *indica* and *japonica* cultivated rice subspecies is one of the most interesting and successful examples (Huang et al., 2012). In rice, phylogenetic and population genetics analyses suggested independent origins of *indica* and *japonica* (He et al., 2011; Ma and Bennetzen, 2004; Xu et al., 2012). By contrast, the same alleles of the well-characterized domestication genes have been found in both subspecies, so that the domestication event appeared monophyletic (Molina et al., 2011; Zhang et al., 2009). Re-sequencing of more than 1500 wild and cultivated genotypes coupled with the demographic modeling suggested a single domestication event in Southern China, which gave rise to *japonica* subspecies, and an introgression event, which transferred the domesticated genes into the South Asian wild rice population, a progenitor of *indica* (Huang et al., 2012). Hence, inferring demographic parameters using the genome-wide data obtained from multiple accessions is critical in understanding the history of domestication.

A single domestication event has long been thought to give rise to cultivated barley (Badr et al., 2000; Salamini et al., 2002). Molecular evidence in support of this hypothesis came from the phylogenetic analysis of the amplified fragment length polymorphism (AFLP) markers (Badr et al., 2000). This finding went in line with the long-standing hypothesis of the monophyletic rapid domestication of crops in the Fertile Crescent area, the so-called core model of domestication. This model stems from the concept of the centers of origin developed by Vavilov (1926) and even echoes back to the original Darwin's hypothesis of the monocentric origin of crops (Darwin, 1868). The finding that the brittleness is controlled by two distinct genetic loci was one of the first facts that did not fit into the core model (Kandemir et al., 2004). Two tightly linked tandem genes *Btr1* and *2*, underlying the brittleness, have been cloned (Haberer and Mayer, 2015; Pourkheirandish et al., 2015). The distribution of their variation suggested a polyphyletic origin of barley from two distinct events in the South and in the North Levant. Morrell and Clegg (2006) suggested another center of barley domestication east of the Zagros Mountains in modern Iran based on the

analysis of 684 SNPs from 18 loci and 25 barley accessions. Tibet has been proposed as another center of barley domestication (Dai et al., 2012). Recently, based on the chip genotyping data of 7,864 SNPs, Poets et al. (2015) suggested even a more complex scenario of domestication. They demonstrated that the genomes of the modern landrace barley genotypes are a mosaic of five distinct populations of wild barley of different geographical origin. These findings together with the archaeological evidence, which for example suggested the slow fixation rate of the brittleness phenotype, led to the development of the so-called protracted model of crop domestication in the Fertile Crescent, as an alternative to the core model (Allaby et al., 2008a; Allaby, 2015; Fuller, 2007; Fuller et al., 2012). The protracted model postulates that the domestication in the Middle East has been a complex and polycentric process, which continued several thousand years. The conflict between these models has been a subject of numerous critical reviews, simulation and demographic modeling experiments, which fit the data into one or another hypothesis (Abbo et al., 2014; Allaby et al., 2008b; Allaby and Brown, 2003, 2004; Brown et al., 2009; Gopher et al., 2013; Heun et al., 2012; Ross-Ibarra and Gaut, 2008; Salamini et al., 2004; Zohary et al., 1999). However, the genome-wide re-sequencing analysis complementing the domestication models is still lacking.

This study is the first example of the enrichment sequencing applied to characterize genome-wide diversity of barley in the domestication context. Here, I generated 560 Gb of deep sequencing data (median depth > 45x) from 345 wild barley and 87 barley landraces and cultivars representing different geographical regions and breeding programs. The analyses performed in this study are based on ~330,000 SNPs. The SNPs originated from ~ 12,800 loci, enriched for homologs of flowering time, meristem and inflorescence development, and domestication-related genes. Using instruments of population genetics, I compared various aspects of genetic diversity between wild and cultivated barley. The results demonstrate that admixture is an important factor to account for in the future barley domestication studies in order to avoid incorrect interpretations of evolutionary history of genotypes or specific loci. I identified ~50% loss of diversity and extended LD in domesticated barley compared to the wild genotypes; the patterns apparently associated with the recent domestication bottleneck. The genome scans revealed multiple regions and individual genes affected by selection under domestication. Analysis of the candidate domestication genes suggested that selection under domestication targeted homologous pathways in different plant species. This invites further research into convergence of the traits and molecular pathways affected by independent domestication events in different crops.

2 Results

2.1 Design and performance of the barley gene capture assay enriched for flowering genes

Searching in barley genomic and transcript databases yielded 666 genic sequences, 526 of which were related to flowering time and flower development, 118 genes to domestication-related processes, e.g. tillering, carbohydrate metabolism, seed dormancy; and 22 abiotic stress genes (Fig. 1A, Table S1). Only for 12% of the target sequences, including intronless genes, the enrichment baits were selected based on gDNA, whereas for the other 88% of the genes the baits were designed based on the cDNA (Fig. 1B). Approximately, 85 % of selected sequences comprised putative promoter regulatory regions longer than 100 bp in size. The predicted ORFs of 126 selected genes were longer than the ORFs of the MLOC genes currently used as a barley reference gene set. Whereas the complete ORFs of 52 % of the genes could be mapped to the IBGSC Morex contigs, the rest of the ORFs were only partially present on or completely absent from the IBGSC reference genome. The latter group apparently represents the genic regions not yet incorporated in the Morex reference genome or the unique allelic variants. To attenuate the effects of biased selection of genes related to the specific pathways on estimates of the genetic diversity, I included into the dataset fragments of 1000 randomly selected genes evenly spread over the seven barley chromosomes. In total, the enrichment design covered 94% of the selected sequences with the target enrichment size of 2.42 Mbp (Table 1).

To map the reads, I created a composite reference genome composed of 666 sequences corresponding to the target capture genes, 125 and 944 genomic contigs corresponding to the additional promoter sequences and the randomly selected sequences, respectively, and 21,673 additional Morex genomic contigs, which carried at least a single mapped read, totaling 23,142 reference sequences (Table S2). Such approach enabled discovery of more polymorphisms compared with mapping only to the target capture reference and, on the other hand, greatly reduces computational load compared with mapping to the complete IBGSC barley genome reference.

Illumina sequencing of 433 barley genotypes yielded a total of 8 billion of 100-bp reads (0.56 Tb of data). The mapping revealed that de facto captured regions comprised

approximately 13.8 Mbp of the reference genome, cumulatively covering 93% (2.24 Mbp) of the target capture regions (Table 1). Of the overall captured regions, 1.33 Mbp resided in the CDS regions. Per sample analysis of the coverage revealed that approximately 87% of the target capture regions were covered at the SNP calling threshold of 8 reads and between-sample variation was relatively low, indicating the robustness of the assay (Fig. 2A). The median depth of coverage varied between samples from 45 to 130.

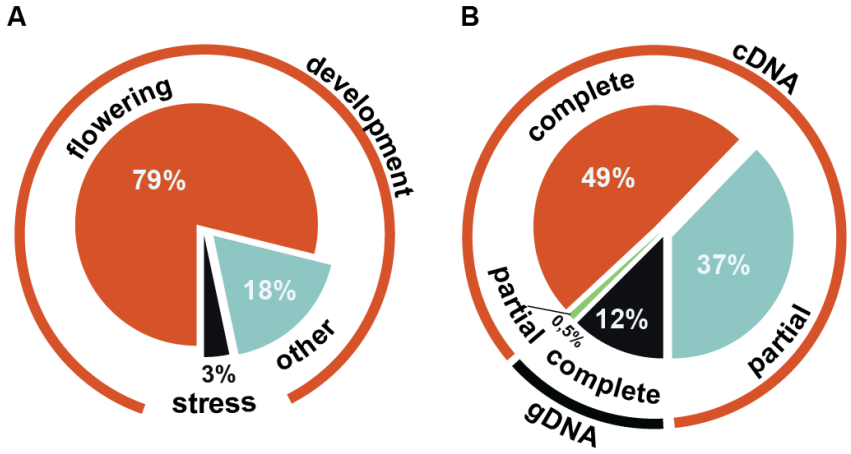


Figure 1. Selection of target genes for enrichment.

Putative functional (A) and structural (B) classes of genes selected for enrichment are shown as colored slices of the pie charts in accordance with the legend.

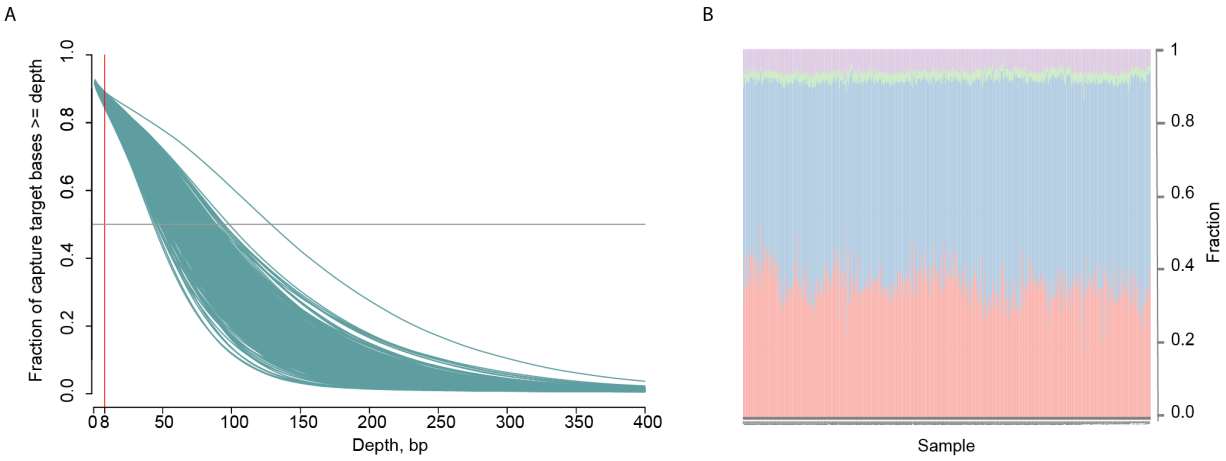


Figure 2. Characteristics of coverage and polymorphisms.

(A) A fraction of target nucleotides covered at a certain depth in the individual samples shown as cyan curves. A cut-off coverage threshold for the SNP calling and the median coverage are shown as vertical red and horizontal gray lines, respectively.

(B) Proportion of different call categories in the individual samples. Fractions of missing data, reference, hetero- and homozygous alleles are shown as red, blue, green and magenta bar charts, respectively.

In total, sequencing of the panel of 433 wild and cultivated barley genotypes

discovered 544,318 high-quality SNPs, including 189,708 singletons. On average, each sample carried 6% of the homozygous SNPs, 3% of the heterozygous SNPs, 55% of the reference alleles and 36% of missing data (Fig. 2B). To investigate whether the large number of heterozygous SNPs was due to pooling of the samples during enrichment, I mapped the exome libraries of three barley genotypes (acc. Nos. ERR271694, ERR271717 and ERR271720; Mascher et al., 2013b), which were individually enriched and sequenced, and extracted the SNPs using the pipeline describe in this study. The amount of heterozygous genotypes derived from the individually-enriched samples was comparable to the number of heterozygotes observed in the pooled-enriched dataset described in this study (Fig. S1).

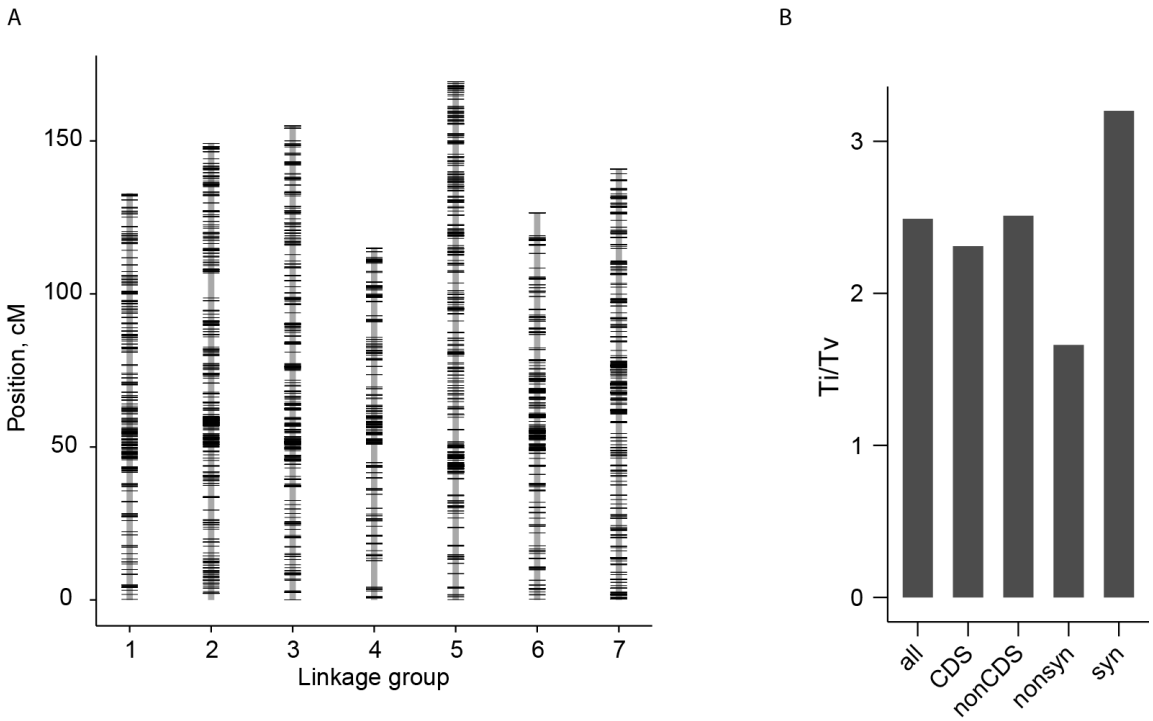


Figure 3 Distribution of SNP markers over the barley chromosomes and transition / transversion (Ti / Tv) ratio.

(A) Mapping location of the SNP markers on barley linkage group based on the PopSeq map (Mascher et al., 2013a). The linkage groups and marker positions are shown as vertical gray and horizontal black bars, respectively.

(B) Ti/Tv ratio of the SNP markers in different structural subgroups – coding (CDS), non-coding (nonCDS), coding synonymous (syn) and coding non-synonymous (nonsyn).

Approximately 22% of the total homozygous SNPs originated from the target enrichment regions, whereas their proportion increased to 38% in the SNP subset that was

filtered for the MAF and missing data (Table 1). 37,870 of all the SNPs resided in the CDS regions and approximately 43% of them fell into the non-neutral category based on the snpEff predictions. The coding regions were more conserved than the non-coding regions as evidenced by the SNP density, which was on average 29 and 41 SNPs per captured Kbp in the coding and non-coding regions, respectively.

Of all the reference sequences that carried SNPs, 24% were located on the barley genetic map, whereas, for 42% of the reference sequences, only the chromosome assignment was known (Suppl. Table R). The SNPs densely covered all seven barley chromosomes with the average distance between the SNPs of 0.59 cM (median 0.28 cM) (Fig. 3A).

To estimate the genome-wide dynamics of point substitutions in the barley genome, I calculated the Ti/Tv ratio for different genomic regions that carried SNPs. The Ti/Tv ratios varied in the range from 1.66 to 3.2 with the genome-wide value of 2.48, which is five times higher than expected if all substitutions would happen at the equal rate (Fig. 3B). The synonymous and non-synonymous SNPs had the highest and lowest Ti/Tv values, respectively, whereas the ratio did not vary dramatically between the coding and non-coding regions with the only slightly higher bias towards transitions in the latter group. The number of different transitions was balanced (191,635 A:G and 196,835 C:T), whereas the number of possible transversions varied in the range from 30,601 to 46,346.

The MAF spectra did not reveal any systematic bias, e.g. lack of rare variants often attributed to the ascertainment bias (Clark et al., 2005; Rosenblum & Novembre, 2007), and resembled the MAF distributions simulated based on the standard neutral coalescent model, i.e. large proportion of rare polymorphisms and rapid exponential decrease in the number of SNPs with the higher MAFs (cf. Nielsen et al., 2004). Interestingly, the rare SNPs (MAF < 0.01) were significantly enriched in the CDSs compared with the non-coding regions (Fig. S2, Fig. S3).

2.2 Patterns of admixture in wild and domesticated barley

Distinguishing wild and domesticated genotypes is critical in a study of domestication, where patterns of genetic variation are contrasted between these two sub-groups. Descriptions of genetic stocks obtained from the seed banks contain taxonomic information, which was recorded by the original collectors and during *ex-situ* reproduction. To verify this information,

I applied PCA on the filtered SNP dataset and estimated global ancestry of the genotypes for a finer-scale refinement. The PCA revealed two distinct clusters corresponding to the domesticated and wild subspecies with the multiple genotypes residing in between these clusters (Fig. S4). The split between the wild and domesticated genotypes was further confirmed by the global ancestry estimates. Surprisingly, the STRUCTURE analysis revealed patterns of recent admixture in 36% and 12% of the domesticated and wild genotypes, respectively (Fig. 4A). These samples were removed from the further population genetic analyses to include only genuine wild and domesticated subgroups.

After pruning, the germplasm subsets consisted of 302 and 58 samples of wild and domesticated genotypes, respectively. The PCA of the filtered genotypes related the genotypes that resided in between the subspecies clusters exclusively to the admixed germplasm (Fig. 4B). Interestingly, in domesticates, the landraces constituted 95% of the admixed individuals and they did not correspond to any specific locality (Table S3). Similarly, in wild subspecies, the admixed genotypes were spread all over the Fertile Crescent (Fig. 5B), indicating that the admixture was not restricted to any particular geographical area.

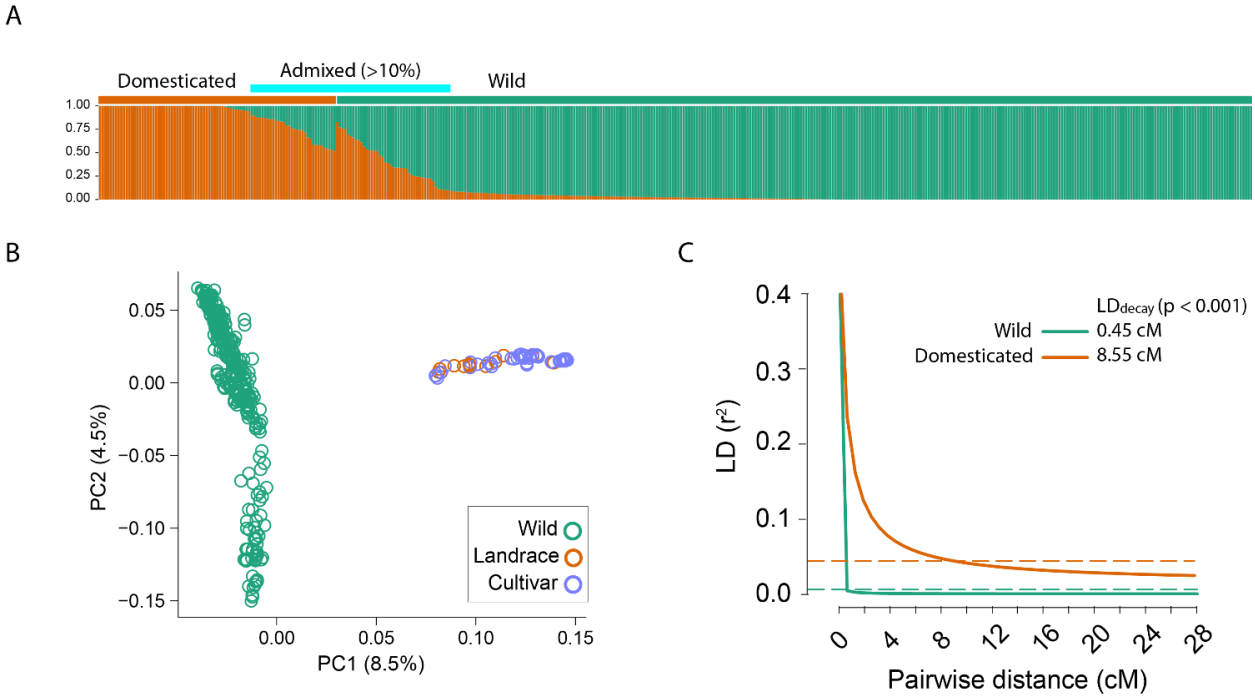


Figure 4. Structure, principal component analysis (PCA) and linkage disequilibrium (LD) in wild and domesticated barley.

(A) Global genetic ancestry of the wild and domesticated barley genotypes as determined by the population structure analysis. Wild and domesticated genotypes are shown as green and orange vertical bars, respectively. Admixed genotypes, which were defined as carrying more that 10 % of mixed ancestry, are shown under the cyan bar.

(B) Decomposition of genetic variation of non-admixed barley genotypes by PCA. Only two major PCs are

shown. The percentages of variation explained by the PCs are shown in parentheses. Wild, landrace and cultivated genotypes are illustrated by green, orange and blue pictograms.

(C) LD decay as a function of the genetic distance between the SNPs. The non-linear regression curves for pairwise r2 values are shown for non-admixed wild (green) and domesticated (orange) barley genotypes. The background levels of LD are shown as horizontal dashed lines. The distances at which the LD decays to the background values are shown in the legend.

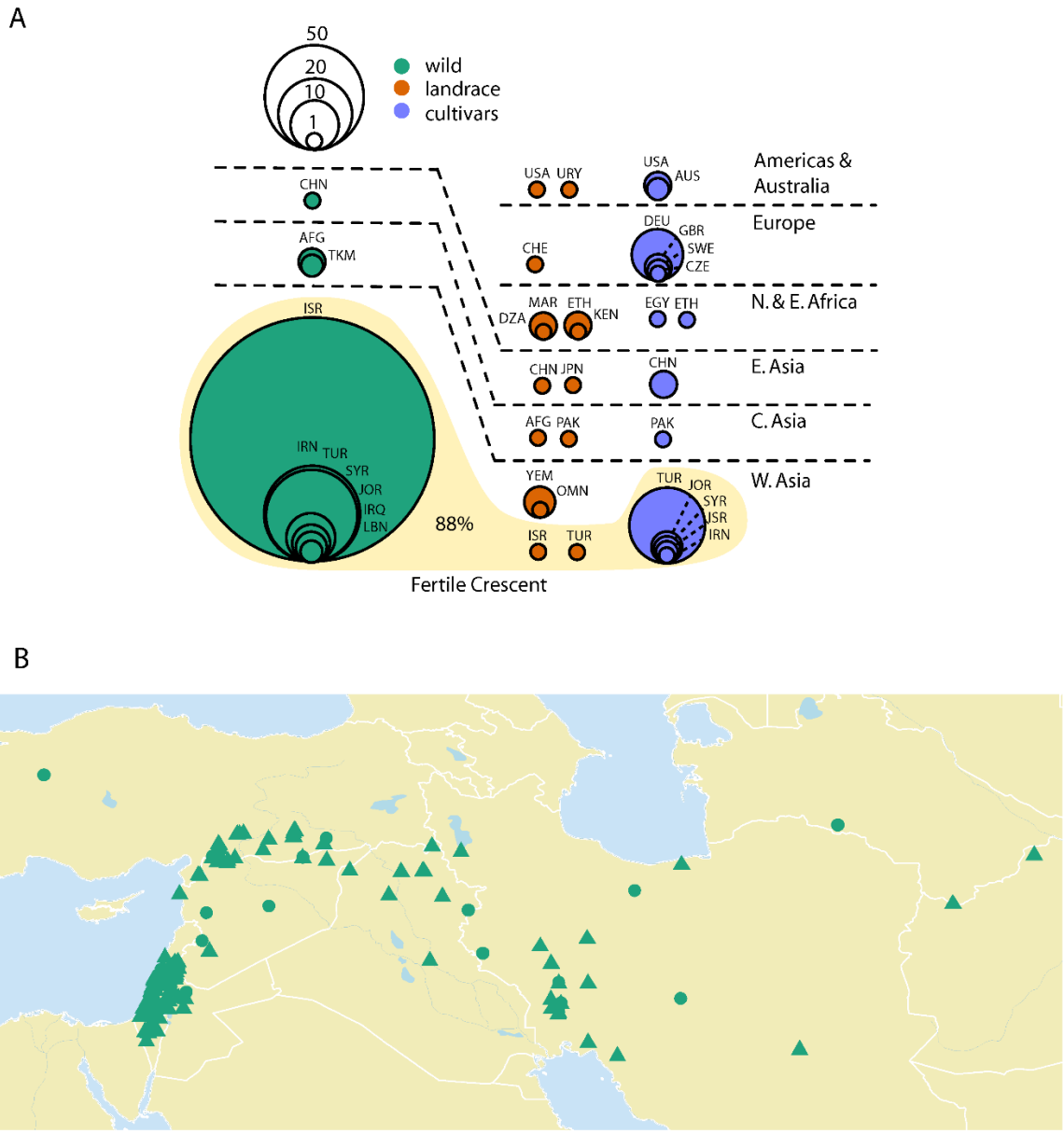


Figure 5. Wild and domesticated barley germplasm.

(A) Geographic representation of wild, landrace and cultivated barley genotypes used in this study. The bubble size is proportional to the number of samples. Colors of the pictograms are as described in the legend.

(B) Geographical distribution of wild barley collecting sites within the Fertile Crescent. Non-admixed and admixed (> 10% of domesticated barley alleles) genotypes are shown as the triangle and round pictograms, respectively.

Landraces and cultivars are two recognized groups of domesticated barley. The former are tentatively defined as locally adapted varieties traditionally cultivated and selected by farmers in the field, whereas the latter are the products of the breeding programs (Zeven, 1998). Despite these generally accepted differences in definitions, sorting extant domesticated genotypes into these two groups is not without controversy, partly owing to the use of landrace material in modern breeding. In this study, the landraces did not differentiate from the cultivars based on the result of both PCA and STRUCTURE analysis (Fig. 4B; Fig. S5). Since the landraces and cultivars were not genetically distinct, these subgroups were treated as a single group of domestication barley in the further analyses.

2.3 Linkage disequilibrium in wild and domesticated barley

Extent of linkage disequilibrium (LD), which is the non-random association of alleles, characterizes the recombination landscape and haplotype diversity of a species or a population. The LD is mostly maintained by the physical properties of a chromosome, as a function of physical distance between markers. Nevertheless, the other processes, such as adaptive selection and varying demographic histories, may create peculiar patterns of LD. With this in mind, I set out to investigate whether the process of domestication modified the patterns of LD in domesticated barley compared with that of the wild genotypes.

The LD decayed to the background levels at the distances of 0.45 cM and 8.55 cM in wild and domesticated sub-groups, respectively (Fig. 4C). Such ~ 20-fold difference in the extent of LD between the groups apparently resulted from the limited amount of historical recombination and therefore retention of longer haplotype blocks in the domesticated barley. Since the wild subset was ~ 5-fold larger than the domesticates, I further examined whether the remarkable difference in the rate of LD decay can be attributed to the unbalanced number of individuals in the two groups. The bootstrapping of the wild subset resulted only in slightly higher estimates of LD decay (median ~ 0.6 cM, p-value < 0.001), demonstrating the robustness of the LD comparisons in the subgroups unbalanced in the number of samples (Fig. S6). The rate of LD decay varied between the individual chromosomes in a range from ~ 0.2 to 0.8 cM in the wild barley and in a much bigger range from ~ 2 cM to 26 cM in the domesticated subspecies. Interestingly, the order of chromosomes within these ranges differed between the two subspecies (Fig. S7). This suggested that the loci targeted by the non-neutral processes, such as adaptation and breeding, which shaped the global patterns of LD, were different in wild and domestication barley. In the case of domesticated barley, the chromosomes bearing the longest LD blocks might carry domestication genes and other loci

targeted by a further improvement.

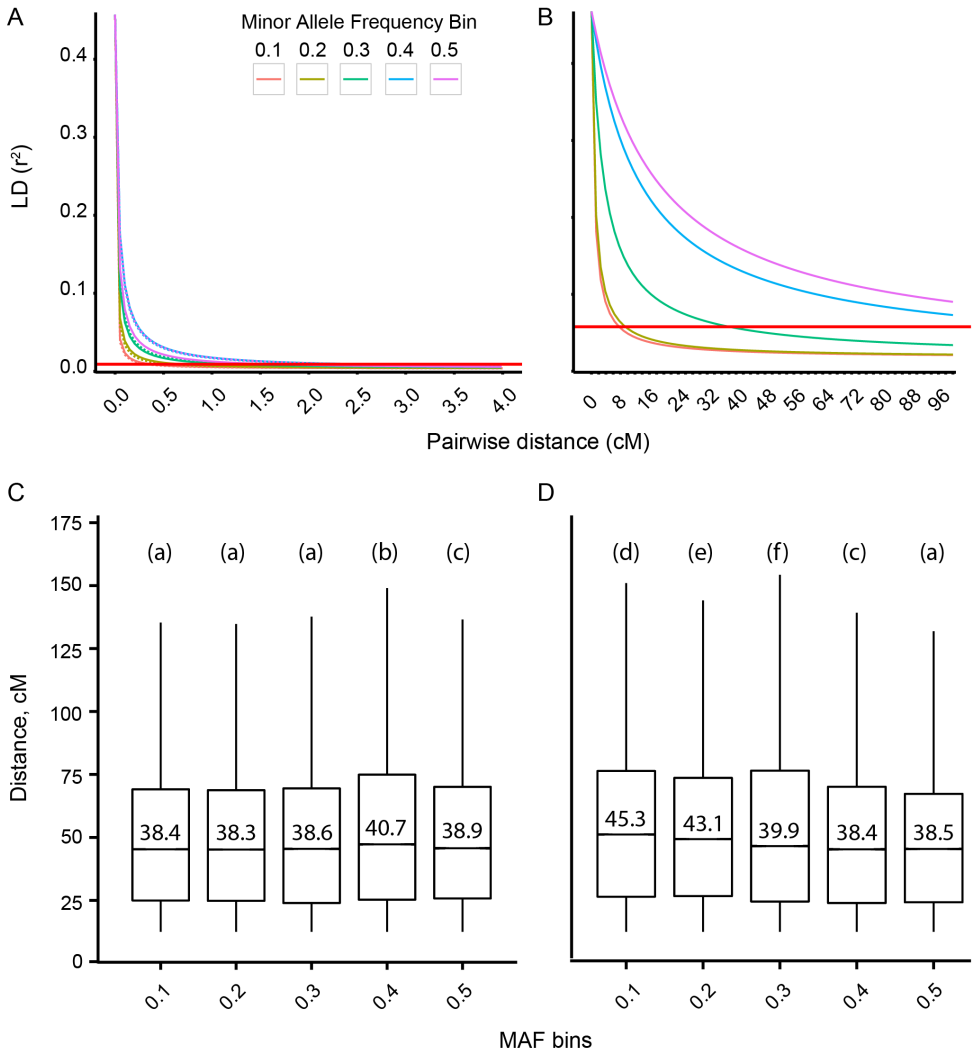


Figure 6. LD dependency on the allele frequency.
(A, B) LD decay curves obtained using the SNP datasets matched by the minor allele frequency (MAF) in bins in wild (A) and domesticated (B) barley (missing data < 50%). The color of the curves corresponds to the MAF bins as shown in the legend. Red horizontal lines indicate the background LD levels. The dashed curves illustrate the LD decay in the wild barley dataset (A) without missing data.

(C, D) Distance covered by perfectly correlated SNPs ($r^2 = 1$) matched by the allele frequency in bins in wild (C) and domesticated (D) barley. The letters in parentheses indicate the significantly different groups as determined by pairwise t-test (FDR adjusted p-value < 0.001).

Averaging LD via regression curves obscures important aspects of the LD structure since LD is often discontinuous and intermittent rather than smoothly declining with the distance. Exploration of the refined LD patterns may offer insights into the causes of LD and help understand its application to diversity analysis. It has been suggested that the measures of

LD show some dependency on the allele frequency (Hedrick, 1987) and that this phenomenon may be related to the age of alleles and selection (Chakravarti, 1999). With that in mind, I reconstructed the LD decay curve for the marker pairs with matching MAF in 0.1 bins and compared it to the unmatched LD calculations in wild barley. Surprisingly, the rate of LD decay of the matched SNP pairs did not markedly differ from the LD decay of all the SNP pairs (Fig. S8). On the contrary, LD decay curves reconstructed for the individual MAF-matched bins revealed that the dynamics of LD decay was frequency dependent. In both subspecies, LD of the low-frequency alleles decayed faster than of the medium-frequency alleles (Fig. 6AB). The number of SNPs in the MAF bins greatly varied, e.g. 19,772 and 3,510 SNPs in the 0.1 and 0.2 MAF bins, respectively. The bootstrapping revealed that this difference in the number of SNP did not notably affect the LD calculations (Fig. S9). To further test the robustness of the LD estimates, I compared the LD decay curves obtained using the standard dataset, which contains less than 50% of missing data, and the dataset without any missing data (Fig. 6A). This experiment revealed that the varying amount of missing data did not significantly alter the LD decay estimates.

It has been suggested that low-frequency alleles are generally younger than medium frequency alleles and therefore may reside on longer LD blocks owing to the lack of historical recombination (Chakravarti, 1999; Slatkin & Rannala, 2000). To illustrate this, I compared the distributions of distances between perfectly correlated alleles across the allele frequency spectrum (Fig. 6CD). Indeed, in the domesticated subspecies, the longer span of perfect LD was evident for low-frequency alleles and gradually decreased toward the medium-frequency alleles, indicating that low-frequency alleles span significantly larger distances and may thus occur on longer LD blocks. Intriguingly, in the wild subspecies, the median distance between perfectly correlated SNPs was similar to that of the domesticates; however, the similar relationship between MAF and the size of the LD blocks was not observed.

2.4 Effect of domestication on genetic diversity

It has been long recognized that domestication results in loss of genetic diversity via the so-called domestication bottleneck (Doebley et al., 2006). To unravel effects of domestication on barley genetic diversity, 315,892 biallelic SNPs were used to describe and compare various population genetics parameters in the wild and domesticated barley subgroups. Wild barley comprised ~ 7x more segregating sites than domesticated barley (Table 2) and ~88% of the sites resided in the non-coding regions. As measured by Watterson's θ_w, an unbiased estimator,

which provides correction for the sample size, the mutation rate in wild barley (θ_w=7.36 x 10^{-3}) was 5x higher than in the domesticates (θ_w=1.47 x 10^{-3}). Nei's nucleotide diversity π_n (also known as θ_π), an estimator of an average number of pairwise differences between two randomly drawn sequences per nucleotide, suggested that the domesticates (π_n=1.53 x 10^{-3}) retained only ~52% of the nucleotide diversity found within wild barley (π_n=2.97 x 10^{-3}). These findings suggested that the distributions of allele frequencies greatly differed between the wild and domesticated barley. Indeed, the MAF spectra revealed enrichment of rare alleles in wild barley and a strong shift toward common alleles in domesticates (Fig. 7). The diversity of the CDS polymorphisms was ~63% of the diversity of the non-coding SNPs in both wild and domesticated species. In addition, a share of the rare CDS SNPs was consistently higher than that of the rare non-coding polymorphisms.

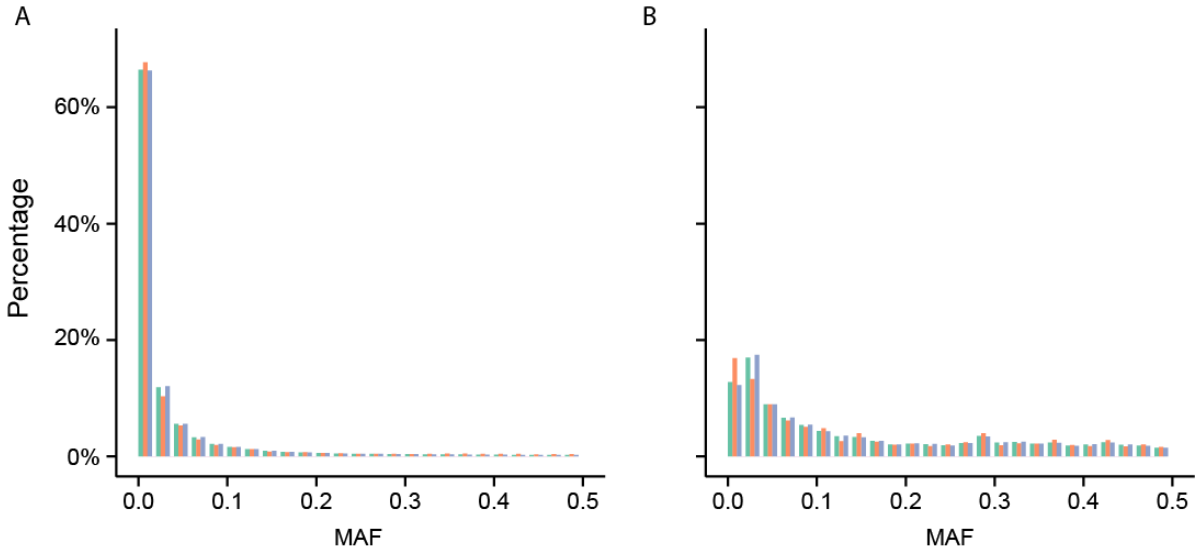


Figure 7. Distribution of minor allele frequencies (MAF) in wild (A) and domesticated (B) barley.

The folded spectra include SNPs with < 50% of missing data, including singletons. Cyan, red and blue bars represent the MAF distributions in all, coding and non-coding SNP subsets, respectively.

In wild barley, the strongly negative genome-wide value of D (-1.908) provided strong evidence for the enrichment of rare alleles, whereas the higher D value in domesticates (0.147) confirmed the shift of MAF toward the common alleles (Table 2). The difference between the D values in wild and domesticated barley was consistent along the individual chromosomes (Fig. S10). Coalescent simulations performed with the assumption of no selective pressure and the characteristics of idealized Wright-Fisher population estimated the range of neutral variation in D between -1.74 and 2.46 for the wild population (p-value < 0.01). Intriguingly, crossing the simulated thresholds, the basal levels of D in wild barley

rejected assumptions of the neutral model. Significantly negative D values are traditionally interpreted as the evidence of either the background (negative) selection (BS) or of a departure of population characteristics, chiefly in terms of demographic history and recombination properties, from the Wright-Fisher's assumptions. The latter seemed to play a bigger role in shifting D to the negative values, which were observed in this study (see Discussion). To evaluate the influence of the considerable amount of missing data present in the dataset and of the singleton alleles, which are the fraction most prone to false positive SNP calls, I estimated D for the wild barley SNP subsets without missing data and separately without singletons. The complete removal of missing data resulted in even lower D value (-2.49), whereas cutting off singletons expectedly elevated D (-1.18), which nevertheless remained strongly negative.

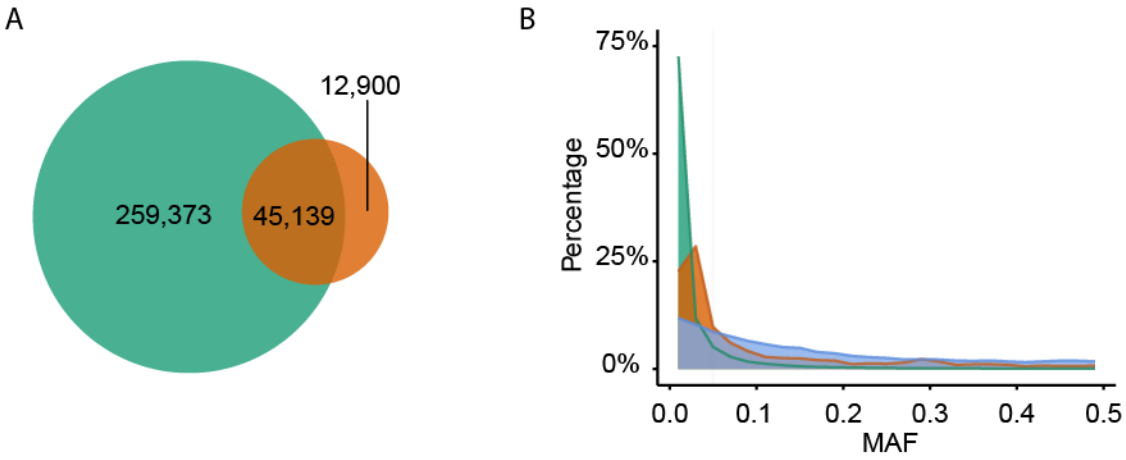


Figure 8. Private and shared SNPs in wild and domesticated barley.
(A) Distribution of private and shared SNPs in the subgroups of wild (green) and domesticated (orange) barley.

(B) The folded minor allele frequency (MAF) spectra of shared, private wild and private domesticated SNP alleles are shown by blue, green and orange ribbon plots, respectively.

Fay and Wu's H_{norm} test requires discrimination between low- and high-frequency polymorphisms and thus allows detection of hitchhiking events, resulting from the episodes of positive selection also known as a selective sweep. The negative and positive H_{norm} values distinguish, respectively, enrichment and depletion of the high-frequency polymorphisms compared to the neutral expectations. In this study, the allelic ancestral status, which allows distinguishing between low- and high-frequency alleles, could be assigned to 64,977 SNPs based on the status of the alleles in two wild *Hordeum* relatives. The amplitude of interspecies variation of H_{norm} was low compared to D (Table 2). In both wild and domesticated subspecies, the H_{norm} values were lower in coding than in non-coding SNPs, which suggest a

link between functionality and the abundance of the high frequency derived alleles.

In wild barley, the private alleles constituted 81% of the total number of SNPs, whereas the share of private alleles in domesticates was much lower (14%). Wild barley contained ~ 27x more private polymorphisms than the domesticates (Fig. 8A). To verify these ratios, I estimated the influence of the unbalanced numbers of genotypes in the wild and domesticated subgroups on the amount of discovered private alleles. This experiment revealed that, in the equalized wild barley subsets, the private allele ratios, which varied between 57% and 65%, remained markedly higher than in the domesticates (Fig. S11). MAF distributions of the private alleles showed patterns similar to those of the complete SNP datasets. On the contrary, the MAF distribution of the shared alleles was severely skewed toward the more common alleles (Fig. 8B). This strongly suggests that among the shared polymorphisms most alleles are identical by descent, i.e. originating from the common ancestor, rather than by state, i.e. an independent occurrence of the same mutation in both subspecies.

The analysis presented above described only intrapopulation diversity. To assess the proportion of diversity due to the divergence between wild and domesticated barley, I calculated the fixation index, also known as Fst statistics, genome-wide and separately for the individual SNPs. The Fst values range from 0 to 1 corresponds respectively to the absence of and to complete differentiation, i.e. no alleles are common between populations. The genome-wide divergence between wild and domesticated barley (Fst=0.29) was to the previously reported differentiation between wild and domesticated populations of barley (Fst=0.26; Russell et al., 2011), soybean (Fst=0.29; Zhou et al., 2015) and rice (Fst=0.27; Huang et al., 2012), but higher than that of maize (Fst=0.11; Hufford et al., 2012). The differentiation between the non-synonymous polymorphisms (Fst=0.27) was higher than that of the synonymous SNPs (Fst=0.25), suggesting the action of adaptive selection under barley domestication. It has been demonstrated in tomato that individual chromosomes may disproportionately contribute to the divergence between populations with the dramatically unequal distribution of highly divergent sites (e.g. Lin et al., 2014). However, such patterns were not evident in this study (Fig. S10). The Fst patterns did not visually differ between the chromosomes, whereas the statistical analysis revealed slightly elevated mean levels of divergence of chromosomes 4 and 7 (p-value < 0.01). The question whether these chromosomes exhibit peculiar patterns of elevated Fst on a finer scale, i.e. at specific loci, will be addressed in future studies.

2.5 Footprints of domestication-related selection

The genetic studies of barley domestication have provided several isolated examples of mapping domestication-related traits and of isolation of the underlying genes and characterization of their diversity in the domestication context (reviewed in Pourkheirandish and Komatsuda, 2007; Comadran et al., 2012; Jones et al., 2008; Pourkheirandish et al., 2015). However, the genome-wide architecture of a domestication syndrome has not been investigated as yet. Selection under domestication generates a whole variety of molecular signatures, which may escape detection if only a single test of selection, revealing particular patterns, is applied. Therefore, in this study, I employed a combination of tests, namely genome scans using mean r^2, H_{norm} and π_w/π_d tests, to obtain a catalog of candidate regions that likely experienced selection under domestication.

Patterns of LD vary along chromosomes due to the heterogenous recombination rate at different structural domains of a chromosome. On the other hand, both selective sweeps and background selection may result in elevated local LD manifested by positively correlated recombination and nucleotide variation rates (Maynard and Haigh, 1974; Charlesworth et al., 1993) and distinction between the two in the regions of low recombination has long been a subject of debates (reviewed in Stephan, 2010). In this study, LD strongly negatively correlated with Nei's nucleotide diversity (Pearson's r= -0.68; p-value < 0.001). Assuming that structural properties of the chromosomes, e.g. location of centromeres, are not likely to vary between closely related subspecies, the outliers in the local LD patterns, distinguishing wild and cultivated barley, tentatively represent the instances of selection. To identify such outliers, I performed genomic scans of the rolling mean r^2 values along the individual chromosomes. Consistently with the genome-wide LD estimates, along each chromosome, the average LD was remarkably higher in domesticated than in wild subspecies (Fig. 9). Moreover, the patterns of LD were heterogeneous along the chromosomes, and, in the domesticates, the amplitude of variation was much higher than in wild barley. The scan of the z-normalized r^2 distributions identified twelve regions on chromosomes 1H, 2H, 3H, 4H and 5H, which significantly deviated from the mean values in either wild (four regions) or domesticated (eight regions) subspecies (Fig. 9; Table 3). It is noteworthy, that in wild barley two of the LD outliers co-located with the major flowering time regulators *PpdH1* and *VRN-H1*. Whereas, in the domesticates, two of the outliers were found in the vicinity of the genes that determine row type of a spike, *HvTB1* and *HvVRS1*. Interestingly, in all the cases, the outliers had

elevated average r^2 and no deviating regions of decreased r^2, which could indicate an increase in the recombination rate in either of the subspecies, were identified.

The power of selection tests that measure deviations of allele frequency distributions from the neutral expectations is often compromised by the demographic factors. Here, the distributions of the D values in wild and domesticated barley displayed sensitivity to the confounding demographic histories, whereas the H_{norm} test appeared to be more robust. The screen for the outliers of H_{norm} identified eight regions of 10 - 31 cM, where the H_{norm} values in domesticated barley crossed the statistical thresholds of neutral variation (Fig. 10; Table 3). These regions comprised from 60 to 1718 'high confidence' genes. To refine the regions, I computed H_{norm} values for the individual loci separately for the complete sequences and only for CDS. This screen identified 147 and 10 outlier genes with significant enrichment of high frequency polymorphisms in domesticated and wild barley, respectively (Table S4). The outlier genes resided in all the sweep regions as well as the outside of them, suggesting additional regions, which were targeted by positive selection. Interestingly, the spike row-type gene, *HvRA2*, and the brittle rachis genes, *Btr1* and *2*, which closely collocate on 3H (40 – 45 cM), resided within the selected region. The sweep region on the chromosome 7H comprised another domestication-related gene NUD, which underlies hull-less phenotype of domesticated barley.

A selective sweep is generally defined as the reduction of nucleotide diversity around the positively selected mutation. Therefore, a significant reduction of nucleotide diversity at specific loci in domesticated compared to the wild barley, which is measured by nucleotide diversity reduction index π_w/π_d, is a footprint of the positive selection under domestication. It is noteworthy that the π_w/π_d scans seem to produce results markedly similar to the cross-population composite likelihood ratio (XP-CLR) scan (Zhou et al., 2014); the XP-CLR test is designed to provide the best spacial resolution but requires resolved physical map for implementation (Chen et al., 2010b). Here, the π_w/π_d scan discovered 11 putative sweep regions varying in breadth from 10 to 23 cM with the π_w/π_d values from 4 to 9 (Fig. 11; Table 3). Three of the π_w/π_d sweep regions on the chromosomes Hv 3 and 7 largely overlapped with three of the H_{norm} signature, including the regions containing the *HvRA2-Btr1/2* and *NUD* loci. 91 of the individual genes were statistical outliers with the π_w/π_d index in the range from 22 to 122 (Table S4). Unlike the H_{norm} sweep regions, several π_w/π_d outlier regions exhibited elevated π_w/π_d but did not comprise individual outlier genes. This discrepancy was apparently due to the much larger range of π_w/π_d variation in the individual genes compared to the sliding windows; therefore higher statistical thresholds defined the outliers. In addition, the individual

genes underlying some of the π_w/π_d sweeped regions may be missing from the capture assay.

The literature search yielded 13 genes, which have been implicated in the regulation of the domestication-related traits and some of them carried patterns of selection (see Introduction). Of these, 9 have been included in the enrichment assay and were polymorphic in this study. Most of these genes resided outside the detected sweep regions and did not carry the signatures of selection on their own. Locations of the *HvRA2* and *NUD* genes matched those of the sweeps, but neither of the tests identified them as putative selection targets (π_w/π_d: *HvRA2*, 1.6; *NUD*, 0.7). Not only the number of outlier genes differed between the screens, but the patterns of their distribution along the chromosomes were also different. With the few exceptions, the π_w/π_d outliers formed the characteristic peaks or stacks of genes, where the test values gradually increased. Conversely, in many cases, the H_{norm} outliers did not form such clusters and were scattered along the chromosomes. Whereas the larger regions identified by different test markedly overlapped (~ 50% in each test), in the tests for individual outliers only 8 out of 229 genes overlapped (Fig. 12).

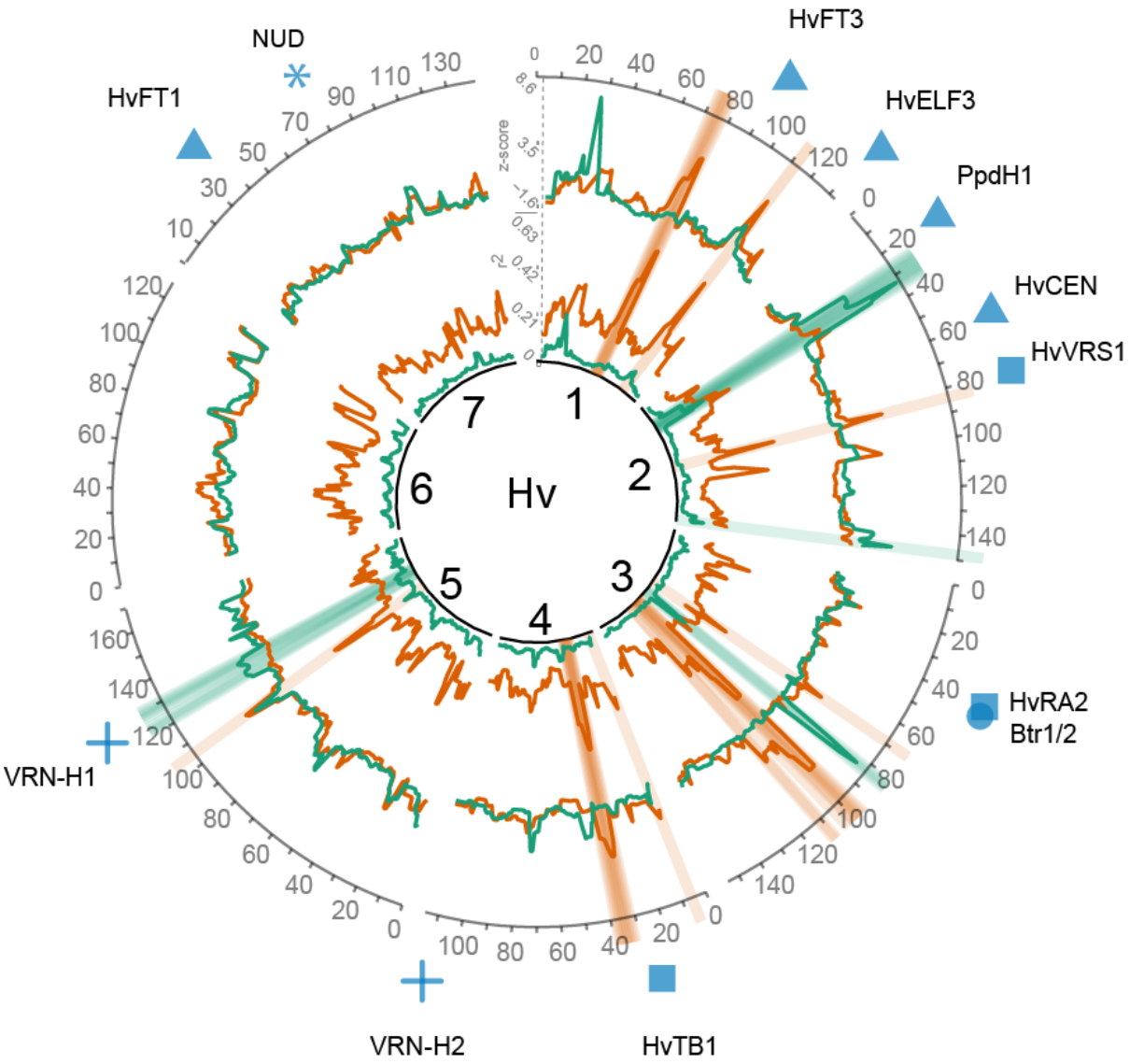


Figure 9. Domestication-related selection signatures based on patterns of linkage disequilibrium (LD).

The numbers inside the innermost circle indicate barley linkage groups and genetic distances in cM are shown on the outermost gray scale. LD patterns (sliding window 10 cM, 1-cM step) are shown for the wild (green) and domesticated (orange) barley subgroups. The inner circle displays the original data, whereas the outer circle shows the data normalized by the z-scores. The selection sweeps defined as significant deviations ($p < 0.05$) in LD between wild and domesticated subgroups are shown by green and orange segments, respectively. Genetic locations of barley genes that have been implicated in domestication and adaptation are shown as the blue triangle, plus, square, round and star symbols, depicting flowering, vernalization, row-type, brittleness and naked caryopsis genes, respectively.

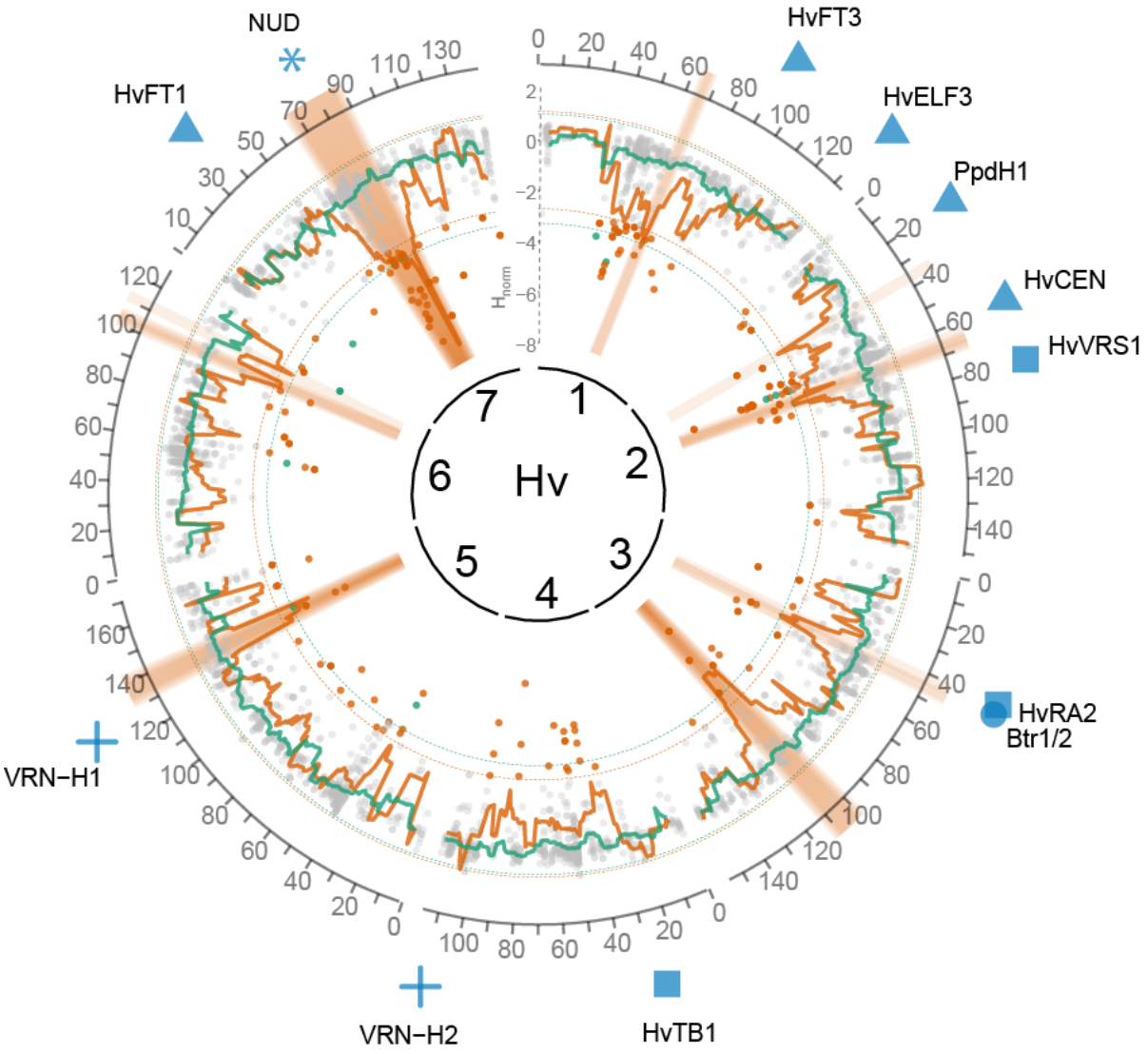


Figure 10. Domestication-related selection signatures based on patterns of normalized Fay & Wu's H values (H_{norm}).

The numbers inside the innermost circle indicate barley linkage groups and genetic distances in cM are shown on the outermost gray scale. Distribution of H_{norm} along the chromosomes (sliding window 10 cM, 1-cM step) is shown for wild and domesticated subgroups by the green and orange lines, respectively. The window-based selection sweeps in domesticated barley are shown by the orange segments. The H_{norm} values of the individual loci that exceed the significance thresholds are shown by the green and orange points for the wild and domesticated subgroups, respectively. Whereas, the H_{norm} values below the thresholds are shown in gray. The green and orange dashed lines are the thresholds of H_{norm} neutral variation (p-value < 0.01) in the wild and domesticated barley, respectively. Genetic locations of barley genes that have been implicated in domestication and adaptation are shown as the blue triangle, plus, square, round and star symbols, depicting flowering, vernalization, row-type, brittleness and naked caryopsis genes, respectively.

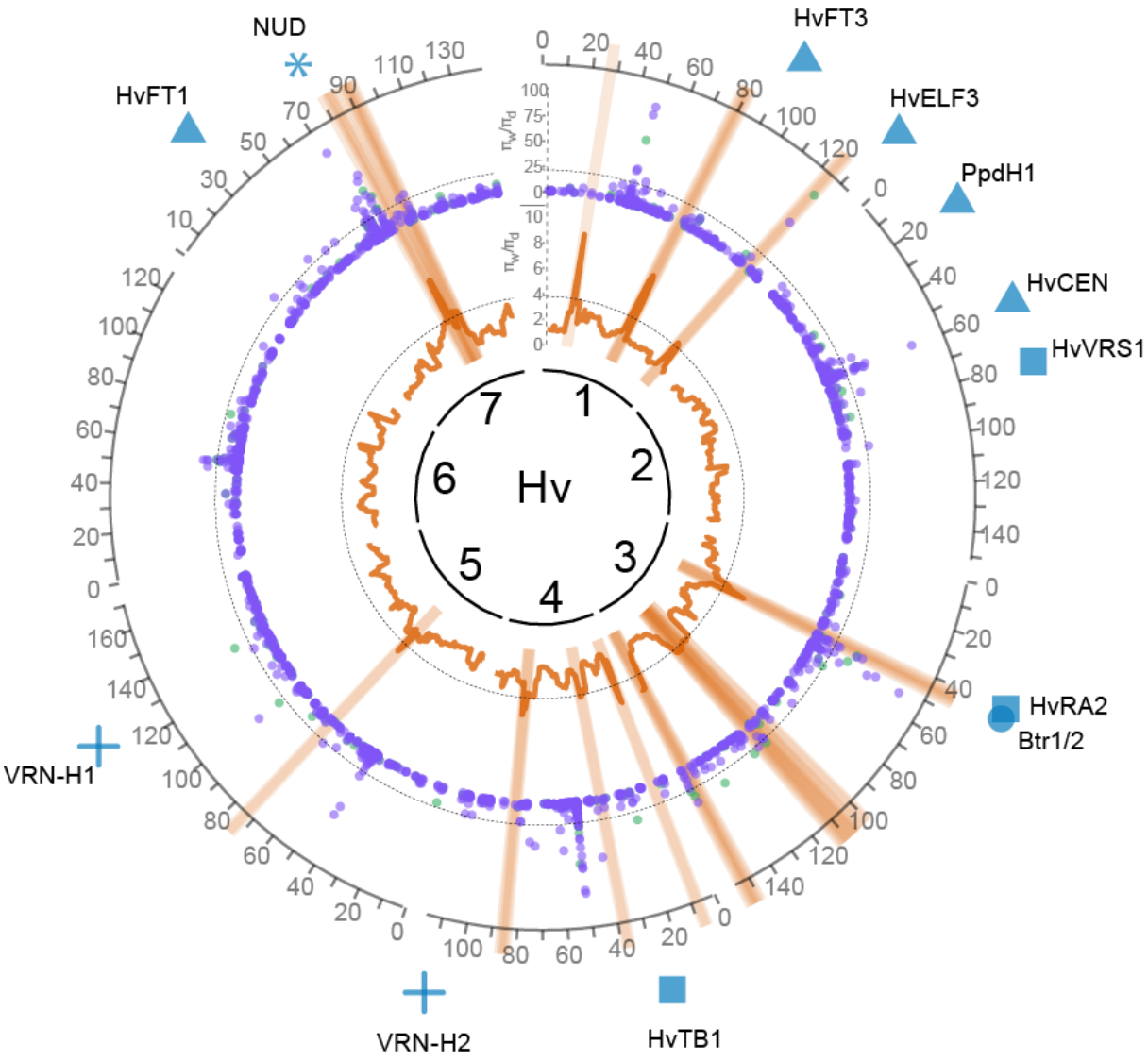


Figure 11. Domestication-related selection sweeps based on reduction of genetic diversity in domesticated (π_d) compared to wild (π_w) genotypes.

The numbers inside the innermost circle indicate barley linkage groups and genetic distances in cM are shown on the outermost gray scale. Distribution of π_w/π_d ratio along the chromosomes (sliding window 10 cM, 1-cM step) is shown as the orange line. The window-based selection sweeps are shown by the orange segments. On the outer circle, the points indicate π_w/π_d ratio values for individual loci, where the lightgreen and violet points indicate the ratios obtained from the SNPs located in the CDS and in the entire locus, respectively. The dashed lines are the thresholds defining the statistical outliers at p-value < 0.05. Genetic locations of barley genes that have been implicated in domestication and adaptation are shown as the blue triangle, plus, square, round and star symbols, depicting flowering, vernalization, row-type, brittleness and naked caryopsis genes, respectively.

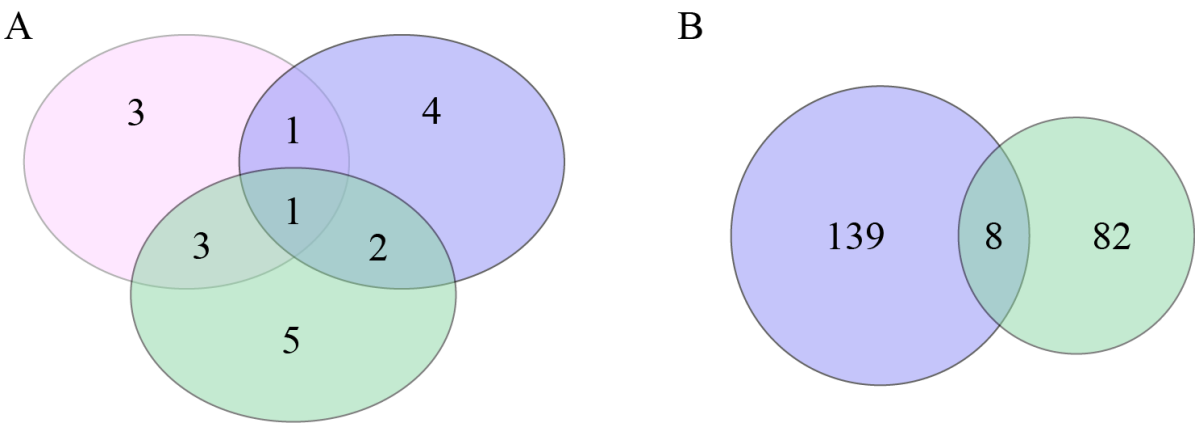


Figure 12. The overlap of the selection signature identified by different methods.
A number of candidate domestication regions (A) and genes (B) identified by the mean r^2, Fay and Wu's H_{norm} and π_w/π_d methods as shown by pink, blue and green circles, respectively.

3 Discussion

3.1 > 300,000 unascertained SNPs identified by the reduced representation deep sequencing

Systematic description of genetic diversity and population structure of a species requires reliable genotype information from a multitude of loci surveyed in a representative set of individuals. Recent advancements in high-throughput genotyping and sequencing facilitated generation of such polymorphism panels in many plant species, including large-genome cereals, such as barley and wheat (Henry et al., 2014b; Close et al., 2009; Winfield et al., 2012). In barley, reduced representation sequencing methods, such as genotyping by sequencing and exome sequencing, are gaining popularity due to the advent of the advanced barley genome resources (IBGSC, 2012; Mascher et al., 2013; Poland et al., 2012). In this study, I further reduced the complexity of barley gene space by selecting and sequencing genes related only to specific developmental pathways, for example, the flowering pathway, which contains candidate loci relevant to domestication and local adaptation (Buckler, 2009; Izawa, 2007; Verhoeven et al., 2008). Many of the traits that vary between wild and cultivated barley are related to the developmental aspects of phenotype, therefore the genes implicated in developmental processes are probable targets of selection.

It has been shown that the hybridization-based enrichment assays, and in particular NimbleGen SeqCap, are prone to generate off-target reads in the human exome capture assays (Bodi et al., 2013). In the human exome sequencing, large high-quality SNP datasets that originate from the off-target enrichment regions have been documented (Guo et al., 2012). Likewise, in this study, the size of the off-target captured regions yielding high-quality SNPs was approximately six times larger than the size of the selected target capture. However, most of the off-target captured regions resided outside of the predicted CDS. To investigate the origin of a large number of heterozygous SNP calls, which was unexpected in barley as a self-pollinated species, I performed the comparison of the pooled and individually sequenced DNA samples. The results suggested that the heterozygous calls apparently originated from the stacking of conserved paralogs, missing from the barley reference genome, rather than from the artefactual “jumping PCR”, commonly contaminating samples that were enriched in pools (Kircher et al., 2011). These findings strongly suggest that the barley genome contains multiple probably pseudogenized copies of the developmental genes, which, according to the earlier proposed model, presumably associate with the mobile elements (Wicker et al., 2011).

The transition to transversion bias also referred to as Ti/Tv ratio is a characteristic of mutational dynamics of genomes. This measure varies between different species and genomic regions. It also serves as a SNP calling quality parameter; it should not strongly deviate from the species average values (Guo et al., 2013). The estimation of the Ti/Tv bias is important not only for our understanding of species-specific patterns of variation, but also for tailoring nucleotide substitution likelihood models used in the selection tests such as the synonymous to non-synonymous substitution ratio (K_a/K_s, also known as d_n/d_s or ω) (Yang and Bielawski, 2000). In most species, transitions are more common than transversions, whereas they constitute only half the number of transversions in the unbiased scenario. In this study, the Ti/Tv genome-wide ratio (2.48) was in par with the genome-wide Arabidopsis estimates (2.4) but higher than previously reported values in other crops (0.92 – 1.83) (Bus et al., 2012; Morton et al., 2006; Ossowski et al., 2010; Simko et al., 2006; Zhu et al., 2003) and in barley ESTs (1.15) (Duran et al., 2009).

To suit the population genomics analyses, the SNP dataset should not suffer from systematic biases that compromise measures of genetic diversity and relatedness, since these tests often rely on accurate estimates of the population allele frequencies (Clark et al., 2005; Nielsen et al., 2011). Whereas the sequencing-based SNP discovery assay is generally void of the ascertainment bias, which is inherent in SNP chip arrays, the other sources of bias, e.g. sequencing error and mapping bias, have been documented (Brandt et al., 2015; Johnson and Slatkin, 2008). These biases may lead to overestimation of reference allele frequencies but can be partially tackled by adjusting the stringency of the SNP calling and filtering algorithms. Here, both the Ti/Tv ratio and the MAF spectra, which did not noticeably deviate from the expected neutral distribution, did not reveal any strong systematic biases in the SNP discovery procedure. It is noteworthy that the assay provided approximately 45-fold more SNPs than the most commonly used barley iSelect chip (Close et al., 2009) but only a two-fold improvement in resolution, apparently due to the mapping of multiple SNPs to the same position. This indicated that the resolution of the assay will be greatly improved with the update of the barley physical and genetic maps.

3.2 Structure analysis revealed cross-pollination between wild and cultivated barley

Both wild and domesticated barley have long been regarded as predominantly selfing species (selfing rate > 99%) (Brown et al., 1978; Wagner and Allard, 1991), dispersing pollen only at a short distance, which has been estimated not to exceed 50 meters (Ritala et al., 2002).

Nevertheless, several studies provided molecular evidence that outcrossing between these subspecies occasionally occurs in the overlapping habitats (Abdel-Ghani et al., 2004; Badr et al., 2000; Pourkheirandish and Komatsuda, 2007; Russell et al., 2011). Hübner et al. (2011) proposed that the historical gene flow resulted in a substantial exchange of alleles between wild and domesticated barley. However, the gene flow appeared to be unidirectional, from the domesticated to wild populations. In this study, the wild and domesticated barley were clearly genetically differentiated with admixed genotypes present in both groups. The patterns of admixture suggested that the gene flow occurred in both directions. The rate of admixture in the wild and domesticated barley (17%) was ~10 times higher than the previously reported values (1.8%) (Russell et al., 2011) estimated using the genome-wide BOPA1 SNP panel. Russell et al. (2011) explained the observed admixture patterns as the outcrossing between the landrace and wild barley in sympatric stands within the Fertile Crescent. Jakob et al. (2014) proposed an alternative scenario for the admixture origin, where most of the outcrossing events occur in course of the *ex situ* reproduction in germplasm banks. The outcrossing rate of the wild barley from the germplasm banks (22%) was even higher than the rate estimated in this study (12%). The analysis presented here provided evidence for both hypotheses.

The landraces were significantly overrepresented among the domesticated admixed genotypes (Fisher's exact test p-value < 0.001). The seeds of these landraces were obtained from different genetic collections and, according to the passport data, most of them originated from the areas, where wild barley populations have been observed (Harlan and Zochary, 1966; Molina-Cano et al., 1987; Xu, 1982). Unlike advanced cultivars, landraces are rarely selected for uniformity (Ceccarelli et al., 1987), therefore admixed alleles that originate from wild barley may persist in cultivated stands. Based on these facts, it is tempting to suggest that systematic introgression of wild barley alleles into landraces occurred *in situ* from the secondary contact. This may be an example of the evolutionary mechanism, which increases the adaptive potential of the landrace genotypes (Warschefsky et al., 2014).

On the other hand, two facts supported the hypothesis of the *ex-situ* origin of the admixture. First, wild genotypes from the so-called HID collection were significantly overrepresented in the admixed wild barley subgroup (Fisher's exact test p-value < 0.001). This correlates with the timescale of the genbank reproduction history of the two collections. The HID genotypes originated from the ICARDA and USDA collections sampled in the years 1960-1990 (Badr et al., 2000), much earlier than the Barley1K set, which was collected by Hübner and associates in 2007 (Hübner et al., 2009). Second, the estimated admixture rate in

the wild barley greatly exceeded the previous *in situ* gene flow estimates and more closely matched those for the 'genebank' genotypes (Jakob et al., 2014). To summarize, I found that admixture between wild and domesticated barley germplasm is more common than previously suggested and, if not accounted for, apparently may lead to erroneous interpretations of the diversity patterns.

3.3 Diversity analysis predicted recent domestication bottleneck

The genetic bottleneck is a major demographic event associated with domestication, in which only a limited number of allelic combinations are passed onto the domesticated genepool (Doebley et al., 2006). It is associated with the reduction of genetic diversity and generates skewed patterns of LD and allele frequencies in the bottlenecked populations (Eyre-Walker et al., 1998; Flint-Garcia et al., 2003; Tenaillon et al., 2004). Domesticated barley showed more differentiation from its wild progenitor and retained less nucleotide diversity (50%) than other crop species such as maize (83%), rice (80%) and common bean (83%), which were characterized by the genome-wide assays (Huang et al., 2012; Hufford et al., 2012; Schmutz et al., 2014). Whereas, in soybean and tomato, retention of diversity (~50%) in domesticates was similar to the values reported here (Lin et al., 2014; Zhou et al., 2014). In barley, previous diversity retention estimates varied dramatically in the range from ~36% to ~81% (Caldwell et al., 2006; Fu, 2012b; Kilian et al., 2006; Morrell et al., 2013; Saisho and Purugganan, 2007). Considering the inflated variance of diversity along the chromosomes of the domesticates compared with the wild barley, such variation of the diversity estimates, based on the analysis of a few isolated loci, is expected and apparently stems from the arbitrary selection of genes and a small number of surveyed genotypes.

LD increased dramatically as a result of barley domestication; LD decayed to the background levels at ~ 20 times larger distances in the domesticates than in the wild barley. These results are in line with the previous notions of low and extreme LD in wild and cultivated barley, respectively (Caldwell et al., 2006; Morrell et al., 2005). The increase of LD in barley domesticates was higher than in domesticated maize, rice and soybean (~ 4x – 7x), but much smaller than in *lycopersicum* tomato (~ 30x – 98x) (Huang et al., 2012; Hufford et al., 2012; Lin et al., 2014; Zhou et al., 2014). Among these crops, only maize is considered an outcrosser; the others are mostly inbred species with varying degree of self-pollination. Therefore, in this study, the link between the mating system and dynamics of LD and diversity

under domestication was not evident. Based on these results, I propose that the demography of domestication, for example, strength of the bottleneck, could have played a larger role than mating systems in shaping diversity of crops. This hypothesis finds support in a study of intraspecies LD variation in *Arabidopsis thaliana*, where extreme LD variation between the populations (1 - 50 cM) has been attributed to the founder effect (Nordborg et al., 2002). It is necessary to mention that the variety of parameters used to measure LD in the aforementioned studies, such as population recombination rate, average haplotype length, r^2 drop to a half of the maximum value or to the background level, hinders direct comparisons of the LD properties between species.

In both barley and maize, the genetic and archaeological evidence suggests that domestication dates back approximately 10,000 years (Piperno et al., 2009; Saisho and Purugganan, 2007; van Heerwaarden, 2011; Zohary and Hopf, 1993). However, in contrast with maize, the depletion of rare alleles in domesticated barley indicated that the variation has not shown signs of recovery after the bottleneck (Hufford et al., 2012). This suggests a tentative link between the mating system and rate of post-domestication recovery of variation (cf. Jarne, 1995).

Dependency of LD (r^2) on the allele frequency has been established in humans (Eberle et al., 2006); however, in plants, understanding of this phenomenon is extremely limited. In both Arabidopsis and barley, the rate of LD decay negatively correlated with the allele frequency (Gan et al., 2011), whereas, in human data, no difference between the different allele frequency bins was found (Eberle et al., 2006). It has been suggested that, compared to common alleles, rare alleles are younger and occur on longer LD blocks due to the lack of historical recombination (Chakravarti, 1999; Eberle et al., 2006). The decay curves did not reflect this dependency apparently due to the fact that the r^2 measure summarizes both recombinational and mutational histories. To illustrate the link between the allele age and frequency, I surveyed the average span of the perfectly correlated SNP pairs, as a proxy for age, across the whole range of allele frequencies. The negative correlation between these measures was evident in the domesticated but not in wild barley. Apparently, only in domesticated barley, a distribution of rare alleles of different ages is skewed toward the younger alleles. This goes in line with the scenario of the dramatic loss of rare alleles during the bottleneck and their subsequent recovery via young *de novo* mutations in the bottlenecked populations (Luikart et al., 1998). These findings draw an interesting parallel between the properties of LD in domesticated barley and humans, as a species that is thought to experience

several severe bottlenecks (Garrigan and Hammer, 2006). I suggest that the negative correlation between the allele frequency and the average distance between perfectly correlated SNP pairs, may be a signature of a population bottleneck.

Taken together, multiple signatures, such as the reduction of diversity, increase of LD, depletion of rare alleles, a small share of private alleles and inflated variance of π_d and Tajima's D along the chromosomes, strongly indicate the existence of a historical bottleneck in barley demographic history (Maruyama and Fuerst, 1985; McVean, 2002; Nei et al., 1975). The evidence from other crops suggests only minor improvement bottlenecks, which accompanied transitions from landraces to advanced cultivars (Cavanagh et al., 2013; Hufford et al., 2012; Zhou et al., 2015). Thus, I propose that the bottleneck observed in this study was associated with domestication but not with further improvement of the crop. The extent of diversity variation between the landraces and cultivars as separated groups requires further investigation.

3.4 Domestication targeted homologous developmental pathways in different crops

Selection under domestication causes beneficial alleles to reach high frequencies together with the surrounding variants in a process of selective sweep also referred to as genetic hitchhiking (Smith and Haigh, 1974). This produces regions of reduced nucleotide diversity and increased LD. The size of the sweep regions gradually reduces over time due to the breakdown of LD and the diversity gradually restores. Numerous tests have been developed to reveal selection and discern it from the effects of demographic history, which often mimic signatures of selection. The tests for selection differ in their power to detect various selection scenarios and their sensitivity to the confounding demographic factors (see reviews by Vitti et al., 2013; Nielsen, 2005). In this study, I used a combination of tests based on the LD and the allele frequency patterns to assess their performance in barley genome scans and catalog candidate regions and genes selected during barley domestication.

In wild barley, the negative genome-wide levels of D departed from the neutral model of DNA sequence variation in a way similar to reports in Arabidopsis (Nordborg et al., 2005; Schmid et al., 2005). Genetic and geographic structure and autogamy of wild barley populations violate the assumptions of the Wright-Fisher population (Hübner et al., 2009; Morrell et al., 2003). These factors are known to mimic signatures of background selection and apparently contributed to the observed shift in the D distributions toward the negative

values (see reviews by Charlesworth et al., 2003; Cutter and Payseur, 2013). Therefore, to detect departures from neutrality using the D test, it is critical to reassess the thresholds of D neutral variation by estimating the realistic demographic models for wild and cultivated barley in the future studies. The difference in the basal levels of D between the wild and domesticated barley indicated the sensitivity of this test to the demographic history of the populations, such as the genetic bottleneck. By contrast, another frequency-based test, H_{norm}, was less sensitive to the confounding demography. Simulation data confirms that the H_{norm} statistics is much less sensitive to the bottleneck than the D test and the compromising effect of the bottleneck diminishes rapidly with time (Zeng et al., 2007).

In their review of the domestication syndrome (DS) concept, Abbo et al. (2014) propose that a unified glossary should be adopted in domestication studies in order to optimize interpretation of the genomic scans, searching for the footprints of domestication. To this end, they coined the concept of a crucial DS trait, which is present in a derived form in all domesticates and either segregates or is fixed in its wild form in the wild ancestor of a crop. Spike brittleness is the most studied example of such trait in barley, in which brittle and non-brittle spikes are strictly specific to domesticated and wild subspecies, respectively (Pourkheirandish and Komatsuda, 2007). On the other hand, improvement traits, such as row type, photoperiod sensitivity and responsiveness to vernalization, which have been studied in the domestication and adaptation contexts, segregate in domesticated barley (Cockram et al., 2007; Hemming et al., 2009; Turner et al., 2005). Here, the selective scans were designed to capture signatures of selection on the crucial DS traits. Indeed, the signature of positive selection was identified at the *Btr1/2* locus, which underlies the brittleness trait, but not in the loci associated with the improvement traits with one notable exception. The genomic region on the chromosome H7 that comprise the *nud* gene, controlling the naked (hulless) grain phenotype, carried the signature of selection (Taketa et al., 2008). However, the *nud* gene itself did not carry the selection signatures and thus was not the target of selection at this locus. This confirms the suggestion that naked grain phenotype is not a crucial DS trait, since both hulless and hulled genotypes are ubiquitously present in the domesticated barley genepool (Saisho and Purugganan, 2007).

Whereas responsiveness of barley to photoperiod and vernalization is an improvement trait, the other flowering-related traits, for example synchronous flowering of barley tillers, could be fixed in domesticated barley (Meyer and Purugganan, 2013; Doebley et al., 2006). Systematic analysis of such flowering crucial DS traits in barley is missing and therefore

flowering pathways and genes regulating these traits are not known. None of the candidate domestication genes identified in this study have been functionally characterized, however, a putative function can often be inferred from homology. In other crops, several flowering loci have been reported among the candidate domestication genes (Hufford et al., 2012; Schmutz et al., 2014). In common bean, the orthologs of photoperiod pathway regulator genes, encoding two different members of the same protein complex CONSTITUTIEVLY PHOTOMORPHOGENIC 1 (COP1) and CULLIN4 (CUL4) have been independently targeted by selection in two separate domestication events Mesoamerican and Andean, respectively (Chen et al., 2010a; Schmutz et al., 2014). Here, the ortholog encoding Arabidopsis SUPRESSOR OF PHYA 2 (SPA2), which is another member of the COP1 complex, carried the strongest signature of selection identified in the relative diversity scan (seq108; π_w/π_d = 122) (Zhu et al., 2008). This raises an intriguing question whether domestication involved independent modification of orthologous flowering pathways and genes in different crops. In maize, a homolog of Arabidopsis *AGAMOUS-LIKE20* gene, encoding SUPRESSOR OF OVEREXPRESSION OF CONSTANS1 (SOC1) protein, was a domestication candidate (mentioned as an ortholog of rice OsMADS56; LOC_Os10g39130) (Hufford et al., 2012). A barley ortholog of OsMADS56 (seq411) resided within a sharp selection π_w/π_d signal on the chromosome 1H (~50 cM) but did not carry a signature of selection itself. Intriguingly, three (seq104, 223, 231) out of ten barley homologs of GIBBERELLIN 2-OXIDASE (GA2ox) genes, involved in the biosynthesis of growth hormone gibberellin, were among the domestication candidates. Whereas, in maize, GA2ox gene was an improvement candidate (Hufford et al., 2012). These findings provide another example of tentative interspecies convergence of domestication targets.

Another notable domestication candidate, *HvGCN5* (seq612; π_w/π_d = 104), encoding a homolog of the GNAT/MYST histone acetyltransferases, has been implicated in regulation of seed maturation, dormancy and germination based on the expression analysis and its regulation by the phytohormone ABA (Papaefthimiou et al., 2010). Synchronous seed germination and reduced dormancy have been proposed as crucial DS traits (Abbo et al., 2014; Doebley et al., 2006).

Location of the *VRN-H1* gene, a key barley regulator of vernalization response, within the region of extended LD on the chromosome 5H in wild barley is noteworthy. Cockram et al. (2011) demonstrated that 98% of wild barley possess the wild-type winter *VRN-H1* alleles, which delay flowering until the vernalization requirement is fulfilled. In chapter 2 of this

dissertation, I demonstrate that the *VRN-H1* gene is tightly linked to several flowering-related genes, including *HvPHYC*, a homolog of Arabidopsis *PHYTOCHROME C* gene. The variation in *HvPHYC* modulates photoperiodic flowering and the mutant early flowering allele is private to the domesticates. It is tempting to speculate that, in this case, the extended LD at the *VRN-H1* locus in wild barley is a signature of the background selection, purging novel mutations and maintaining the integrity of this gene cluster, which is apparently critical for flowering.

3.5 Variety and number of selection signatures support the protracted model of domestication

In three selection scans based on the sliding-window LD, D and H_{norm} tests, the large selected regions notably overlapped between the scans. Conversely, the test for selection signatures in the individual genes displayed only small overlap of the outliers. This discrepancy may reflect the difference in the hypotheses behind the Fay&Wu's H_{norm} and relative diversity π_w/π_d tests. The H_{norm} measure tests for the enrichment of high-frequency derived alleles in the unfolded allele frequency spectra of a locus. It is assumed that, under the neutral scenario, relatively young derived alleles persist at the low frequencies over the long time period and reach intermedium and high frequencies at random locations due to the genetic drift. By contrast, under selection, the derived alleles flanking the beneficial mutation rapidly raise to the high frequencies – the signature captured by the H_{norm} test. Therefore, when H_{norm} suggests selection at a locus in domesticated barley, no assumption about the diversity of this locus in the wild barley is made. Conversely, the π_w/π_d test contrasts the diversity between the wild and domesticated subgroups and integrates the number of polymorphisms and their frequency. It captures a signal of selection only when the diversity is high in the wild but low in the domesticated barley. The highly diverse locus contains a larger pool of alleles segregating at a medium frequency than the one with low diversity. Thus, it is intuitive to assume that, compared to the H_{norm}, the π_w/π_d scan is more sensitive to the episodes of selection on standing variation, a scenario in which an allele that is beneficial and fixed in the domesticates already originated and persisted in wild population but without conferring immediate selective advantage (Barrett and Schluter, 2008). Selection on standing variation results in smaller sharper troughs of diversity, owing to the older age of such alleles and therefore their location within the smaller LD blocks. The finding that the signatures of selection were more compact in the π_w/π_d scan compared to the H_{norm} scan and frequently resided outside of the larger

selected regions identified by the sliding-window scans confirms this hypothesis.

Accumulating molecular and archaeological evidence has been shifting our view on the process of barley domestication from an early model of a fast monophyletic event toward a more complex protracted model of domestication (Badr et al., 2000; Allaby et al., 2008). The protracted model postulates slower gradual process of domestication and polyphyletic origin of the barley crop (Brown et al., 2009; Purugganan and Fuller, 2011). Accordingly, domestication traits may reach fixation in a cultigen on a varying timescale and archaeological data corroborate this assumption (Fuller, 2007). Distinctions between the two models and their conceptualization has been a subject of recent debates (Fuller et al., 2012; Heun et al., 2012).

The variety and high number of domestication loci discovered in this study indicates that selection for the crucial DS traits was not achieved in a single rapid event. The crucial DS traits might have evolved in episodes of selection separated in time and varied in intensity. Given the remarkably high proportion of common identical-by-descent SNPs in domesticated barley (~ 87%), it is reasonable to assume that these SNPs might originate from the distinct wild barley sub-populations. Indeed, Poets et al. (2015) demonstrated that the domesticated barley genomes consist of mosaic segments, which originated from the geographically distant and genetically distinct wild barley sub-populations. These findings advocate the protracted polycentric model of barley domestication (Allaby, 2015). The origin of the common alleles discovered in this study and assessment of the relevance of the historical introgressions to the formation of domestication traits are pending further investigation.

The ultimate goal of the selection scans is to identify the selected alleles, which underlie the domestication traits. The selection scans conducted under the conservative criteria and limitations of the current barley genetic map may underestimate the number of crucial DS loci in barley (Teshima et al., 2006). Nevertheless, a combination of tests, which examine various facets of selection, provided the initial appraisal of loci associated with barley domestication. The catalog of barley genomic loci affected by the domestication is instrumental in future efforts directed to map the crucial DS traits on the identified selected regions and to prioritize and functionally verify the candidate domestication genes.

In addition to being the first example of the enrichment sequencing applied to characterize the genome-wide diversity of barley in the domestication context, presented results carry practical implications for barley breeding. Due to the absence of cytogenetic crossability barriers, wild *spontaneum* barley constitutes the so-called primary genepool used

in breeding (Bothmer et al., 1995; Harlan and de Wet, 1970; Nevo, 1992; Pickering and Johnston, 2005). The goal of the introgressive breeding has been to introduce desirable agronomic traits, most notably disease resistance and tolerance to abiotic stresses, into cultivated barley genepool (Abbot et al., 1991; Eglinton et al., 1999); yet its global impact on barley improvement has been contested (Hajjar and Hodgkin, 2007). Another strategy would be to shift the focus of barley breeding from specific traits to a more global alleviation of detrimental effects of genetic bottlenecks and selfing on genetic diversity. In this case, exploiting wild germplasm may help introduce fitter alleles in cultivated barley at regions of low diversity and promote recombination of haplotype blocks. I demonstrated that the population genomics facilitates identification of loci that require enrichment and, therefore, I expect that the results of this study will inform future breeding programs.

4 Materials and Methods

4.1 Plant material

A panel consisting of 345 wild *H. vulgare* ssp. s*pontaneum* and 87 domesticated *H. vulgare* ssp. v*ulgare* lines and a single genotype of *H. vulgare* ssp. a*griocrithon* were selected from the MPIPZ and Barley1K wild barley collections (Badr et al., 2000; Hübner et al., 2007) to cover the entire range of the wild barley habitats (Fig. 5A; Table S3). The landrace barley genotypes were sampled from the ICARDA collection (Syria) of North African and Middle Eastern landraces and from the collection sampled in Central Turkey in 2004 by Özkan and associates (unpublished). The advanced barley cultivars were sampled to represent Northern European, East Asian, North American and Australian breeding programs. The largest part of the germplasm set, 98% of wild and 40% of domesticated barley genotypes originated from the countries of the Fertile Crescent area. The selection of domesticated barley originated from various breeding programs and represented the whole variety of cultivated barley lifeforms, namely two- (71%) and six-row (29%) genotypes with the so-called winter (45%) and spring (55%) growth habits based on the passport data.

Leaf samples for DNA extraction were collected from a single 3-week plant of every genotype. The DNA extraction was performed using DNeasy Plant Mini kit (QIAGEN, Hilden, Germany) following the manufacturer's recommendations with minor modifications of incubation times and with double amount of the input leaf material. The DNA samples were quantified using the NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and the DNA integrity was assessed using electrophoresis in the 0.8% agarose gel.

4.2 Selection of target genes

A set of genic sequences selected for the analysis of genetic diversity comprised a comprehensive subset of loci related to flowering time and development of meristem and inflorescences. Additionally, it contained a selection of genes related to carbohydrate metabolism, to agronomic traits putatively affected by domestication, e.g. tillering, seed dormancy; and to abiotic stress response (Fig. 1; Table S1). To select these genes, first, I mined the scientific literature for the genes implicated in the aforementioned processes and extracted corresponding nucleotide sequences from NCBI GenBank. Second, I selected

flowering genes from the other grass species, such as Brachypodium and rice (Higgins et al., 2010). Third, I extracted 259 Arabidopsis genes characterized by the flowering-related gene ontology (GO) terms, which have been confirmed experimentally (Table S5). The homologs of these genes were extracted from the NCBI barley UniGene set (Hv cDNA, cv. Haruna Nijo, build 59; Matsumoto et al., 2011) either by the BLASTN search (Camacho et al., 2008) at the e-value threshold of 1e-7 or, in case of Arabidopsis genes, by searching the annotation table downloaded from the NCBI UniGene server (ftp://ftp.ncbi.nih.gov/repository/UniGene/Hordeum_vulgare/). This table was further used to reciprocally extract additional Hv homologs based on the Arabidopsis gene identifiers. If the BLAST search failed to identify a reliable Hv homolog, the search was performed against a set of barley High and Low confidence genes (MLOC cDNA; IBGSC, 2011) and the HarvEST unigene assembly 35 (http://harvest.ucr.edu/).

Open reading frames (ORF) of Hv cDNA were predicted using OrfPredictor guided by the BLASTX search against Arabidopsis TAIR 10 database (Min et al., 2005). The predicted ORFs were aligned to the genomic contigs of barley cultivars Morex, Bowman and Barke using the Spidey algorithm implemented in the NCBI toolkit (Wheelan et al., 2001). The ORF of the selected sequences were categorized as complete or partial based on the presence or absence of putative start and stop codons. The complete complementary DNA (cDNA) were selected and, if the complete cDNA was absent, partial gDNA and cDNA were included in the dataset. For several genes with previously characterized intronic regions, e.g. predicted to contain regulatory elements, complete genomic DNA (gDNA) were selected. In the case when only partial cDNA was available, chimeric sequences were assembled from the Hv, MLOC and HarvEST cDNA using SeqMan software (DNASTAR Lasergene®8 Core Suite, Madison, WI, USA). The selected sequences were cross-annotated with NCBI UniGene Hv and IBGSC MLOC identifiers using reciprocal BLASTN (e-value cut-off 1e-05). In addition to the coding regions and introns, the selection contained sequences up to 3 kilobase pairs (Kbp) upstream of the predicted start codon, which presumably corresponded to regulatory promoter regions. The target selection workflow is schematically described on Fig. 13.

A set of 1000 additional HarvEST genes was randomly selected such that they had no homology to target genes as determined by BLASTN and evenly spread over all barley linkage groups according to the GenomeZipper map (Mayer et al., 2011). The 100-bp stretches of each of these genes were included in the enrichment library.

4.3 Enrichment design and sequencing

The target sequences were filtered and tiled with 100-bp selection baits using Nimblegen proprietary algorithm and the library of baits was synthesized as a part of the SeqCap EZ enrichment kit (design name 130830_BARLEY_MVK_EZ_HX3; Roche NimbleGene, Madison, WI). Barcoded Illumina libraries were individually prepared, then enriched and sequenced in 24-sample pools at the Cologne Center for Genomics facilities following the standard protocols.

4.4 Selection of mapping reference

The genic sequences from a variety of barley genotypes were used to design the enrichment library to ensure that the longest ORF and promoter regions were selected. However, most advanced physical and genetics maps have been developed for the barley cultivar Morex. Since mapping information is essential for the downstream analyses, the so-called Morex genomic contigs were used as a mapping reference provided that they comprised the entire regions tiled by the baits (Table S2). If such contigs were not available, the genomic contigs of the barley genotypes Bowman and Barke or the templates that were used for the bait design were included in the mapping reference.

Targeted enrichment assays are known to capture large amount of sequences, which are homologous to the selected targets but not included in the original enrichment design. To identify such regions, the Illumina reads from 10 randomly selected barley genotypes were mapped to the complete Morex genome reference set (IBGSC, 2011). All genomic contigs that had at least one read mapped to them were included in the mapping reference. This thinning of the complete Morex genome dataset helped avoid excessive computational load in the downstream steps of the SNP calling pipeline. These Morex contigs were masked with “N”s at the regions of longer than 100 bp that exhibited more than 97% homology with the original capture targets. Up to now, the so-called PopSeq and IBGSC ‘Morex” genetic maps provide the most complete information on the genetic positions of barley genomic contigs. Therefore, these maps were used to extract mapping positions of the reference sequences (IBGSC, 2001; Mascher et al., 2013a).

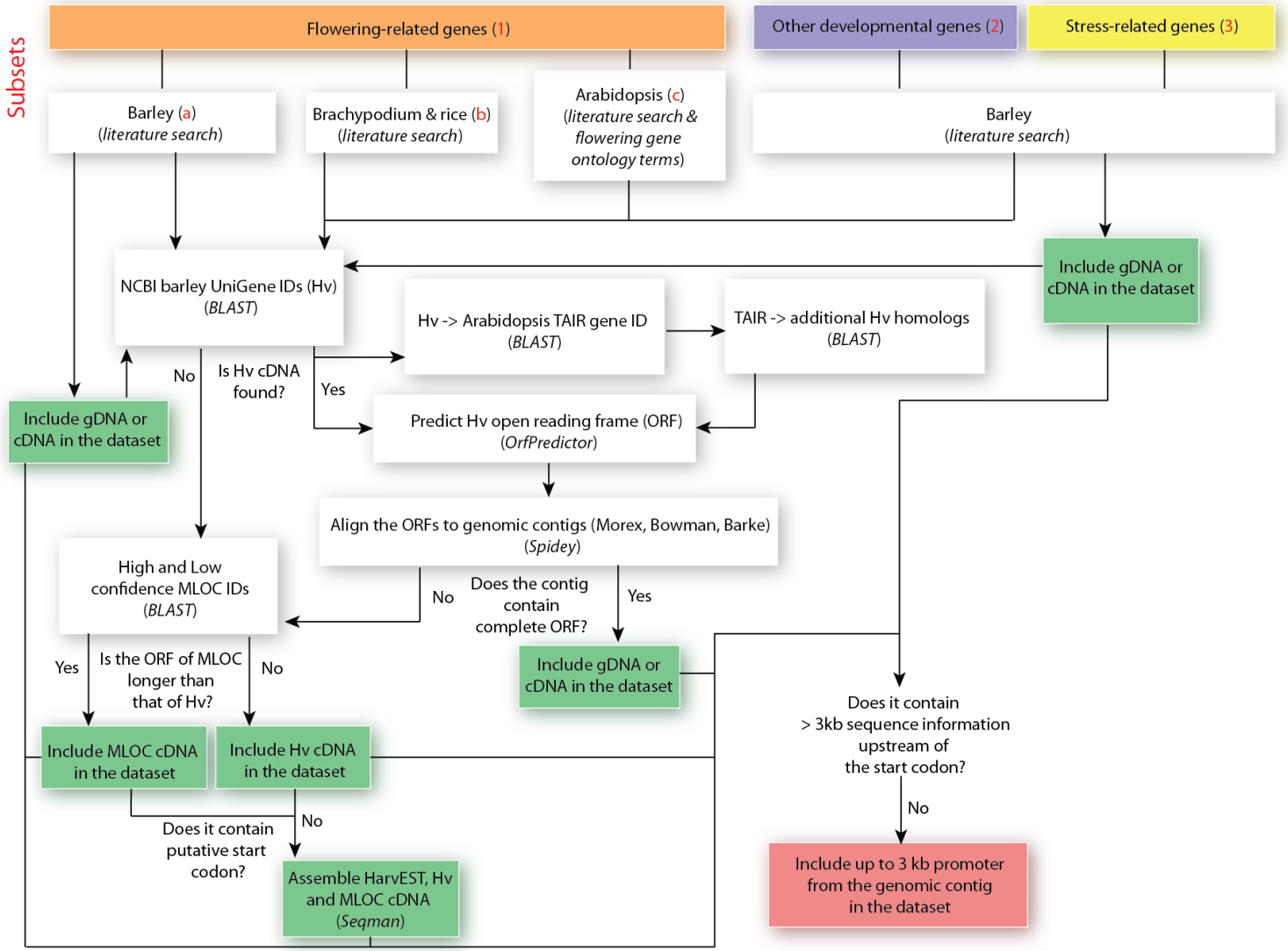


Figure 15. Selection of target genes. Subsets of three different functional categories of genes are highlighted in orange, violet and yellow. The output steps of the decision-making processes of selecting gene body sequences and promoter regions are highlighted in green and red, respectively.

4.5 Read mapping and SNP calling

The SNP calling pipeline consisted of three modules: quality control and filtering of Illumina read libraries; mapping the reads to the reference; and SNP calling, genotyping and filtering (a step-by-step pipeline is shown in Fig S11). The quality parameters of the paired-end Illumina libraries were assessed using FastQC tool (v. 0.11.2; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). After filtering out optical duplicates, resulting from a PCR amplification, using the CD-HIT-DUP software (v. 0.5; Fu et al., 2012a), the paired-end read files were merged and henceforth treated as a single-end dataset. Next, based on the FastQC results, the reads were trimmed from both ends to remove low quality sequencing data, filtered to remove the remaining adaptor sequences and low-complexity artifacts using the FASTX toolkit (v. 0.0.14; http://hannonlab.cshl.edu/fastx_toolkit/). The sequencing errors in the dataset were corrected using the Bloom-filter tool Lighter with the conservative set of parameters: k-mer size 23, alpha 0.2, and maximum corrections per read 2 (Song et al., 2014). The reference file was indexed for the downstream processing using Burrows-Wheeler Aligner 0.5.9-r16 (BWA), SAMtools and Picard tools (Li and Durbin, 2009; Li et al., 2009; http://broadinstitute.github.io/picard/). The groomed read datasets were mapped onto the reference genome using BWA (modules ‘aln’ and ‘samse’; Li and Durbin, 2009) with the following stringency parameters: missing probability (-n) 0.05, maximum number of gaps (-o) 2, and gap extensions (-e) 12. Some of the reference loci were present in the form of cDNA and the gDNA-derived reads mapped onto such targets may generate false positive SNP calls at the intron-exon junctions. To alleviate this problem, the reads that mapped to cDNA-derived targets were extracted, additionally trimmed by 14 bp from each end and remapped following the described procedure. Reads that mapped to several locations were filtered out.

The regions containing INDELs are prone to alignment errors and thus may generate false positive polymorphism calls. To tackle this issue, the reads were locally realigned around INDELs using RealignerTargetCreator and IndelRealigner walkers of the GATK suite (McKenna et al., 2010). Raw SNP calling was performed for each sample library separately using the GATK UnifiedGenotyper walker with the default parameters. Afterwards, the output lists of polymorphisms, the so-called VCF files, were merged into a multi-sample VCF file using the GATK CombineVariants walker. The *de novo* SNP discovery using GATK emits a call only if there was a nucleotide substitution compared with the reference genome without distinction between a reference allele and zero coverage (missing data). To obtain a dataset

containing both reference and non-reference calls, the genotyping mode of the GATK UnifiedGenotyper was applied to the individual bam files using the raw calls as the reference set of alleles. The output VCF files were merged into a multi-sample VCF file, which contained only the biallelic homozygous SNPs passing the following filters: depth of coverage (DP) > 8, mapping quality (MQ) > 20, Fisher strand (FS) < 60. For the downstream analyses, all heterozygous SNPs were assumed to be technical artifacts and treated as missing data. This pipeline was implemented in a series of bash scripts adapted for high-performance parallelized computation.

4.6 Characterization of the assay

To describe the capture quality parameters, two different sets of reference regions were defined as following: target capture regions tiled by the baits and the regions covered by the reads outside of the target and predicted capture regions. *De facto* captured regions were defined as those with the depth of coverage ≥ 8, set as the SNP calling threshold, in at least one of the samples. The depth of coverage was analyzed using bedtools v.2.16.2, vcftools v.0.1.11 and R (Danecek et al., 2011; Quinlan, 2014). Functional effects of the SNPs were predicted using SnpEff 3.6b software using the custom CDS coordinates as a reference genome (Cingolani et al., 2012). The CDS coordinates were mapped on the target genomic contigs based on the Spidey predictions and extracted from the IBGSC annotation file for the additional genomic contigs. Transition / transversion ratios (Ti/Tv) were calculated using VariantEval walker of the GATK package.

4.7 Population genetics analyses

Minor allele frequency (MAF) spectra for various genomic regions and bootstrapping of the rare SNPs (1000 random draws) were calculated using R. The SNPs were tentatively divided into neutral and non-neutral subsets defined by the SnpEff flags, which, for the neutral subgroup, carried the UTR, DOWNSTREAM, UPSTREAM, INTERGENIC, INTRON and SILENT SnpEff flags. The vcf files were converted into the ped format using tabix utility of Samtools, PLINK 1.9 (Chang et al., 2014). For estimations of population parameters, only a subset of SNPs with MAF > 0.05, missing data frequency (MDF) < 0.5 was selected. For the structure and principal component analyses (PCA), the SNPs in very high LD (r^2 >0.99) were pruned using PLINK 1.9. The PCA was performed using smartpca utility of the EIGENSOFT software version 5.0.2 (Patterson et al., 2006).

The linkage disequilibrium (LD) estimator r^2 was calculated for each SNP pair separately in the wild and domesticated barley subsets using PLINK 1.9. The background LD was defined as an average of the interchromosomal r^2 values (95th percentile). Rate of LD decay was estimated using a nonlinear least-square (nls) regression fit to the intrachromosomal or intergenic r^2 values using Hill and Weir's formula, providing adjustment for sample size (n):

$$E(r^2)=\left(\frac{10+C}{(2+C)*(11+C)}\right)*\left(1+\left(\frac{(3+C)*(12+12*C+(C)^2)}{n*(2+C)*(11+C)}\right)\right)$$

, where $E(r^2)$ is the expected value of r^2 and C equals to a theoretical parameter 4Nc, where N is the effective population size and c is the recombination frequency (see Hill and Weir 1986 for theoretical considerations underlying this formula). The nls regression analysis was implemented in R. The LD decay value was defined at the intersection point of the regression curve with the background LD. To estimate the robustness of LD estimated in unbalanced samples, i.e. varying number of individuals or markers, the balanced sub-samples were 1000x randomly drawn from the larger sub-group. Variation of the LD estimates in these bootstrap experiments was assessed using standard summary statistics. To screen LD patterns along the chromosomes the average r^2 values were calculated in sliding windows of 10 cM with 1-cM steps.

To identify admixture between wild and domesticated genotypes, the structure of barley populations was inferred using fastSTRUCTURE software, which implements Pritchard's STRUCTURE algorithm in a fast and resource-efficient manner (Raj et al., 2014). This algorithm very efficiently detects recent gene flow events but not the historical admixture (Pritchard et al., 2000). The runs were executed with 20 iterations for a predefined number of population (K) set at 2.

The diversity parameters, such as number of segregating sites (S), Watterson's estimator (θ_w) per genotyped site (Watterson, 1975), Nei's (sometimes referred as Tajima's) nucleotide diversity (π) per genotyped site (Nei and Li, 1979; Tajima, 198), fixation index (F_{st}; Weir and Cockerham, 1984), as well as the frequency-based selection tests, such as Tajima's D (Tajima, 1989) and normalized Fay and Wu's H_{norm} (Fay and Wu, 2000; Zeng et al., 2006) were calculated separately for the wild and domesticated barley using mstatspop software with 1000 permutations (release 0.1b 20150803;

http://bioinformatics.cragenomica.es/numgenomics/people/sebas/software/software.html). This software performs the neutrality tests taking into account missing data using the models suggested by Ferretti et al. (2012). To determine ancestral status of the SNPs, which is a prerequisite for the H test, the SNPs were genotyped in two wild barley species, *H. bulbosum* and *H. pubiflorum*, and alleles that were identical in both species were tentatively assigned as ancestral. The genotyping was performed following the mapping and SNP calling pipeline described above using the *Hordeum* exome Illumina datasets (Mascher et al., 2013b).

The D and H_{norm} values vary greatly at different genomic regions due to the neutral random processes, e.g. genetic drift, and the range of this variation depends on the properties of the examined populations, such as the population size and demographic history. To estimate confidence intervals for the distribution of the D and H_{norm} under a Wright-Fisher neutral model in the wild and domesticated barley, coalescent simulations of 1000 datasets were performed using the ms software (Hudson, 2002) with the number of samples (n) and θ_w used as the variable parameters describing the populations. Variation of the D and H_{norm} in the simulated neutral datasets was assessed using the msstats and statsPs software (https://github.com/molpopgen/msstats).

The selective sweeps and selection signatures in individual loci were discovered using a combination of methods, namely the diversity reduction index ($\pi_{w(ild)}/\pi_{d(omesticated)}$), the average r^2 and the D and H_{norm} tests. The scans were performed in the wild and domesticated barley sub-sets genome-wide in the 10-cM windows with a sliding step of 1 cM and separately for the individual loci (cut-off > 5 SNPs). The sweeps and targets of selection were statistically defined based on the z-score test for outliers (p-value < 0.05) for the π_w/π_d and r^2 tests and on the simulated thresholds of neutral variation for the D and H_{norm} tests (p-value < 0.01). The overlapping outlier windows were merged into putatively selected regions.

The described analyses and visualization of the results were implemented in a series of custom scripts, consisting of a combination of bash, perl, awk and R programming languages and the standard Linux tools.

5 Tables

Table 1. Characteristics of the enrichment assay and SNP calling.

	Selected size, Mbp	**Captured, Mbp**	**Captured CDS, Mbp**	**Homozygous SNPs**		
				Total, with singletons	Total, w/o singletons	Filtered set*
Target	2.42	2.24	0.85	121,294	83,752	20,954
Non-target	-	11.56	0.48	423,024	270,858	34,682
Total	9.91	13.80	1.33	544,318	354,610	55,636

* - minor allele frequency < 0.05; missing data frequency < 0.5

Table 2. Nucleotide diversity parameters for wild and domesticated barley.

	Type	**S**	**Θ_w (x10^{-3}) ***	**π_n (x10^{-3}) ****	**D*****	**H_{norm}********
Wild	CDS	31,563	4.51	1.93	-1.804	-0.127
	nonCDS	265,990	7.95	3.18	-1.919	0.001
	all	297,553	7.36	2.97	-1.908	-0.050
Domesticated	CDS	5,125	1.01	1.05	0.165	-1.231
	nonCDS	37,750	1.56	1.62	0.145	-0.856
	all	42,875	1.47	1.53	0.147	-1.018

* - per genotyped nucleotide; ** - Nei's π_n; *** - Tajima’s D; **** - Fay&Wu's H_{norm}

Table 3. Selection signatures based on the sliding window genome scans.

ID*	Chromosome	Start, cM	End, cM	Size, cM	Total number of genes**	
					HighConfidence	LowConfidence
FWH_feat1_1	1	60	72	12	177	265
FWH_feat1_2	2	30	40	10	60	76
FWH_feat2_2	2	58	73	15	1200	2294
FWH_feat1_3	3	37	51	14	521	820
FWH_feat2_3	3	95	116	21	209	260
FWH_feat1_5	5	127	146	19	417	967
FWH_feat1_6	6	96	116	13	161	294
FWH_feat1_7	7	61	92	31	1718	3094
NPR_feat1_1	1	22	32	10	44	154
NPR_feat2_1	1	73	86	13	104	90
NPR_feat3_1	1	118	130	12	160	214
NPR_feat1_3	3	37	51	14	521	820
NPR_feat2_3	3	95	118	23	258	306
NPR_feat3_3	3	145	159	14	137	297
NPR_feat1_4	4	2	13	11	26	47
NPR_feat2_4	4	32	43	11	64	66
NPR_feat3_4	4	79	91	12	149	209
NPR_feat1_5	5	68	79	11	111	144
NPR_feat1_7	7	75	97	22	534	806
dLD_feat1_1	1	67	82	15	126	126
dLD_feat2_1	1	107	117	10	46	109
dLD_feat1_2	2	78	88	10	125	147
dLD_feat1_3	3	64	74	10	193	215
dLD_feat2_3	3	91	116	25	228	289
dLD_feat1_4	4	1	11	10	66	92
dLD_feat2_4	4	27	43	16	86	111
dLD_feat1_5	5	100	110	10	62	133

* - feat(ure) – putatively selected region; FWH – Fay&Wu's H_{norm}, NPR – π_w/π_d, dLD and wLD – LD outlier regions in domesticated and wild barley, respectively; ** - High and l confidence genes as defined in IBGSC, 2011

CHAPTER 2: 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

1 Introduction

Many plants use seasonal cues, such as photoperiod or vernalization, to coincide the timing of reproductive development with optimal climate conditions. Cultivated barley (*H. vulgare* L. subsp. *vulgare*), like most temperate cereal crops, is a long day (LD) plant with two growth types, winter and spring. Winter types accelerate flowering after a prolonged period of cold (vernalization), whereas spring barley does not respond to vernalization. The growth habit is determined by the interaction of two genes, *Vrn-H2*, a strong inhibitor of flowering under long day conditions and *Vrn-H1* (also known as *HvVRN1*). *Vrn-H1* is upregulated during vernalization and represses *Vrn-H2* (Yan *et al.*, 2003; 2004). A deletion of the *Vrn-H2* locus and deletions in a regulatory region of *Vrn-H1* cause a spring growth habit (Hemming *et al.*, 2009, Rollins *et al.*, 2013). In spring or vernalized winter barley, LDs strongly promote flowering, whereas short days (SD) delay reproductive development. Flowering under LDs is controlled by the major photoperiod response gene *Ppd-H1* (Turner *et al.*, 2005). *Ppd-H1* is a homolog of the *PSEUDO-RESPONSE REGULATOR* (*PRR*) genes implicated in the circadian clock of the model species *Arabidopsis thaliana* (L.) Heynh. (hereafter Arabidopsis). A natural mutation in the conserved CCT domain of *Ppd-H1* causes a reduced response to LDs and was selected in cultivation areas with long growing seasons (Turner *et al.*, 2005; von Korff *et al.*, 2006, 2010; Jones *et al.*, 2008; Wang *et al.*, 2010). *Ppd-H1*, *Vrn-H1* and *Vrn-H2* converge on the floral inducer *HvFT1* (*Vrn-H3*), a homolog of Arabidopsis florigen *FLOWERING LOCUS T* (*FT*). In barley, expression levels of *HvFT1* in the leaf correlate with flowering time. *Vrn-H2* represses *HvFT1* to counteract induction of *HvFT1* by *Ppd-H1* in LDs before vernalization (Hemming *et al.*, 2008). After vernalization, *Ppd-H1* becomes dominant and controls *HvFT1* expression and flowering time under LDs (Turner *et al.*, 2005; Campoli *et al.*, 2012).

In addition to the vernalization and photoperiod response genes, reproductive development is controlled by the *early maturity* (*eam*; also referred to as *earliness per se*) loci. The environmental effect on flowering phenotypes controlled by variation at these genes is reduced or completely removed. Two barley *eam* genes, *HvELF3* and *HvLUX1*, have recently been identified as homologs of the Arabidopsis circadian clock regulators *EARLY FLOWERING 3* (*ELF3*), and

LUX/ARRHYTHMO (*LUX*), respectively (Faure *et al.*, 2012; Zakhrabekova *et al.*, 2012; see Chapter 2). Mutations in these genes were linked to reduced photoperiod response and early flowering under both LD and non-inductive SD conditions.

The circadian clock is an autonomous oscillator that produces endogenous biological rhythms with a period of about 24 hours. Conceptually, a circadian system can be divided into three parts: the central oscillator, input and output pathways. In Arabidopsis, the circadian system comprises at least three interlocking feedback loops. The core oscillator is composed of three negative feedback loops: (a) the inhibition of evening complex (EC) genes *ELF3*, *EARLY FLOWERING 4* (*ELF4*), and *LUX* by the rise of CIRCADIAN CLOCK ASSOCIATED1 (CCA1) and LATE ELONGATED HYPOCOTYL (LHY) late at night, (b) the inhibition of *PRR* genes by the EC early at night, and (c) the inhibition of *LHY*/*CCA1* by TIMING OF CAB EXPRESSION1 (TOC1) in the morning (Kolmos *et al.*, 2009; Huang *et al.*, 2012; Pokhilko *et al.*, 2012). Furthermore, the evening-expressed GIGANTEA (GI) protein was proposed as a negative regulator of the EC, which in turn inhibits *TOC1* expression (Herrero *et al.*, 2012; Pokhilko *et al.*, 2012).

Light provides the circadian clock with diurnal entrainment signals. In plants, light is perceived and transduced by multiple photoreceptors including phytochromes, cryptochromes and phototropins (Davis, 2002). The phytochromes (PHY), which are apoproteins covalently bound to the chromophore, primarily detect and interpret the levels and ratio of red and far-red light in the environment. In darkness, phytochrome is synthesized as the physiologically inactive red-absorbing form, Pr. Upon illumination with red light, Pr is converted to the active far red absorbing form, Pfr, which can be transformed back to Pr by far red light. In the dark, Pfr-active phytochrome reverts to the inactive Pr form in a process referred to as “dark reversion” (Butler and Lane, 1965). These inter-convertible forms provide the plant with a cellular switch that can interpret information on spectral quantity and quality into a suitable response.

The diversity of phytochromes is organized in three major clades of *PHYA*, *PHYB* and *PHYC*, with the former two present in all studied seed plant taxa (Mathews, 2010). Plant phytochromes have been intensively studied, since they contribute to various developmental processes in plants such as flowering, shade avoidance, dormancy, germination and stomatal development (Franklin and Quail, 2010). Mutations in phytochrome genes affect flowering time in Arabidopsis, sorghum and rice (Childs *et al.*, 1997; Takano *et al.*, 2005; Balasubramanian *et al.*, 2006; Saidou *et al.*, 2009; Osugi *et al.*, 2011; Hu *et al.*, 2013). However, very little is known about the diversity, function and signaling pathways of barley phytochromes. Szücs *et al.* (2006) mapped

the barley orthologs of *PHYA* (*HvPHYA*) and *B* (*HvPHYB*) both to the short arm of chromosome 4H, and *PHYC* (*HvPHYC*) to the long arm of chromosome 5H at the same location as the vernalization gene Vrn-H1. According to Hanumappa *et al.* (1999), a chemically mutagenized barley genotype BMDR-1 contains a light-labile phyB. They demonstrated that it was responsible for the photoperiod insensitivity of this genotype and additionally implicated phyA in regulation of flowering via a distinct but interrelated pathway. A recent study in barley has suggested that variation at HvPHYC, affected flowering time only under LDs and independently of the circadian clock (Nishida *et al.*, 2013).

I identified the same candidate mutation in HvPHYC, which underlies the *early maturity 5* (*eam5)* locus, using whole-exome capture and a mapping-by-sequencing approach (Schneeberger *et al.*, 2009; Mascher *et al.*, 2013a) applied on a backcross population between the spring barley Bowman and the introgression line Bowman(*eam5*) (Druka *et al.*, 2010). This study adds important information to the recent findings of Nishida *et al.* (2013) by demonstrating that *HvPHYC-eam5* genetically interacts with the major photoperiod response gene *Ppd-H1* to accelerate flowering under LDs and in particular under SDs. In contrast to findings by Nishida *et al.* (2013), expression analyses showed that HvPHYC-eam5 disrupts the circadian clock and acts in the same pathway as the evening complex genes *HvELF3* and *HvLUX1 (*data not shown; courtesy of Chiara Campoli). Diversity analysis indicated the presence of two major HvPHYC haplotypes separated by a synonymous SNP and reduced nucleotide diversity at this locus. Interestingly, the HvPHYC-*eam5* allele was selected in barley cultivars from Japan despite of the strong effect of this mutation on the barley clock. This invites further research on comparing physiological effects and the overall significance of the circadian clock on plant adaptation.

2 Results

2.1 Bowman(*eam5)* is early flowering under short- and long-day conditions

Bowman, Bowman(*eam5*), Bowman(*Ppd-H1*) and Bowman(*Ppd-H1* + *eam5*) plants were scored for flowering time under LD and SD conditions in controlled greenhouse settings (Fig. 1AB). Under LDs, Bowman(*Ppd-H1*) and Bowman(*Ppd-H1* + *eam5*) flowered first, both at 28 days after sowing (DAS). Bowman(*eam5*) flowered at 42 DAS, followed by Bowman at 45 DAS. Under SDs, Bowman(*Ppd-H1* + *eam5*) flowered 57 DAS, followed by Bowman(*eam5*) with 65 DAS, while Bowman and Bowman(*Ppd-H1*) flowered on average 88 and 91 DAS, respectively. Thus *eam5* accelerated flowering time under both LD and SD conditions. It is notable that, under SDs, *Ppd-H1* accelerated flowering time in the background of *eam5*.

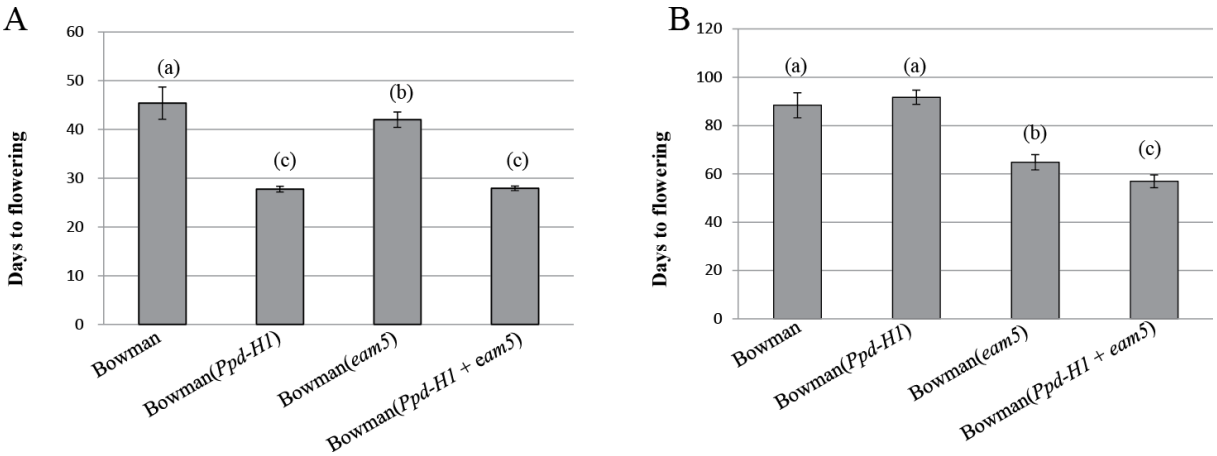


Figure 1 Flowering time of wild type and introgression lines.
Bowman and Bowman introgression lines carrying different combination of *Ppd-H1* and *eam5* alleles were grown under long days (16-h light/8-h darkness; LD, panel A) or short days (10-h light/14-h darkness, panel B). (A-B) Flowering time is shown as mean days to awn emergence on the main stem. Error bars are standard deviations for 8 to 15 plants of each genotype.

2.2 Identification of barley *PHYC* as a candidate gene underlying *eam5*

The *eam5* locus was described as a mutation of unknown origin isolated from an ICARDA/CIMMYT selection CMB85-533 (Higuerilla*2/Gobernadora) and mapped onto chromosome 5H (Jain, 1961; Franckowiak, 2002). Like many other barley QTLs, e*am5* has been introgressed into the spring barley cultivar Bowman (Druka *et al.*, 2011). The resultant BC6 introgression line Bowman(*eam5*) (also referred to as BW285) was genotyped using the single nucleotide polymorphism (SNP)-based array (Illumina's Barley Oligo Pool Array, BOPA) with 3072 SNPs (Close *et al.*, 2009). This revealed a single introgression on chromosome 5H of

Bowman(*eam5*) flanked by the BOPA markers 2_0533 (9.3 cM as morex_contig_64122) and 1_0336 (149.8 cM as morex_contig_2550061). *eam5* was mapped between RFLP markers MWG522 (80.3 cM as morex_contig_2549712) and MWG583 (89.9 cM as morex_contig_1583223) (Fig. 2A). To delineate the candidate gene underlying the *eam5* mutation, the introgression line Bowman(*eam5*) was backcrossed with Bowman and heading date was scored in the field. Of the 846 phenotyped BC_1F_2 lines, 204 genotypes were selected as flowering at the same time as Bowman(*eam5*), which flowered on average 4 days earlier than Bowman and the remaining population.

To refine the *eam5* interval, I used a mapping-by-sequencing approach applied on the parental lines Bowman and Bowman(*eam5*) and the pool of BC_1F_2 lines enriched for early flowering genotypes. The complexity of barley genome was reduced using whole-exome capture (Mascher *et al.*, 2013b), which targeted sequencing on the gene space. On average 92 % of the Illumina reads aligned against the targeted regions and were used to discover SNPs between the samples and the reference sequence. I identified 3,884 SNPs that were specific for either Bowman or Bowman(*eam5*). Out of those, 2,929 SNPs resided in 640 contigs located on the barley physical map (IBGSC, 2012). Using the median-allele frequency of SNPs within each contig and their physical location, SHOREmap, applied on the pool of BC_1F_2, identified a probabilistic QTL mapping interval of 8 Mb located on the chromosome arm 5HL and comprising 210 genes according to the most current POPSEQ barley map (Fig. 2B, Mascher *et al.*, 2013a). The maximum SNP allele frequencies of Bowman(*eam5*) approached 75% due to the presence of heterozygous genotypes in the pool of BC_1F_2 lines selected from the field experiment. The presence of heterozygotes in the pools was verified by phenotyping and genotyping in $BC_1F_{2:3}$ lines derived from the pooled plants as explained below.

The GO analysis identified five candidate genes related to flowering time or circadian clock within this 8-Mb interval. Four of them, *HvPHYC* (MLOC_824), *VRN-H1* (AK360697), *HvVIP4.1* (MLOC_17672), and *HvVIP4.2* (MLOC_17943), which are, respectively, homologs of Arabidopsis genes *PHYC* (AT5G35840), *APETALA 1* (AT1G69120) and *VERNALIZATION INDEPENDENCE 4* (AT5G61150), mapped to the same location on the chromosome arm 5HL. The fifth gene *HvCK2α* (MLOC_55943), a homologue of Arabidopsis *CASEIN KINASE 2 ALPHA* (AT2G23070), resided 2 Mb downstream of this cluster of four genes. Another flowering-related gene *HvPIF*-like (AK362162), a homologue of Arabidopsis *PHYTOCHROME-INTERACTING FACTOR* genes, located 9 Mb upstream of the unresolved cluster of the four candidate genes was also selected for the fine-mapping.

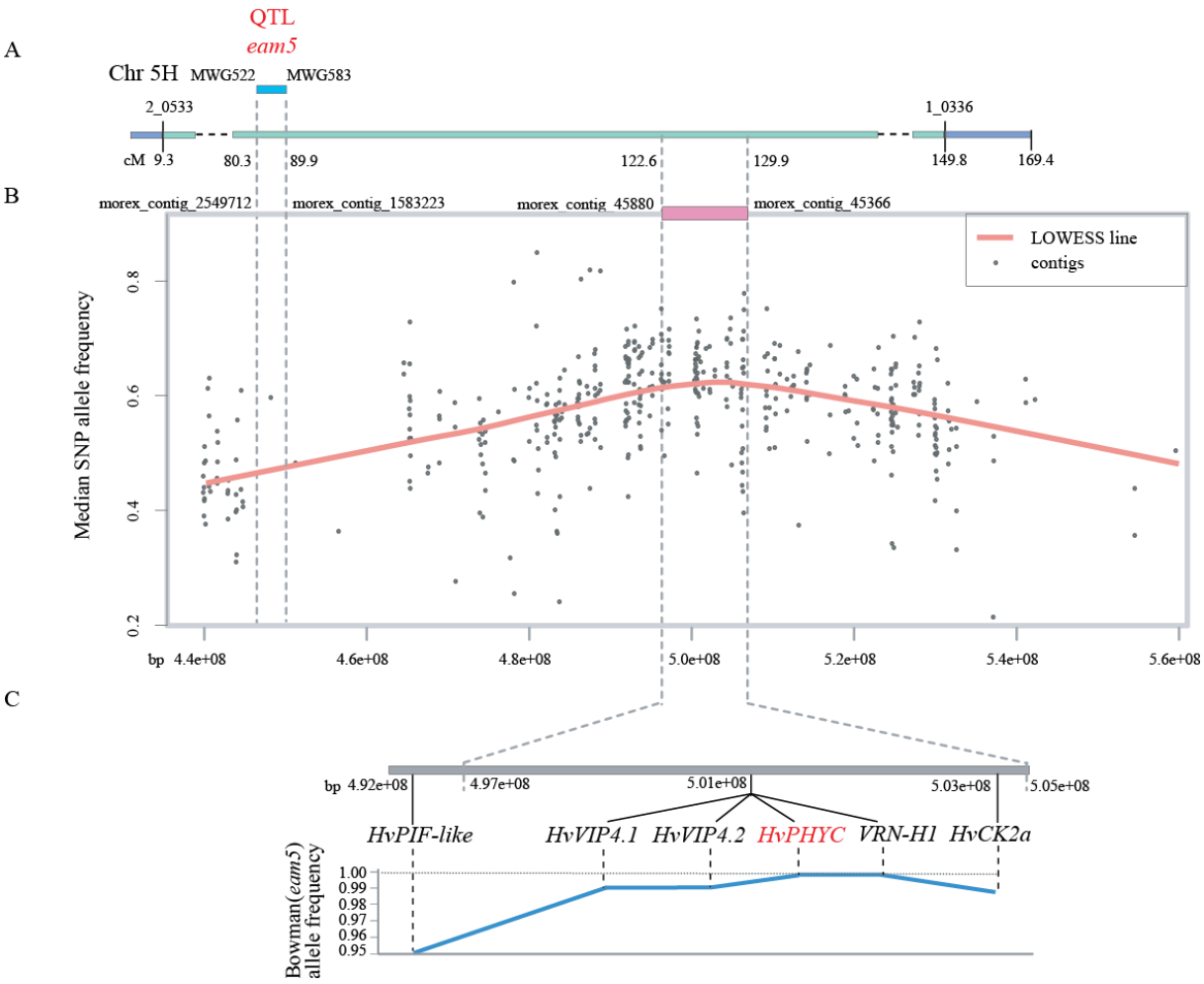


Figure 2 *Identification of* HvPHYC *as a candidate gene for the eam5 QTL.*

Corresponding regions of the genetic and physical maps are connected with grey dashed lines.

(A) Location of the *eam5* QTL on the Bowman(*eam5*) introgression. A 144-cM introgression on chromosome 5 of the Bowman(*eam5*) line and Bowman background are shown respectively as green and blue bars. The BOPA markers flanking the introgression and the corresponding genetic distances (cM) are above and below the bars, respectively. The tentative location of the *eam5* QTL is shown as a blue rectangle with genetic markers flanking the QTL.

(B) QTL position refinement using bulk segregant high-throughput sequencing and SHOREmap (Schneeberger *et al.*, 2009). Median SNP allele frequencies are plotted along the physical map of barley chromosome 5H in the vicinity of the *eam5* QTL. The QTL interval as calculated by SHOREmap is shown as a magenta bar together with the flanking contigs. Numbers on the x-axis are distances on barley physical map (IBGSC, 2012) in pairs of nucleotides. LOWESS, locally weighted scatterplot smoothing.

(C) Segregation analysis of the candidate genes within the SHOREmap QTL region. Six flowering-related candidate genes are located in or near the SHOREmap QTL region (grey bar). Frequencies of the Bowman(*eam5*) alleles in BC1F2:3 plants that flowered, excluding heterozygous genotypes, are plotted along the barley physical map (IBGSC, 2012). The highest allele frequency indicates the best candidate genes for the *eam5* QTL.

Segregation analysis of the candidate genes was conducted in $BC_1F_{2:3}$ lines, derived from the early flowering BC_1F_2 plants. Analysis of flowering time in 180 $BC_1F_{2:3}$ under SD conditions in the glasshouse revealed that the phenotype followed a trimodal distribution. A total of 108 lines were scored as early, which flowered within 10 days after Bowman(*eam5*), whereas 17 lines flowered significantly later frequently showing abnormalities in the spike development. The rest of 55 lines did not flower until the end of the experiment. This suggested that *eam5* was a semi-dominant locus with the heterozygote exhibiting an intermediate phenotype.

I reconstructed Bowman and Bowman(*eam5*) alleles of the candidate genes from the *de novo* assembly of the exome reads. Allele-specific polymorphisms were tagged with co-dominant sequence characterized amplified region (SCAR) and cleaved amplified polymorphic sequence (CAPS) markers used for screening of the $BC_1F_{2:3}$ lines. The segregation analysis revealed that both *Vrn-H1* and *HvPHYC* were tightly linked to the early flowering phenotype; all plants carrying Bowman(*eam5*) alleles at these genes flowered, plants with Bowman alleles did not flower until the end of the experiment, and heterozygotes exhibited an intermediate phenotype. No recombinants were detected between *Vrn-H1* and *HvPHYC* (Fig. 2C). *Vrn-H1* and *HvPHYC* reside only five gene models apart based on the better resolved Brachypodium map. Using allele-specific markers reported by Hemming *et al.* (2009), I discovered that Bowman carries the *HvVRN1-1* spring allele with a 5,154 bp deletion in the first intron, whereas Bowman(*eam5*) contains the *HvVRN1* winter allele without an intron deletion. This winter allele is known to strongly delay flowering in the absence of vernalization (Hemming *et al.*, 2009).

HvPHYC carried a non-synonymous mutation (T/C) in Bowman(*eam5*). This caused the missense substitution that leads to a change of the hydrophobic phenylalanine to the hydrophilic serine (mutation F380S) within a previously uncharacterized extremely conserved motif in the GAF domain of phytochromes (Fig. 3). At this position, phenylalanine is exclusively present in 4267 phytochrome homologs from 2799 species of Plantae and Bacteria kingdoms present in GenBank. The conservation analysis using the PROVEAN tool predicted that the F380S mutation could be functional (observed score -7.673; cut-off score -2.5). Taken together, our analyses strongly suggested that the F380S substitution could be critical for the HvPHYC function and thus I propose *HvPHYC* as the candidate gene underlying *eam5*.

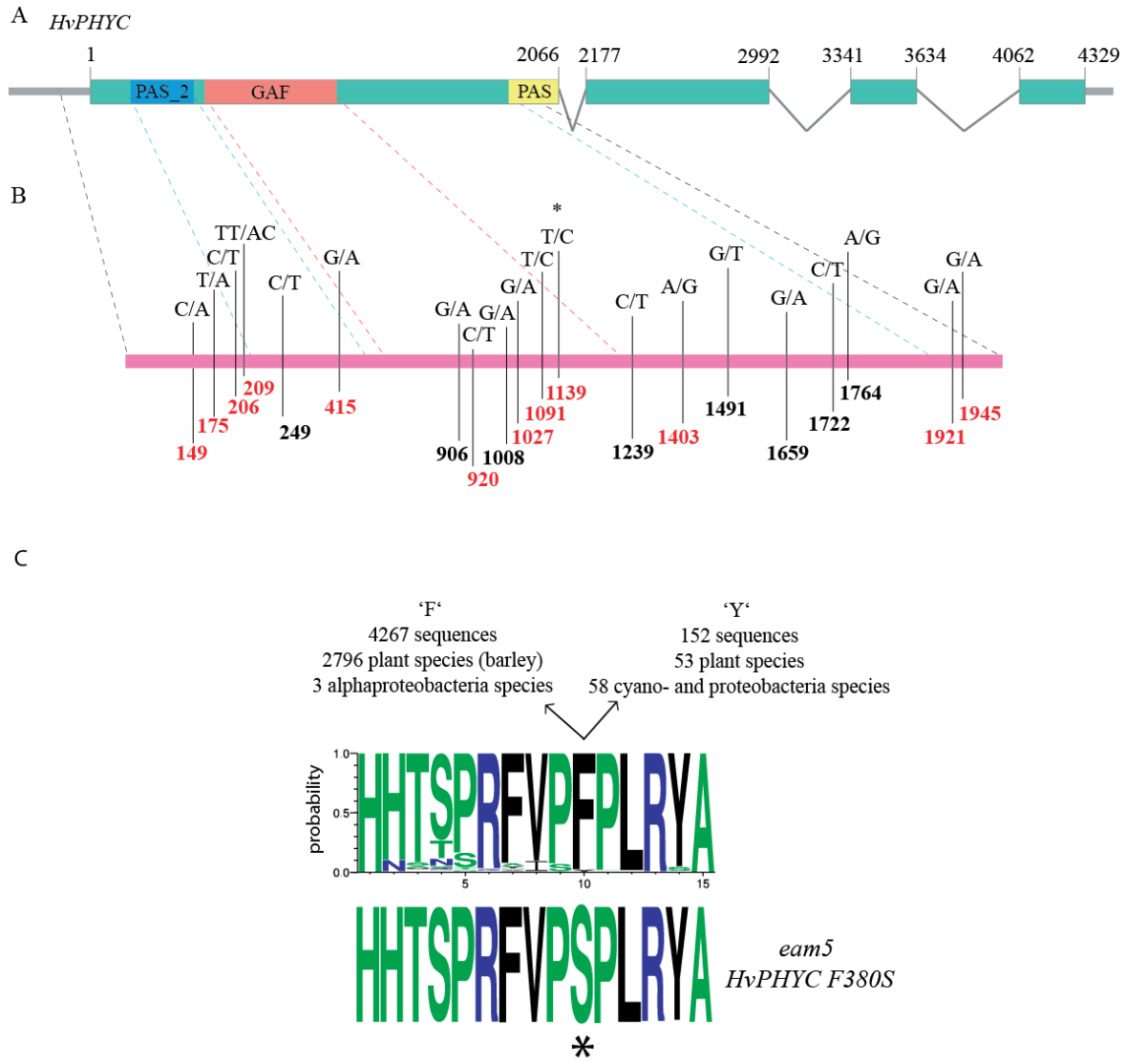


Figure 3 The structure of the barley (Hordeum vulgare) HvPHYC gene, polymorphisms in the exon 1 and effect of the eam5 mutation.

(A) The diagrammatic structure of the *HvPHYC* gene and location of the conserved domains on the exon 1. The exons and introns are respectively shown by green rectangles and lines angled downwards. The coordinates are shown in pairs of nucleotides relative to the first coding nucleotide of the full-length 'Morex' *HvPHYC* sequence (DQ238106). The conserved domains PAS_2 (IPR013654), PAS (IPR000014) and GAF (IPR03018) are highlighted in blue, yellow and red, respectively.

(B) Nucleotide variation in a 2045-bp fragment of the *HvPHYC* exon 1. The re-sequenced fragment comprising exon 1 of *HvPHYC* is shown as a magenta bar. SNPs are shown above the bar (major/minor alleles) and the coordinates, which are also a unique SNP IDs, below the bar. Non-synonymous SNPs are highlighted in red font. Correspondence of the polymorphic sites to the conserved domains is depicted by the dashed lines.

(C) Conservation of the motif in the GAF domain of the phytochrome gene family and the effect of the *eam5* mutation. An alignment of 4419 sequences of the extremely conserved motif in the GAF domain from the plant and bacteria *PHY* genes is shown as a sequence logo. Green letters, neutral amino-acid residues (a.a.); black, hydrophobic a.a.; blue, hydrophilic a.a. The a.a. substitution in the *eam5* line is shown by an asterisk.

2.3 Diversity analysis of *HvPHYC* and its linkage with *VRN-H1* alleles

To explore natural diversity of barley *HvPHYC* alleles, I sequenced a 2045-bp fragment of the first exon comprising conserved domains from 52 wild (*H. vulgare* subsp. *spontaneum* (K. Koch) Thell.), 56 cultivar and landrace, and 3 *H. agriocrithon* A.E. Åberg genotypes. In addition, three *HvPHYC* sequences from cultivars were extracted from NCBI Genbank. I identified 15 haplotypes, out of which 6 haplotypes were specific for cultivated accessions, 6 haplotypes for wild and 3 haplotypes were common to both groups (Fig. 4; Table S1). The nucleotide diversity of *HvPHYC* haplotypes, which were defined by 13 non-synonymous (ns) and 8 synonymous (s) SNPs within the coding region, was low ($\pi = 0.46 \times 10^{-3}$). Haplotypes 1 and 2 were most frequent; 84 out of 114 genotypes carried these two haplotypes.

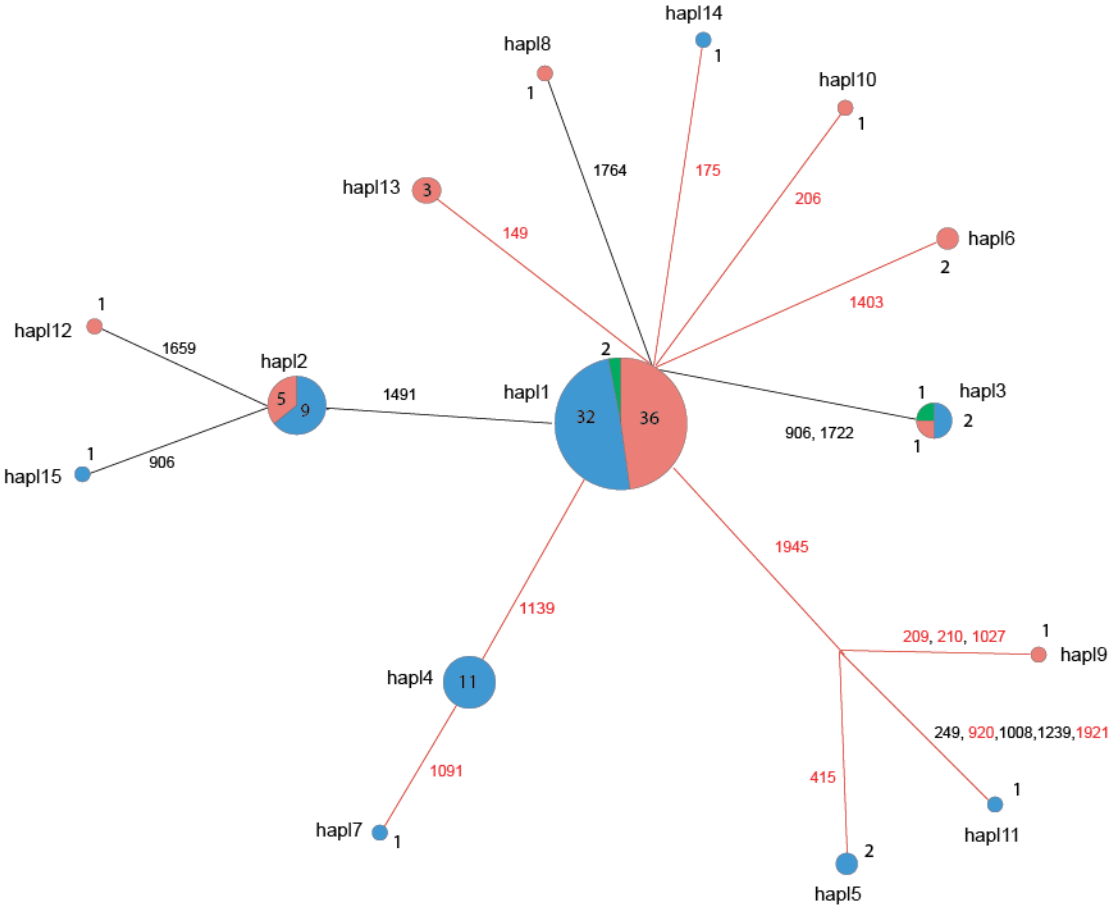


Figure 4 *Median-joining network of 15 HvPHYC haplotypes.*
Numbers at the nodes indicate the number of genotypes carrying the corresponding haplotype (out of 114 accessions). The haplotype frequency is also reflected in the relative size of a node. The color of a node corresponds to the different species: blue, *Hordeum vulgare* subsp. v*ulgare*; red, *H. vulgare* subsp. *spontaneum*; green, *H. agriocrithon*. Numbers at the branches denote synonymous (in black) and non-synonymous (in red) SNPs (as on Figure 3) separating the haplotypes.

I attempted to predict a functional effect of the observed non-synonymous SNPs based on the protein conservation patterns. The motif conservation analysis revealed two extremely conserved amino-acid substitutions, which were additionally identified by PROVEAN as deleterious (Table S2). One of the mutations, F380S, which was found in Bowman(*eam5*), also appeared in ten Japanese cultivars (haplotype 4; Table S3), whereas another mutation, L364D was found in addition to F380S in Japanese cultivar Azumamugi (haplotype 7).

Based on the result of segregation analysis, I assumed that the Bowman(*eam5*) *HvPHYC* allele and wild type Hv*VRN1* are tightly linked. To verify this fact, I screened 10 other 'haplotype 4' genotypes with the markers specific for the wild type *HvVNR1* allele. Without exception, the Bowman(*eam5*) *HvPHYC* allele was linked to the winter type *HvVRN1* allele (Table S3).

3 Discussion

In this chapter, I describe the barley locus *eam5*, which accelerated flowering under LDs and in addition led to flowering under non-inductive SDs. In order to fine-map the *eam5* mutation, I used mapping-by-sequencing of bulked early flowering BC_1F_2 lines, followed by candidate-gene mapping in $BC_1F_{2:3}$. Fine-mapping by deep sequencing has been successfully applied in model species to map and identify induced mutations underlying a specific phenotype in a single step (James *et al.*, 2013). I demonstrate that this method is also effective in fine-mapping a mutation in the large genome of a crop plant. I found that the use of an introgression line with prior mapping information and exome-enrichment greatly reduced the complexity of the task (Druka *et al.*, 2011, IBGSC, 2012, Mascher *et al.*, 2013b). Mascher *et al.* (2013b) have demonstrated that fine-mapping of barley genes has become feasible due to the recent release of the barley gene space reference sequence and advances in the physical and genetic mapping (IBGSC, 2012; Mascher *et al.*, 2013a). As a proof of concept, using a simulated *in silico* bulk-segregant analysis, they showed that a qualitative row-type gene *Vrs1* could be fine-mapped to a relatively small interval containing 128 genes (Mascher *et al.*, 2013b). I demonstrate that fine-mapping through exome capture and deep sequencing of a BC_1F_2 pool was successful, even though the phenotype was quantitative, subtle and obscured by the segregation of another tightly linked flowering gene. The identification of a mutation in the extremely conserved motif of the PHY GAF domain strongly suggested that *HvPHYC* is the gene underlying the *eam5* locus. A mutation in this region might not be expected to alter the photochromic behavior of the protein. Future work could examine whether the reversion rates from Pfr to Pr are compromised in the HvPHYC F380S isoform.

This study is an example of the efficient exploitation of the comprehensive barley mutant resources. More than 800 natural and induced barley mutants introgressed into the common genetic background Bowman await fine-mapping and functional analysis of the causative genes (Druka *et al.*, 2010).

The F380S mutation did not correspond to any known loss- or gain-of-function phytochrome alleles in Arabidopsis (Nagatani *et al.*, 2010). However, the mutation was identical to the HvPHYC-e allele recently described in the Japanese winter barley Hayakiso 2 (Nishida *et al.*, 2013). The authors of that study found that this allele accelerated flowering time under 16-h and 20-h LDs, but not under 12-h SDs. However, I observed the strongest effect of *eam5* under 8-h SDs, where flowering time of Bowman(*eam5*) and Bowman differed by more than 60 days. It is noteworthy that the loss-of-function *PHYC* mutants in Arabidopsis and rice also showed acceleration of flowering in non-inductive photoperiods (Monte *et al.*, 2003; Takano *et al.*, 2005;

Balasubramanian *et al.*, 2006). Changes in clock gene expression, especially the absence of *HvCCA1* oscillations in Bowman(*eam5*), were reminiscent of expression changes observed in the *eam8* and *eam10* mutants carrying mutations in *HvELF3* and *HvLUX1*, respectively (Faure *et al.*, 2012; see Chapter 2; expression data not shown, courtesy of Chiara Campoli).

In Arabidopsis, *ELF3* together with *ELF4* and *LUX* form the so-called evening complex (EC) and repress transcription of PRRs (Dixon *et al.*, 2011; Herrero *et al.*, 2012). The transcriptional targets of the EC genes seem conserved in barley. HvELF3 and HvLUX1 act as repressors of the barley PRR gene *Ppd-H1*, which in turn affects *HvFT1* expression and flowering time (Faure *et al.*, 2012). Indeed, I observed genetic interactions between *eam5* and *Ppd-H1* as demonstrated for *eam8* and *eam10* (Fig. 1).

To investigate genetic diversity at *HvPHYC*, I re-sequenced in addition to Bowman and Bowman(*eam5*) a diverse set of 109 *HvPHYC* genotypes and added 3 *HvPHYC* alleles from the GenBank. The prevalence of two major haplotypes and the nucleotide diversity index, which was 2 to 16 times lower than reported for other barley genes (Russell *et al.*, 2004; Morrell *et al.*, 2005, 2013; Kilian *et al.*, 2006; Xia *et al.*, 2012), indicated that PhyC was conserved and presumably under selective constraints. It is tempting to speculate that reduced nucleotide diversity at *HvPHYC* could be indirect effect of its tight linkage to *VRN-H1*, which mediates an adaptive trait such as vernalization response (Beales et al., 2005; Hemming *et al.*, 2009). It is intriguing that the mutant PhyC380 allele was detected in eleven genotypes with a common geographic origin in Japan, suggesting that this mutant allele was targeted by breeders.

Molecular taxonomists have widely used phytochromes in phylogenetic studies (Mathews *et al.*, 1995). This led to the accumulation of a large number of PHY sequences from very diverse plant and bacteria species. I used these data to infer a functional effect of the observed amino-acid substitutions based on the extent of their conservation. Remarkably, except for the two amino-acid substitutions, L364D and F380S (haplotypes 4 and 7), other substitutions were within variable motifs and thus presumably non-functional.

In summary, I successfully applied a mapping-by-sequencing approach to pinpoint a mutation in *HvPHYC* as a candidate underlying the *eam5* locus in barley. The mutation disrupts the circadian clock and results in an acceleration of flowering under LDs and non-inductive SDs. Such interaction of phytochromes and the circadian clock has not been reported before and opens new perspectives on the role of PHYC in controlling the circadian clock and downstream light signaling pathways. I demonstrate that *HvPHYC* is characterized by low levels of genetic diversity in wild and cultivated barley germplasm. Interestingly, HvPHYC F380S was selected in cultivars from Japan, and may thus provide a selective advantage in these environments.

4 Materials and Methods

4.1 Plant material and growth conditions

Flowering time (days to awn emergence on the main spike) of the spring cultivar Bowman and three Bowman-derived introgression lines Bowman(*Ppd-H1)*, Bowman(*eam5)* and Bowman(*Ppd-H1+eam5)* (kindly provided by R. Waugh, James Hutton Institute and by David Laurie, John Innes Centre and developed by Chiara Campoli) was recorded for 10-15 plants per genotype. To score flowering, plants were grown in soil in a glasshouse under SDs (10 h light, 20°C : 14 h dark, 18°C) and LDs (16 h light, 20°C : 8 h dark, 18°C). Significant differences in flowering time, meristem development, and gene expression between Bowman, Bowman(*eam5*), and Bowman(*eam5 + Ppd-H1*) were calculated using a paired t-test ($P < 0.05$).

To investigate natural diversity of the candidate gene, a set of 110 wild and cultivated barley genotypes were selected from germplasm collections at the Max Planck Institute of Plant Breeding Research (MPIPZ) in Cologne (Badr *et al.*, 2000), the Barley 1K collection (Hübner *et al.*, 2009), the Barley Germplasm Center at Okayama University (http://www.shigen.nig.ac.jp/barley) and Hakan Özkan (University of Çukurova, Turkey) (Table S3).

4.2 Identification of a candidate gene using mapping by sequencing and segregation analysis

To determine the position of *eam5* on genetic and physical maps, I found closest flanking RFLP markers with known nucleotide sequences using the GrainGenes CMap browser (BinMap, 2005; http://wheat.pw.usda.gov/cgi-bin/cmap). BOPA markers flanking the Bowman(*eam5*) introgression were as determined by Druka *et al.* (2010). The RFLP and BOPA markers were anchored to the Morex genomic contigs (IBGSC, 2012) using blastn. Genetic and physical locations of the Morex contigs were extracted from the Mascher *et al.* (2013a) and IBGSC (2012) data, respectively. To refine the candidate region carrying the causative mutation, a mapping-by-sequencing approach was applied on Bowman, Bowman(*eam5*) and a pool of 204 BC_1F_2 lines enriched for early-flowering genotypes. The BC_1F_2 lines flowering together or close to Bowman(*eam5*) were selected from 846 BC_1F_2 lines sown in March 2012 in the field at the MPIPZ and scored for heading date together with the parental lines. Genomic DNA was extracted from leaves using the BioSprint 96 kit (QIAGEN) according to the manufacturer's recommendations. DNA samples were quantified using Quant-iT™

PicoGreen assay (Invitrogen) with the Synergy™ 4 microplate reader (Biotek). DNA from 204 BC_1F_2 lines was pooled in equal amounts and along with Bowman, Bowman(*eam5*) enriched for a 61.6 mega base pair coding sequence target using in-solution whole-exome capture (Roche NimbleGen, Madison, WI, USA; Mascher *et al.*, 2013b).

Illumina sequencing of the enriched libraries generated 93M, 80M and 211M reads for the Bowman, Bowman(*eam5*) and the pool of BC_1F_2 lines, respectively. Reads were trimmed and aligned to the barley reference sequence (IBGSC, 2012) using BWA 0.59 with default parameters (Li and Durbin, 2009). Only those reads, which uniquely mapped to the reference sequence, were retained. Samtools 0.1.19 was used to generate consensus pileup information (Li *et al.*, 2009). SNPs distinguishing the two parental samples were extracted using VarScan 2.3.5 (Koboldt *et al.*, 2009). SNPs supported by at least 30 reads with the non-reference allele frequency more than 95% in either Bowman or Bowman(*eam5*) were considered in downstream analyses. In the BC1F2 pool data, I estimated the frequencies of the Bowman(*eam5*) SNP alleles. Next, the median allele frequency for all SNPs was estimated within individual reference contigs that were anchored to the barley physical map (IBGSC, 2013) and calculated a locally weighted scatterplot smoothing (LOWESS) regression. SHOREmap 2.1 (http://www.shoremap.org) was used to calculate a mapping interval based on the median allele frequencies and the corresponding coefficients of variation (Schneeberger *et al.*, 2009; Galvão *et al.*, 2012).

Barley genes within the candidate mapping interval were extracted using the map published by Mascher *et al.* (2013a) and characterized by the gene ontology (GO) terms using Blast2GO 2.5 pipeline (Conesa *et al.*, 2005). Genes related to flowering and circadian clock were selected as candidate genes. To narrow the list of candidate genes down, I designed allele-specific co-dominant markers (SCAR and CAPS) distinguishing candidate gene alleles from Bowman and Bowman(*eam5*) and analyzed their segregation with the flowering phenotype using $BC_1F_{2:3}$ lines. To extract allele-specific polymorphisms in the candidate genes, Bowman and Bowman(*eam5*) reads were assembled *de novo* into contigs using ABySS 1.3.7 assembler (single end, k=25; Simpson *et al.*, 2009). The contigs homologous to the candidate genes were identified using blastn and aligned to extract SNPs to design allele-specific PCR markers. Allele-specific markers for the barley *VIP4*-like genes, which were absent from the exome-enrichment assay, were designed based on the Sanger sequencing data of PCR fragments (primers see in Table S4).

To create the $BC_1F_{2:3}$ population, one seed of each of the early BC_1F_2 plants was sown in the greenhouse under 10-h short days. Flowering time was scored and leaf material was harvested for DNA extraction and genotyping. PCR reactions (1× Colorless GoTaq® Buffer, 0.2 µM dNTPs, 0.5

µM primers, 1 U GoTaq® polymerase (Promega, Mannheim, Germany), 50 ng DNA) were incubated in the PTC DNA Engine thermocycler (Bio-Rad, Hercules, CA, USA) and visualized using agarose gel electrophoresis. The primer sequences and incubation regimes were as in Table S4. Restriction of the CAPS PCR fragments was performed using endonucleases (New England Biolabs, Frankfurt am Main, Germany) following manufacturer's recommendations.

4.3 Natural diversity and population-genetic analyses

Barley genomic DNA was extracted using the DNAeasy 96 Plant kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendations and quantified using NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE). A 2045-bp fragment covering the exon 1 of *HvPHYC* was amplified from a set of 113 wild and cultivated barley genotypes using the primer pairs Ex1seq_1f + Ex1seq_1r and Ex1seq_2f + Ex1seq_2r (Table S4). PCR reactions (1 × Q5® buffer, 0.2 µM dNTPs, 0.5 µM primers, 1 U Q5® High-Fidelity DNA Polymerase (New England Biolabs, Frankfurt am Main, Germany), 50 ng DNA) were incubated in a PTC DNA Engine thermocycler (Bio-Rad, Hercules, CA, USA). PCR fragments were purified using 1,8x Agencourt AMPure XP beads (Beckman Coulter, Krefeld, Germany) following the manufacturer's recommendations and Sanger-sequenced. Three additional *HvPHYC* sequences were extracted from NCBI GenBank (DQ201145, DQ201146 and DQ238106). Haplotype analysis was performed as described in Chapter 2. Nucleotide diversity π was calculated for the coding region using DnaSAM v. 20100621 (Eckert *et al.*, 2010).

4.4 Motif conservation analysis

Positions of the GAF and PAS domains at HvPHYC were determined using InterProScan (Quevillon *et al.*, 2005). A set of 4419 protein sequences of the PHY homologues from plant and bacterial species were extracted using blastp search in the NCBI 'nr' database with a 70-bp conserved fragment of the HvPHY GAF domain as a query (cut-off e-value 1e-7). The PHY homologues were aligned using MAFFT v6.851b ('auto' model selection) and PRANK v.130820 (default parameters, +F) (Loytynoja and Goldman, 2005; Katoh and Toh, 2008). Amino-acid polymorphisms at HvPHYC were discovered by translating the exon 1 fragments comprising non-synonymous SNPs using ExPASy Translate tool (http://web.expasy.org/translate). The 8-15 aa sub-alignments around these polymorphic sites were submitted to the WebLogo generator (Crooks *et al.*,

2004). Visually misaligned regions flanking the polymorphic amino-acid residues outside the GAF domain were iteratively re-aligned in a smaller subset of the PHY sequences. In addition, the functional effect of the amino acid substitutions was predicted using PROVEAN (cut-off score -2.5) (Choi *et al.*, 2012).

LITERATURE CITED

Abbo, Shahal, Ruth Pinhasi van-Oss, Avi Gopher, Yehoshua Saranga, Itai Ofner, and Zvi Peleg. "Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes." *Trends in plant science* 19, no. 6 (2014): 351-360.

Abbott, D. C., A. H. D. Brown, and J. J. Burdon. "Genes for scald resistance from wild barley (Hordeum vulgare ssp spontaneum) and their linkage to isozyme markers." *Euphytica* 61, no. 3 (1991): 225-231.

Abdel-Ghani, Adel H., Heiko K. Parzies, Ayed Omary, and Hartwig H. Geiger. "Estimating the outcrossing rate of barley landraces and wild barley populations collected from ecologically different regions of Jordan." *Theoretical and Applied Genetics* 109, no. 3 (2004): 588-595.

Allaby, Robin G. "Barley domestication: the end of a central dogma?." *Genome biology* 16, no. 1 (2015).

Allaby, Robin G., and Terence A. Brown. "AFLP data and the origins of domesticated crops." *Genome* 46, no. 3 (2003): 448-453.

Allaby, Robin G., and Terence A. Brown. "Reply to the comment by Salamini et al. on" AFLP data and the origins of domesticated crops"." *Genome* 47, no. 3 (2004): 621-622.

Allaby, Robin G., Dorian Q. Fuller, and Terence A. Brown. "The genetic expectations of a protracted model for the origins of domesticated crops." *Proceedings of the National Academy of Sciences* 105, no. 37 (2008a): 13982-13986.

Allaby, Robin, Dorian Q. Fuller, and Terence Brown. "Reply to Ross-Ibarra and Gaut: Multiple domestications do appear monophyletic if an appropriate model is used." *Proceedings of the National Academy of Sciences of the United States of America* 105, no. 49 (2008b): E106.

Badr, Abdelfattah, R. Sch, H. El Rabey, S. Effgen, H. H. Ibrahim, C. Pozzi, W. Rohde, and F. Salamini. "On the origin and domestication history of barley (Hordeum vulgare)." *Molecular Biology and Evolution* 17, no. 4 (2000): 499-510.

Balasubramanian, Sureshkumar, Sridevi Sureshkumar, Mitesh Agrawal, Todd P. Michael, Carrie Wessinger, Julin N. Maloof, Richard Clark, Norman Warthmann, Joanne Chory, and Detlef Weigel. "The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of Arabidopsis thaliana." *Nature genetics* 38, no. 6 (2006): 711-715.

Barrett, Rowan DH, and Dolph Schluter. "Adaptation from standing genetic variation." *Trends in Ecology & Evolution* 23, no. 1 (2008): 38-44.

Beales, J., D. A. Laurie, and K. M. Devos. "Allelic variation at the linked AP1 and PhyC loci in hexaploid wheat is associated but not perfectly correlated with vernalization response." *Theoretical and applied genetics* 110, no. 6 (2005): 1099-1107.

Bodi, K., A. G. Perera, P. S. Adams, D. Bintzler, K. Dewar, D. S. Grove, J. Kieleczawa et al. “Comparison of commercially available target enrichment methods for next-generation sequencing.” *Journal of biomolecular techniques: JBT* 24, no. 2 (2013): 73.

Bothmer R von, Flink J, Jacoben N, Kotimäki M, Landström T: Interspecific hybridization with cultivated barley (*Hordeum vulgare* L.). Hereditas 99:219–244 (1983).

Brandt, Débora YC, Vitor RC Aguiar, Bárbara D. Bitarello, Kelly Nunes, Jérôme Goudet, and Diogo Meyer. “Mapping bias overestimates reference allele frequencies at the HLA genes in

the 1000 Genomes Project phase I data." *G3: Genes| Genomes| Genetics* 5, no. 5 (2015): 931-941.

Brown, A. H. D., E. Nevo, D. Zohary, and O. Dagan. "Genetic variation in natural populations of wild barley (Hordeum spontaneum)." *Genetica* 49, no. 2 (1978): 97-108.

Brown, Terence A., Martin K. Jones, Wayne Powell, and Robin G. Allaby. "The complex origins of domesticated crops in the Fertile Crescent." *Trends in Ecology & Evolution* 24, no. 2 (2009): 103-109.

Buckler, Edward S., James B. Holland, Peter J. Bradbury, Charlotte B. Acharya, Patrick J. Brown, Chris Browne, Elhan Ersoz et al. "The genetic architecture of maize flowering time." *Science* 325, no. 5941 (2009): 714-718.

Bus, Anja, Jochen Hecht, Bruno Huettel, Richard Reinhardt, and Benjamin Stich. "High-throughput polymorphism detection and genotyping in Brassica napus using next-generation RAD sequencing." *BMC genomics* 13, no. 1 (2012): 281.

Butler, W. L., and H. C. Lane. "Dark transformations of phytochrome in vivo. II." *Plant physiology* 40, no. 1 (1965): 13.

Caldwell, Katherine S., Joanne Russell, Peter Langridge, and Wayne Powell. "Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, Hordeum vulgare." *Genetics* 172, no. 1 (2006): 557-567.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., & Madden T.L. (2008) "BLAST+: architecture and applications." BMC Bioinformatics 10:421.

Campoli, Chiara, Benedikt Drosse, Iain Searle, George Coupland, and Maria von Korff. "Functional characterisation of HvCO1, the barley (Hordeum vulgare) flowering time ortholog of CONSTANS." *The Plant Journal* 69, no. 5 (2012a): 868-880.

Campoli, Chiara, Munqez Shtaya, Seth J. Davis, and Maria von Korff. "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, no. 1 (2012b): 97.

Campoli, Chiara, Artem Pankin, Benedikt Drosse, Cristina M. Casao, Seth J. Davis, and Maria Korff. "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, no. 4 (2013): 1045-1059.

Cavanagh, Colin R., Shiaoman Chao, Shichen Wang, Bevan Emma Huang, Stuart Stephen, Seifollah Kiani, Kerrie Forrest et al. "Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars." *Proceedings of the national academy of sciences* 110, no. 20 (2013): 8057-8062.

Ceccarelli, S., S. Grando, and J. A. G. Van Leur. "Genetic diversity in barley landraces from Syria and Jordan." *Euphytica* 36, no. 2 (1987): 389-405.

Chakravarti, A. (1999). Population genetics—making sense out of sequence.*Nature genetics*, *21*, 56-60.

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. "Second-generation PLINK: rising to the challenge of larger and richer datasets." *GigaScience* 4 (2014): 7-7.

Charlesworth, Brian, Deborah Charlesworth, and Nicholas H. Barton. "The effects of genetic and geographic structure on neutral variation." *Annual Review of Ecology, Evolution, and*

Systematics (2003): 99-125.

Charlesworth, Brian, M. T. Morgan, and D. Charlesworth. "The effect of deleterious mutations on neutral molecular variation." *Genetics* 134, no. 4 (1993): 1289-1303.

Chen, Andrew, Chengxia Li, Wei Hu, Mei Yee Lau, Huiqiong Lin, Nathan C. Rockwell, Shelley S. Martin, Judith A. Jernstedt, J. Clark Lagarias, and Jorge Dubcovsky. "PHYTOCHROME C plays a major role in the acceleration of wheat flowering under long-day photoperiod." *Proceedings of the National Academy of Sciences* 111, no. 28 (2014): 10037-10044.

Childs, Kevin L., Frederick R. Miller, Marie-Michele Cordonnier-Pratt, Lee H. Pratt, Page W. Morgan, and John E. Mullet. "The sorghum photoperiod sensitivity gene, Ma3, encodes a phytochrome B." *Plant Physiology* 113, no. 2 (1997): 611-619.

Chen, Haodong, Xi Huang, Giuliana Gusmaroli, William Terzaghi, On Sun Lau, Yuki Yanagawa, Yu Zhang et al. "Arabidopsis CULLIN4-damaged DNA binding protein 1 interacts with CONSTITUTIVELY PHOTOMORPHOGENIC1-SUPPRESSOR OF PHYA complexes to regulate photomorphogenesis and flowering time." *The Plant Cell* 22, no. 1 (2010a): 108-123.

Chen, Hua, Nick Patterson, and David Reich. "Population differentiation as a test for selective sweeps." *Genome research* 20, no. 3 (2010b): 393-402.

Chia, Jer-Ming, Chi Song, Peter J. Bradbury, Denise Costich, Natalia de Leon, John Doebley, Robert J. Elshire et al. "Maize HapMap2 identifies extant variation from a genome in flux." *Nature genetics* 44, no. 7 (2012): 803-807.Fuller, Dorian Q., George Willcox, and Robin G. Allaby. "Early agricultural pathways: moving outside the ‘core area’ hypothesis in Southwest Asia" *Journal of Experimental Botany* 63, no. 2 (2012): 617-633.

Choi, Yongwook, Gregory E. Sims, Sean Murphy, Jason R. Miller, and Agnes P. Chan. "Predicting the functional effect of amino acid substitutions and indels." (2012): e46688.

Cingolani, Pablo, Adrian Platts, Le Lily Wang, Melissa Coon, Tung Nguyen, Luan Wang, Susan J. Land, Xiangyi Lu, and Douglas M. Ruden. “A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3.” *Fly* 6, no. 2 (2012): 80-92.

Clark, Andrew G., Melissa J. Hubisz, Carlos D. Bustamante, Scott H. Williamson, and Rasmus Nielsen. “Ascertainment bias in studies of human genome-wide polymorphism.” *Genome research* 15, no. 11 (2005): 1496-1502.

Close, Timothy J., Prasanna R. Bhat, Stefano Lonardi, Yonghui Wu, Nils Rostoks, Luke Ramsay, Arnis Druka et al. “Development and implementation of high-throughput SNP genotyping in barley.” *BMC genomics* 10, no. 1 (2009): 582.

Cockram, James, Elena Chiapparino, Scott A. Taylor, Konstantina Stamati, Paolo Donini, David A. Laurie, and Donal M. O’Sullivan. "Haplotype analysis of vernalization loci in European barley germplasm reveals novel VRN-H1 alleles and a predominant winter VRN-H1/VRN-H2 multi-locus haplotype." *Theoretical and Applied Genetics* 115, no. 7 (2007): 993-1001.

Cockram, James, Huw Hones, and Donal M. O'Sullivan. "Genetic variation at flowering time loci in wild and cultivated barley." *Plant Genetic Resources* 9, no. 02 (2011): 264-267.

Comadran, Jordi, Benjamin Kilian, Joanne Russell, Luke Ramsay, Nils Stein, Martin Ganal, Paul Shaw et al. "Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley." *Nature genetics* 44, no. 12 (2012): 1388-1392.

Conesa, Ana, Stefan Götz, Juan Miguel García-Gómez, Javier Terol, Manuel Talón, and Montserrat Robles. "Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research." *Bioinformatics* 21, no. 18 (2005): 3674-3676.

Crooks, Gavin E., Gary Hon, John-Marc Chandonia, and Steven E. Brenner. "WebLogo: a sequence logo generator." *Genome research* 14, no. 6 (2004): 1188-1190.

Cutter, Asher D., and Bret A. Payseur. "Genomic signatures of selection at linked sites: unifying the disparity among species." *Nature Reviews Genetics*14, no. 4 (2013): 262-274.

Dai, Fei, Eviatar Nevo, Dezhi Wu, Jordi Comadran, Meixue Zhou, Long Qiu, Zhonghua Chen, Avigdor Beiles, Guoxiong Chen, and Guoping Zhang. "Tibet is one of the centers of domestication of cultivated barley." *Proceedings of the National Academy of Sciences* 109, no. 42 (2012): 16969-16973.

Darwin, Charles. "The Variation of animals and plants under domestication, 2d edition, 2 vols." (1868).

Davis, Seth J. "Photoperiodism: the coincidental perception of the season."*Current Biology* 12, no. 24 (2002): R841-R843.

Dixon, Laura E., Kirsten Knox, Laszlo Kozma-Bognar, Megan M. Southern, Alexandra Pokhilko, and Andrew J. Millar. "Temporal repression of core circadian genes is mediated through EARLY FLOWERING 3 in Arabidopsis."*Current Biology* 21, no. 2 (2011): 120-125.

Doebley, John F., Brandon S. Gaut, and Bruce D. Smith. "The molecular genetics of crop domestication." *Cell* 127, no. 7 (2006): 1309-1321.

Druka, Arnis, Jerome Franckowiak, Udda Lundqvist, Nicola Bonar, Jill Alexander, Kelly Houston, Slobodanka Radovic et al. "Genetic dissection of barley morphology and development." *Plant physiology* (2010): pp-110.

Dubcovsky, Jorge, Chialing Chen, and Liuling Yan. "Molecular characterization of the allelic variation at the VRN-H2 vernalization locus in barley." *Molecular Breeding* 15, no. 4 (2005): 395-407.

Duran, Chris, Nikki Appleby, Megan Vardy, Michael Imelfort, David Edwards, and Jacqueline Batley. “Single nucleotide polymorphism discovery in barley using autoSNPdb.” *Plant biotechnology journal* 7, no. 4 (2009): 326-333.

Eberle, Michael A., Mark J. Rieder, Leonid Kruglyak, and Deborah A. Nickerson. "Allele frequency matching between SNPs reveals an excess of linkage disequilibrium in genic regions of the human genome." *PLoS Genet* 2, no. 9 (2006): e142.

Eckert, Andrew J., John D. Liechty, Brandon R. Tearse, Barnaly Pande, and David B. Neale. "DnaSAM: Software to perform neutrality testing for large datasets with complex null models." *Molecular ecology resources* 10, no. 3 (2010): 542-545.

Eglinton, J. K., D. E. Evans, A. H. D. Brown, P. Langridge, G. McDonald, S. P. Jefferies, and A. R. Barr. "The use of wild barley (Hordeum vulgare ssp. spontaneum) in breeding for quality and adaptation." In *9th Australian barley technical symposium, Melbourne*. 1999.

Eyre-Walker, Adam, Rebecca L. Gaut, Holly Hilton, Dawn L. Feldman, and Brandon S. Gaut. "Investigation of the bottleneck leading to the domestication of maize." *Proceedings of the National Academy of Sciences* 95, no. 8 (1998): 4441-4446.

Faure, Sebastien, Adrian S. Turner, Damian Gruszka, Vangelis Christodoulou, Seth J. Davis, Maria von Korff, and David A. Laurie. "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*109, no. 21 (2012): 8328-8333.

Fay, Justin C., and Chung-I. Wu. "Hitchhiking under positive Darwinian selection." *Genetics* 155.3 (2000): 1405-1413.

Ferretti, Luca, Emanuele Raineri, and Sebastian Ramos-Onsins. "Neutrality tests for sequences with missing data." *Genetics* 191.4 (2012): 1397-1401.

Flint-Garcia, Sherry A., Jeffry M. Thornsberry, and Buckler IV. "Structure of linkage disequilibrium in plants*." *Annual review of plant biology* 54, no. 1 (2003): 357-374.

Franckowiak, J., 2002 Barley Genetic Newsletter 32: 109-109.

Franklin, Keara A., and Peter H. Quail. "Phytochrome functions in Arabidopsis development." *Journal of experimental botany* 61, no. 1 (2010): 11-24.

Fu, Limin Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. “CD-HIT: accelerated for clustering the next-generation sequencing data.” Bioinformatics, 28, (2012a): 3150-3152.

Fu, Yong-Bi. "Population-based resequencing analysis of wild and cultivated barley revealed weak domestication signal of selection and bottleneck in the Rrs2 scald resistance gene region." *Genome* 55, no. 2 (2012b): 93-104.

Fuller, Dorian Q. "Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World." *Annals of Botany*100, no. 5 (2007): 903-924.

Fuller, Dorian Q., Eleni Asouti, and Michael D. Purugganan. "Cultivation as slow evolutionary entanglement: comparative data on rate and sequence of domestication." *Vegetation history and archaeobotany* 21, no. 2 (2012): 131-145.

Galvão, Vinicius C., Karl JV Nordström, Christa Lanz, Patric Sulz, Johannes Mathieu, David Posé, Markus Schmid, Detlef Weigel, and Korbinian Schneeberger. "Synteny-based mapping-by-sequencing enabled by targeted enrichment." *The Plant Journal* 71, no. 3 (2012): 517-526.

Gan, Xiangchao, Oliver Stegle, Jonas Behr, Joshua G. Steffen, Philipp Drewe, Katie L. Hildebrand, Rune Lyngsoe et al. "Multiple reference genomes and transcriptomes for Arabidopsis thaliana." *Nature* 477, no. 7365 (2011): 419-423.

Garrigan, Daniel, and Michael F. Hammer. "Reconstructing human origins in the genomic era." *Nature Reviews Genetics* 7, no. 9 (2006): 669-680.

Gottlieb, L. D. "Genetics and morphological evolution in plants." *American Naturalist* (1984): 681-709.

Guo, Yan, Fei Ye, Quanghu Sheng, Travis Clark, and David C. Samuels. “Three-stage quality control strategies for DNA re-sequencing data.” *Briefings in bioinformatics* (2013): bbt069.

Guo, Yan, Jirong Long, Jing He, Chung-I. Li, Qiuyin Cai, Xiao-Ou Shu, Wei Zheng, and Chun Li. “Exome sequencing generates high quality data in non-target regions.” *BMC genomics* 13, no. 1 (2012): 194.

Haberer, Georg, and Klaus FX Mayer. "Barley: From Brittle to Stable Harvest." *Cell* 162, no. 3 (2015): 469-471.

Habte, E., Müller, L., Shtaya, M., Davis, S.J., M. von Korff, 2013 Osmotic stress at the barley root affects expression of circadian clock genes in the shoot. Plant Cell & Environment 36 (6): 1321-1337.

Hajjar, Reem, and Toby Hodgkin. "The use of wild relatives in crop improvement: a survey of developments over the last 20 years." *Euphytica* 156, no. 1-2 (2007): 1-13.

Hammer, Karl. "Das domestikationssyndrom." *Die Kulturpflanze* 32, no. 1 (1984): 11-34.

Hanumappa, Mamatha, Lee H. Pratt, Marie-Michele Cordonnier-Pratt, and Gerald F. Deitzer. "A

photoperiod-insensitive barley line contains a light-labile phytochrome B." *Plant physiology* 119, no. 3 (1999): 1033-1040.

Harlan, Jack R., and Daniel Zohary. "Distribution of wild wheats and barley." *Science* 153, no. 3740 (1966): 1074-1080.

Harlan, Jack R., and Jan MJ de Wet. "Toward a rational classification of cultivated plants." *Taxon* (1971): 509-517.

He, Ziwen, Weiwei Zhai, Haijun Wen, Tian Tang, Yu Wang, Xuemei Lu, Anthony J. Greenberg, Richard R. Hudson, Chung-I. Wu, and Suhua Shi. "Two evolutionary histories in the genome of rice: the roles of domestication genes." *PLoS Genet* 7, no. 6 (2011): e1002100.

Hedrick, Philip W. "Gametic disequilibrium measures: proceed with caution." *Genetics* 117.2 (1987): 331-341.

Hemming, Megan N., W. James Peacock, Elizabeth S. Dennis, and Ben Trevaskis. "Low-temperature and daylength cues are integrated to regulate FLOWERING LOCUS T in barley." *Plant Physiology* 147, no. 1 (2008): 355-366.

Hemming, Megan N., Sarah Fieg, W. James Peacock, Elizabeth S. Dennis, and Ben Trevaskis. "Regions associated with repression of the barley (Hordeum vulgare) VERNALIZATION1 gene are not required for cold induction." *Molecular Genetics and Genomics* 282, no. 2 (2009): 107-117.

Henry, Amanda G., Alison S. Brooks, and Dolores R. Piperno. "Plant foods and the dietary ecology of Neanderthals and early modern humans." Journal of human evolution 69 (2014a): 44-54.

Henry, Isabelle M., Ugrappa Nagalakshmi, Meric C. Lieberman, Kathie J. Ngo, Ksenia V. Krasileva, Hans Vasquez-Gross, Alina Akhunova et al. “Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing.” *The Plant Cell* 26, no. 4 (2014b): 1382-1397.

Herrero, Eva, Elsebeth Kolmos, Nora Bujdoso, Ye Yuan, Mengmeng Wang, Markus C. Berns, Heike Uhlworm et al. "EARLY FLOWERING4 recruitment of EARLY FLOWERING3 in the nucleus sustains the Arabidopsis circadian clock." *The Plant Cell* 24, no. 2 (2012): 428-443.

Heun, Manfred, Shahal Abbo, Simcha Lev-Yadun, and Avi Gopher. "A critical review of the protracted domestication model for Near-Eastern founder crops: linear regression, long-distance gene flow, archaeological, and archaeobotanical evidence." *Journal of experimental botany* 63, no. 12 (2012): 4333-4341.

Higgins, Janet A., Paul C. Bailey, and David A. Laurie. "Comparative genomics of flowering time pathways using Brachypodium distachyon as a model for the temperate grasses." *PLoS one* 5, no. 4 (2010): e10065.

Hill, W. G., and B. S. Weir. “Variances and covariances of squared linkage disequilibria in finite populations.” *Theoretical population biology* 33, no. 1 (1988): 54-78.

Hirsch, Cory D., Joseph Evans, C. Robin Buell, and Candice N. Hirsch. "Reduced representation approaches to interrogate genome diversity in large repetitive plant genomes." *Briefings in functional genomics* (2014): elt051.

Hu, W., K. Franklin, R. A. Sharrock, M. A. Jones, S. L. Harmer *et al.*, 2013 Unanticipated regulatory roles for Arabidopsis phytochromes revealed by null mutant analysis. Proc Natl Acad Sci U S A 110: 1542-1547.

Huang, W., P. Perez-Garcia, A. Pokhilko, A. J. Millar, I. Antoshechkin, José Luis Riechmann, and

Paloma Mas. "Mapping the core of the Arabidopsis circadian clock defines the network structure of the oscillator." *Science* 336, no. 6077 (2012): 75-79.

Huang, Xuehui, Nori Kurata, Xinghua Wei, Zi-Xuan Wang, Ahong Wang, Qiang Zhao, Yan Zhao et al. "A map of rice genome variation reveals the origin of cultivated rice." *Nature* 490, no. 7421 (2012): 497-501.

Hübner, S., M. Höffken, E. Oren, G. Haseneyer, N. Stein, A. Graner, K. Schmid, and E. Fridman. "Strong correlation of wild barley (Hordeum spontaneum) population structure with temperature and precipitation variation." *Molecular Ecology* 18, no. 7 (2009): 1523-1536.

Hübner, Sariel, Torsten Günther, Andrew Flavell, Eyal Fridman, Andreas Graner, Abraham Korol, and Karl J. Schmid. "Islands and streams: clusters and gene flow in wild barley populations from the Levant." *Molecular ecology* 21, no. 5 (2012): 1115-1129.

Hudson, R. R. (2002). Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinformatics, 18(2), 337-338.

Hufford, Matthew B., Xun Xu, Joost Van Heerwaarden, Tanja Pyhäjärvi, Jer-Ming Chia, Reed A. Cartwright, Robert J. Elshire et al. "Comparative population genomics of maize domestication and improvement." *Nature genetics* 44, no. 7 (2012): 808-811.

International Barley Genome Sequencing Consortium (IBGSC). “A physical, genetic and functional sequence assembly of the barley genome.” Nature 491.7426 (2012): 711-716.

Izawa, Takeshi, Tetsuo Oikawa, Nobuko Sugiyama, Takatoshi Tanisaka, Masahiro Yano, and Ko Shimamoto. "Phytochrome mediates the external light signal to repress FT orthologs in photoperiodic flowering of rice." *Genes & development* 16, no. 15 (2002): 2006-2020.

Izawa, Takeshi. “Adaptation of flowering-time by natural and artificial selection in Arabidopsis and rice.” *Journal of experimental botany* 58, no. 12 (2007): 3091-3097.

Jain, K. B. L. "Genetic Studies in Barley—III. Linkage Relations of some Plant Characters." *Indian Journal of Genetics and Plant Breeding (The)* 21, no. 1 (1961): 23-33.

Jakob, Sabine S., Dennis Rödder, Jan O. Engler, Salar Shaaf, Hakan Özkan, Frank R. Blattner, and Benjamin Kilian. "Evolutionary history of wild barley (Hordeum vulgare subsp. spontaneum) analyzed using multilocus sequence data and paleodistribution modeling." *Genome biology and evolution* 6, no. 3 (2014): 685-702.

James, G. V., V. Patel, K. J. V. Nordström, J. R. Klasen, P. A. Salomé *et al.*, User guide for mapping-by-sequencing in Arabidopsis. *Genome Biology*, 14 (2013): R61.

Jarne, Philippe. "Mating system, bottlenecks and genetic polymorphism in hermaphroditic animals." *Genetical Research* 65, no. 03 (1995): 193-207.

Johnson, Philip LF, and Montgomery Slatkin. “Accounting for bias from sequencing error in population genetic estimates.” *Molecular biology and evolution* 25, no. 1 (2008): 199-206.

Jones, Huw, Fiona J. Leigh, Ian Mackay, Mim A. Bower, Lydia MJ Smith, Michael P. Charles, Glynis Jones, Martin K. Jones, Terence A. Brown, and Wayne Powell. "Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the Fertile Crescent." *Molecular biology and evolution* 25, no. 10 (2008): 2211-2219.

Kandemir, N., A. Yildirim, D. A. Kudrna, P. M. Hayes, and A. Kleinhofs. "Marker assisted genetic analysis of non-brittle rachis trait in barley." *Hereditas* 141, no. 3 (2004): 272-277.

Karsai, I., P. Szűcs, K. Mészáros, T. Filichkina, P. M. Hayes, J. S. Skinner, L. Láng, and Z. Bedő. "The Vrn-H2 locus is a major determinant of flowering time in a facultative× winter growth habit barley (Hordeum vulgare L.) mapping population." *Theoretical and Applied Genetics*

110, no. 8 (2005): 1458-1466.

Katoh, K., and H. Toh, 2008 Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform 9: 286-298.

Kilian, Benjamin, Hakan Özkan, Jochen Kohl, Arndt von Haeseler, Francesca Barale, Oliver Deusch, Andrea Brandolini, Cemal Yucel, William Martin, and Francesco Salamini. "Haplotype structure at seven barley genes: relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication."*Molecular Genetics and Genomics* 276, no. 3 (2006): 230-241.

Kircher, Martin, Susanna Sawyer, and Matthias Meyer. "Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. "*Nucleic acids research* (2011): gkr771.

Koboldt, D., K. Chen, T. Wylie, D. E. Larson, M. D. McLellan *et al.*, 2009 VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 25: 2283-2285.

Kolmos, Elsebeth, Eva Herrero, Nora Bujdoso, Andrew J. Millar, Réka Tóth, Peter Gyula, Ferenc Nagy, and Seth J. Davis. "A reduced-function allele reveals that EARLY FLOWERING3 repressive action on the circadian clock is modulated by phytochrome signals in Arabidopsis." *The plant cell* 23, no. 9 (2011): 3230-3246.

Kolmos, Elsebeth, Monika Nowak, Maria Werner, Katrin Fischer, Guenter Schwarz, Sarah Mathews, Heiko Schoof, Ferenc Nagy, Janusz M. Bujnicki, and Seth J. Davis. "Integrating ELF4 into the circadian system through combined structural and functional studies." *HFSP journal* 3, no. 5 (2009): 350-366.

Komatsuda, Takao, Mohammad Pourkheirandish, Congfen He, Perumal Azhaguvel, Hiroyuki Kanamori, Dragan Perovic, Nils Stein et al. "Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene." *Proceedings of the National Academy of Sciences* 104, no. 4 (2007): 1424-1429.

Koppolu, Ravi, Nadia Anwar, Shun Sakuma, Akemi Tagiri, Udda Lundqvist, Mohammad Pourkheirandish, Twan Rutten et al. "Six-rowed spike4 (Vrs4) controls spikelet determinacy and row-type in barley." *Proceedings of the National Academy of Sciences* 110, no. 32 (2013): 13198-13203.

Lev-Yadun, Simcha, Avi Gopher, and Shahal Abbo. "The cradle of agriculture." Science 288, no. 5471 (2000): 1602.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. "The sequence alignment/map format and SAMtools." Bioinformatics 25, no. 16 (2009): 2078-2079.

Liu, X., M. Covington, C. Fankhauser, J. Chory, and D. R. Wagner, 2001 *ELF3* encodes a circadian clock-regulated nuclear protein that functions in an Arabidopsis PHYB signal transduction pathway. Plant Cell 13: 1293-1304.

Loytynoja, A., and N. Goldman, 2005 An algorithm for progressive multiple alignment of sequences with insertions. Proc Natl Acad Sci U S A 102: 10557-10562.

Luikart, G., F. W. Allendorf, J. M. Cornuet, and W. B. Sherwin. "Distortion of allele frequency distributions provides a test for recent population bottlenecks."*Journal of Heredity* 89, no. 3 (1998): 238-247.

Ma, Jianxin, and Jeffrey L. Bennetzen. "Rapid recent growth and divergence of rice nuclear genomes." *Proceedings of the National Academy of Sciences of the United States of America* 101, no. 34 (2004): 12404-12410.

Maruyama, Takeo, and Paul A. Fuerst. "Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck." *Genetics* 111, no. 3 (1985): 675-689.

Mascher, Martin, Gary J. Muehlbauer, Daniel S. Rokhsar, Jarrod Chapman, Jeremy Schmutz, Kerrie Barry, María Muñoz-Amatriaín et al. “Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ).” The Plant Journal 76, no. 4 (2013a): 718-727.

Mascher, Martin, Todd A. Richmond, Daniel J. Gerhardt, Axel Himmelbach, Leah Clissold, Dharanya Sampath, Sarah Ayling et al. “Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond.” *The Plant Journal* 76, no. 3 (2013b): 494-505.

Mathews, Sarah. "Evolutionary studies illuminate the structural-functional model of plant phytochromes." *The Plant Cell* 22, no. 1 (2010): 4-16.

Mathews, Sarah, Matt Lavin, and Robert A. Sharrock. "Evolution of the phytochrome gene family and its utility for phylogenetic analyses of angiosperms." *Annals of the Missouri Botanical Garden* (1995): 296-321.

Matsumoto, Takashi, Tsuyoshi Tanaka, Hiroaki Sakai, Naoki Amano, Hiroyuki Kanamori, Kanako Kurita, Ari Kikuta et al. “Comprehensive sequence analysis of 24,783 barley full-length cDNAs derived from 12 clone libraries.” Plant Physiology 156, no. 1 (2011): 20-28.

Mayer KF, Martis M, Hedley PE, Simková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H et al. 2011.Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23: 1249–1263.

McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella et al. “The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.” Genome research 20, no. 9 (2010): 1297-1303.

McVean, Gilean AT. "A genealogical interpretation of linkage disequilibrium."*Genetics* 162, no. 2 (2002): 987-991.

Min, X. J., Butler, G., Storms, R., & Tsang, A. (2005). OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic acids research,33(suppl 2), W677-W680.

Molina, Jeanmaire, Martin Sikora, Nandita Garud, Jonathan M. Flowers, Samara Rubinstein, Andy Reynolds, Pu Huang et al. "Molecular evidence for a single evolutionary origin of domesticated rice." *Proceedings of the National Academy of Sciences* 108, no. 20 (2011): 8351-8356.

Molina-Cano, J. L., P. Fra-Mon, G. Salcedo, C. Aragoncillo, F. Roca Togores, and Francisco García-Olmedo. "Morocco as a possible domestication center for barley: biochemical and agromorphological evidence." *TAG Theoretical and Applied Genetics* 73, no. 4 (1987): 531-536.

Monte, E., J. Alonso, J. R. Ecker, Y. Zhang, X. Li *et al.*, 2003 Isolation and characterization of phyC mutants in Arabidopsis reveals complex crosstalk between phytochrome signaling pathways. Plant Cell 15: 1962-1980.

Morrell, Peter L., Ana M. Gonzales, Kapua KT Meyer, and Michael T. Clegg. "Resequencing data indicate a modest effect of domestication on diversity in barley: a cultigen with multiple

origins." *Journal of Heredity* (2014): 105 (2): 253-264.

Morrell, Peter L., and Michael T. Clegg. "Genetic evidence for a second domestication of barley (Hordeum vulgare) east of the Fertile Crescent."*Proceedings of the national academy of sciences* 104, no. 9 (2007): 3289-3294.

Morrell, Peter L., Donna M. Toleno, Karen E. Lundy, and Michael T. Clegg. "Low levels of linkage disequilibrium in wild barley (Hordeum vulgare ssp. spontaneum) despite high rates of self-fertilization." *Proceedings of the National Academy of Sciences of the United States of America* 102, no. 7 (2005): 2442-2447.

Morrell, Peter L., Karen E. Lundy, and Michael T. Clegg. "Distinct geographic patterns of genetic diversity are maintained in wild barley (Hordeum vulgare ssp. spontaneum) despite migration." *Proceedings of the National Academy of Sciences* 100, no. 19 (2003): 10812-10817.

Morton, Brian R., Irie V. Bi, Michael D. McMullen, and Brandon S. Gaut. “Variation in mutation dynamics across the maize genome as a function of regional and flanking base composition.” *Genetics* 172, no. 1 (2006): 569-577.

Muñoz-Amatriaín, María, Stefano Lonardi, MingCheng Luo, Kavitha Madishetty, Jan T. Svensson, Matthew J. Moscou, Steve Wanamaker et al. "Sequencing of 15 622 gene-bearing BACs clarifies the gene-dense regions of the barley genome." *The Plant Journal* 84, no. 1 (2015): 216-227.

Nagatani, A., 2010 Phytochrome: structural basis for its functions. Curr Opin Plant Biol 13: 565-570.

Nei, Masatoshi, and Wen-Hsiung Li. "Mathematical model for studying genetic variation in terms of restriction endonucleases." *Proceedings of the National Academy of Sciences* 76.10 (1979): 5269-5273.

Nei, Masatoshi, Takeo Maruyama, and Ranajit Chakraborty. "The bottleneck effect and genetic variability in populations." *Evolution* (1975): 1-10.

Nevo E: Origin, evolution, population genetics and resources for breeding of wild barley, *Hordeum spontaneum*, in the fertile crescent, in Shewry PR (ed): Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology, pp 19–43 (CAB International, Wallingford 1992).

Nielsen, Rasmus, Joshua S. Paul, Anders Albrechtsen, and Yun S. Song. “Genotype and SNP calling from next-generation sequencing data.” *Nature Reviews Genetics* 12, no. 6 (2011): 443-451.

Nielsen, Rasmus, Melissa J. Hubisz, and Andrew G. Clark. “Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data.”*Genetics* 168, no. 4 (2004): 2373-2382.

Nielsen, Rasmus. "Molecular signatures of natural selection." *Annu. Rev. Genet.* 39 (2005): 197-218.

Nishida, Hidetaka, Daisuke Ishihara, Makoto Ishii, Takuma Kaneko, Hiroyuki Kawahigashi, Yukari Akashi, Daisuke Saisho et al. "Phytochrome C is a key factor controlling long-day flowering in barley." *Plant physiology* 163, no. 2 (2013): 804-814.

Nordborg, Magnus, Justin O. Borevitz, Joy Bergelson, Charles C. Berry, Joanne Chory, Jenny Hagenblad, Martin Kreitman et al. "The extent of linkage disequilibrium in Arabidopsis thaliana." *Nature genetics* 30, no. 2 (2002): 190-193.

Nordborg, Magnus, Tina T. Hu, Yoko Ishino, Jinal Jhaveri, Christopher Toomajian, Honggang Zheng, Erica Bakker et al. "The pattern of polymorphism in Arabidopsis thaliana." *PLoS*

biology 3, no. 7 (2005): 1289.

Onai, Kiyoshi, and Masahiro Ishiura. "PHYTOCLOCK 1 encoding a novel GARP protein essential for the Arabidopsis circadian clock." *Genes to Cells* 10, no. 10 (2005): 963-972.

Ossowski, Stephan, Korbinian Schneeberger, José Ignacio Lucas-Lledó, Norman Warthmann, Richard M. Clark, Ruth G. Shaw, Detlef Weigel, and Michael Lynch. “The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana.” S*cience* 327, no. 5961 (2010): 92-94.

Osugi, Asami, Hironori Itoh, Kyoko Ikeda-Kawakatsu, Makoto Takano, and Takeshi Izawa. "Molecular dissection of the roles of phytochrome in photoperiodic flowering in rice." *Plant physiology* 157, no. 3 (2011): 1128-1137.

Papaefthimiou, Dimitra, Eleni Likotrafiti, Aliki Kapazoglou, Konstantinos Bladenopoulos, and Athanasios Tsaftaris. "Epigenetic chromatin modifiers in barley: III. Isolation and characterization of the barley GNAT-MYST family of histone acetyltransferases and responses to exogenous ABA." *Plant Physiology and Biochemistry* 48, no. 2 (2010): 98-107.

Patterson N, Price AL, Reich D (2006) Population Structure and Eigenanalysis. PloS Genet 2(12): e190.

Pettersen, E., T. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt *et al.*, 2004 UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25: 1605-1612.

Pickering, R., and P. A. Johnston. "Recent progress in barley improvement using wild species of Hordeum." *Cytogenetic and genome research* 109, no. 1-3 (2005): 344-349.

Piperno, Dolores R., Anthony J. Ranere, Irene Holst, Jose Iriarte, and Ruth Dickau. "Starch grain and phytolith evidence for early ninth millennium BP maize from the Central Balsas River Valley, Mexico." *Proceedings of the National Academy of Sciences* 106, no. 13 (2009): 5019-5024.

Poets, Ana M., Zhou Fang, Michael T. Clegg, and Peter L. Morrell. "Barley landraces are characterized by geographically heterogeneous genomic origins."*Genome biology* 16, no. 1 (2015): 1-11.

Pokhilko, A., A. Fernández, K. D. Edwards, M. M. Southern, K. J. Halliday *et al.*, 2012 The clock gene circuit in Arabidopsis includes a repressilator with additional feedback loops. Mol Syst Biol 8: 574-574.

Poland, Jesse A., Patrick J. Brown, Mark E. Sorrells, and Jean-Luc Jannink. “Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach.” *PloS one* 7, no. 2 (2012): e32253.

Pontius JU, Wagner L, Schuler GD. UniGene: a unified view of the transcriptome. In: The NCBI Handbook. Bethesda (MD): National Center for Biotechnology Information; 2003.

Pourkheirandish, Mohammad, and Takao Komatsuda. "The importance of barley genetics and domestication in a global perspective." *Annals of Botany*100, no. 5 (2007): 999-1008.

Pourkheirandish, Mohammad, Goetz Hensel, Benjamin Kilian, Natesan Senthil, Guoxiong Chen, Mohammad Sameri, Perumal Azhaguvel et al. "Evolution of the Grain Dispersal System in Barley." *Cell* 162, no. 3 (2015): 527-539.

Pritchard, Jonathan K., Matthew Stephens, and Peter Donnelly. “Inference of population structure using multilocus genotype data.” *Genetics* 155, no. 2 (2000): 945-959.

Purugganan, Michael D., and Dorian Q. Fuller. "Archaeological data reveal slow rates of evolution

during plant domestication." *Evolution* 65, no. 1 (2011): 171-183.

Quevillon, Emmanuel, Ville Silventoinen, Sharmila Pillai, Nicola Harte, N. Mulder, Rolf Apweiler, and Rodrigo Lopez. "InterProScan: protein domains identifier." *Nucleic acids research* 33, no. suppl 2 (2005): W116-W120.

Quinlan, Aaron R. “BEDTools: The Swiss-Army Tool for Genome Feature Analysis.” Current Protocols in Bioinformatics (2014): 11-12.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Raj, Anil, Matthew Stephens, and Jonathan K. Pritchard. “fastSTRUCTURE: variational inference of population structure in large SNP data sets.” *Genetics*197, no. 2 (2014): 573-589.

Ramsay, Luke, Jordi Comadran, Arnis Druka, David F. Marshall, William TB Thomas, Malcolm Macaulay, Katrin MacKenzie et al. "INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene TEOSINTE BRANCHED 1." *Nature genetics* 43, no. 2 (2011): 169-172.

Ritala, A., A. M. Nuutila, R. Aikasalo, V. Kauppinen, and J. Tammisola. "Measuring gene flow in the cultivation of transgenic barley." *Crop Science* 42, no. 1 (2002): 278-285.

Rollins, Jarod A., B. Drosse, M. A. Mulki, S. Grando, M. Baum, M. Singh, S. Ceccarelli, and M. von Korff. "Variation at the vernalisation genes Vrn-H1 and Vrn-H2 determines growth and yield stability in barley (Hordeum vulgare) grown under dryland conditions in Syria." *Theoretical and applied genetics* 126, no. 11 (2013): 2803-2824.

Rosenblum, E. B., & Novembre, J. (2007). Ascertainment bias in spatially structured populations: a case study in the eastern fence lizard. *Journal of Heredity*, *98*(4), 331-336.

Ross-Ibarra, Jeffrey, and Brandon S. Gaut. "Multiple domestications do not appear monophyletic." *Proceedings of the National Academy of Sciences of the United States of America* 105, no. 49 (2008): E105.

Ross-Ibarra, Jeffrey, Peter L. Morrell, and Brandon S. Gaut. "Plant domestication, a unique opportunity to identify the genetic basis of adaptation."*Proceedings of the National Academy of Sciences* 104, no. suppl 1 (2007): 8641-8648

Russell, Joanne, Allan Booth, John Fuller, Brian Harrower, Peter Hedley, Gordon Machray, and Wayne Powell. "A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome." *Genome* 47, no. 2 (2004): 389-398.

Russell, Joanne, Ian K. Dawson, Andrew J. Flavell, Brian Steffenson, Eva Weltzien, Allan Booth, Salvatore Ceccarelli, Stefania Grando, and Robbie Waugh. "Analysis of> 1000 single nucleotide polymorphisms in geographically matched samples of landrace and wild barley indicates secondary contact and chromosome-level differences in diversity around domestication genes." *New Phytologist* 191, no. 2 (2011): 564-578.

S.A.S. Institute, 2009 The SAS system for Windows, release 9.1.3. SAS Institute Inc., Cary, NC, USA.

Saisho, Daisuke, and Michael D. Purugganan. "Molecular phylogeography of domesticated barley traces expansion of agriculture in the Old World." *Genetics*177, no. 3 (2007): 1765-1776.

Salamini, F., et al. "Comment on" AFLP data and the origins of domesticated crops"." *Genome* 47.3 (2004): 615-620.

Salamini, Francesco, Hakan Özkan, Andrea Brandolini, Ralf Schäfer-Pregl, and William Martin.

"Genetics and geography of wild cereal domestication in the Near East." *Nature Reviews Genetics* 3, no. 6 (2002): 429-441.

Saïdou, Abdoul-Aziz, Cédric Mariac, Vivianne Luong, Jean-Louis Pham, Gilles Bezançon, and Yves Vigouroux. "Association studies identify natural variation at PHYC linked to flowering time and morphological variation in pearl millet."*Genetics* 182, no. 3 (2009): 899-910.

Schmid, Karl J., Sebastian Ramos-Onsins, Henriette Ringys-Beckstein, Bernd Weisshaar, and Thomas Mitchell-Olds. "A multilocus sequence survey in Arabidopsis thaliana reveals a genome-wide departure from a neutral model of DNA sequence polymorphism." *Genetics* 169, no. 3 (2005): 1601-1615.

Schmutz, Jeremy, Phillip E. McClean, Sujan Mamidi, G. Albert Wu, Steven B. Cannon, Jane Grimwood, Jerry Jenkins et al. "A reference genome for common bean and genome-wide analysis of dual domestications." *Nature genetics* 46, no. 7 (2014): 707-713.

Schneeberger, Korbinian, Stephan Ossowski, Christa Lanz, Trine Juul, Annabeth Høgh Petersen, Kåre Lehmann Nielsen, Jan-Elo Jørgensen, Detlef Weigel, and Stig Uggerhø Andersen. "SHOREmap: simultaneous mapping and mutation identification by deep sequencing." *Nature Methods* 6, no. 8 (2009): 550-551. Schöning, J., and D. Staiger, 2005 At the pulse of time: protein interactions determine the pace of circadian clocks. FEBS Lett 579: 3246-3252.

Shi, Junpeng, and Jinsheng Lai. "Patterns of genomic changes with crop domestication and breeding." *Current opinion in plant biology* 24 (2015): 47-53.

Simko, Ivan, Kathleen G. Haynes, and Richard W. Jones. “Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers.” *Genetics* 173, no. 4 (2006): 2237-2245.

Simpson, J., K. Wong, S. D. Jackman, J. E. Schein, S. J. Jones *et al.*, 2009 ABySS: a parallel assembler for short read sequence data. Genome Res 19: 1117-1123.

Slatkin, Montgomery, and Bruce Rannala. "Estimating allele age." *Annual review of genomics and human genetics* 1.1 (2000): 225-249.

Smith, John Maynard, and John Haigh. "The hitch-hiking effect of a favourable gene." *Genetical research* 23, no. 01 (1974): 23-35.

Somers, David E., Paul F. Devlin, and Steve A. Kay. "Phytochromes and cryptochromes in the entrainment of the Arabidopsis circadian clock." *Science*282, no. 5393 (1998): 1488-1490.

Song, L., Florea, L., & Langmead, B. (2014). Lighter: fast and memory-efficient sequencing error correction without counting. Genome biology, 15(11), 1-13.

Stephan, Wolfgang. "Genetic hitchhiking versus background selection: the controversy and its implications." *Philosophical Transactions of the Royal Society B: Biological Sciences* 365, no. 1544 (2010): 1245-1253.

Szücs, P., I. Karsai, J. von Zitzewitz, K. Mészáros, L. L. D. Cooper *et al.*, 2006 Positional relationships between photoperiod response QTL and photoreceptor and vernalization genes in barley. Theor Appl Genet 112: 1277-1285.

Tajima, Fumio. "Evolutionary relationship of DNA sequences in finite populations." *Genetics* 105.2 (1983): 437-460.

Tajima, Fumio. "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism." *Genetics* 123.3 (1989): 585-595.

Takano, M., N. Inagaki, X. Xie, N. Yuzurihara, F. Hihara *et al.*, 2005 Distinct and cooperative

functions of phytochromes A, B, and C in the control of deetiolation and flowering in rice. Plant Cell 17: 3311-3325.

Taketa, S., S. Kikuchi, T. Awayama, S. Yamamoto, M. Ichii, and S. Kawasaki. "Monophyletic origin of naked barley inferred from molecular analyses of a marker closely linked to the naked caryopsis gene (nud)." *Theoretical and Applied Genetics* 108, no. 7 (2004): 1236-1242.

Taketa, Shin, Satoko Amano, Yasuhiro Tsujino, Tomohiko Sato, Daisuke Saisho, Katsuyuki Kakeda, Mika Nomura et al. "Barley grain with adhering hulls is controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway." *Proceedings of the National Academy of Sciences* 105, no. 10 (2008): 4062-4067.

Tenaillon, Maud I., Jana U'Ren, Olivier Tenaillon, and Brandon S. Gaut. "Selection versus demography: a multilocus investigation of the domestication process in maize." *Molecular Biology and Evolution* 21, no. 7 (2004): 1214-1225.

Teshima, Kosuke M., Graham Coop, and Molly Przeworski. "How reliable are empirical genomic scans for selective sweeps?." *Genome research* 16, no. 6 (2006): 702-712.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker et al. "The variant call format and VCFtools." *Bioinformatics* 27, no. 15 (2011): 2156-2158.

Tolbert, D. M., C. O. Qualset, S. K. Jain, and J. C. Craddock. "A diversity analysis of a world collection of barley." *Crop Science* 19, no. 6 (1979): 789-794.

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

USDA. “Market Review January 2014: Outlook for the Agricultural Markets in 2015.” ADM Germany GmbH, (2015)

van Heerwaarden, Joost, John Doebley, William H. Briggs, Jeffrey C. Glaubitz, Major M. Goodman, Jose de Jesus Sanchez Gonzalez, and Jeffrey Ross-Ibarra. "Genetic signals of origin, spread, and introgression in a large sample of maize landraces." *Proceedings of the National Academy of Sciences* 108, no. 3 (2011): 1088-1092.

Vavilov NI, Studies on the origin of cultivated plants. Institute of Applied Botany and Plant Breeding (1926), Leningrad

Verhoeven, K. J. F., H. Poorter, E. Nevo, and A. Biere. “Habitat-specific natural selection at a flowering-time QTL is a main driver of local adaptation in two wild barley populations.” *Molecular Ecology* 17, no. 14 (2008): 3416-3424.

Vitti, Joseph J., Sharon R. Grossman, and Pardis C. Sabeti. "Detecting natural selection in genomic data." *Annual review of genetics* 47 (2013): 97-120.

von Korff, M., H. Wang, J. Léon, and K. Pillen. "AB-QTL analysis in spring barley: II. Detection of favourable exotic alleles for agronomic traits introgressed from wild barley (H. vulgare ssp. spontaneum)." *Theoretical and Applied Genetics* 112, no. 7 (2006): 1221-1231.

von Korff, Maria, Jens Léon, and Klaus Pillen. "Detection of epistatic interactions between exotic alleles introgressed from wild barley (H. vulgare ssp. spontaneum)." *Theoretical and applied genetics* 121, no. 8 (2010): 1455-1464.

Wagner, D. B., and R. W. Allard. "Pollen migration in predominantly self-fertilizing plants: barley." *Journal of Heredity* 82, no. 4 (1991): 302-304.

Wang, Gongwei, Inga Schmalenbach, Maria von Korff, Jens Léon, Benjamin Kilian, Jeannette Rode, and Klaus Pillen. "Association of barley photoperiod and vernalization genes with

QTLs for flowering time and agronomic traits in a BC2DH population and a set of wild barley introgression lines." *Theoretical and Applied Genetics* 120, no. 8 (2010): 1559-1574.

Warschefsky, Emily, R. Varma Penmetsa, Douglas R. Cook, and Eric JB von Wettberg. "Back to the wilds: Tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives." *American journal of botany* 101, no. 10 (2014): 1791-1800.

Watterson, G. A. "On the number of segregating sites in genetical models without recombination." *Theoretical population biology* 7.2 (1975): 256-276.

Weir, Bruce S., and C. Clark Cockerham. "Estimating F-statistics for the analysis of population structure." E*volution* (1984): 1358-1370.

Wheelan, Sarah J., Deanna M. Church, and James M. Ostell. “Spidey: a tool for mRNA-to-genomic alignments.” Genome Research 11.11 (2001): 1952-1957.

Wicker, Thomas, Klaus FX Mayer, Heidrun Gundlach, Mihaela Martis, Burkhard Steuernagel, Uwe Scholz, Hana Šimková et al. „Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives.” *The Plant Cell* 23, no. 5 (2011): 1706-1718.

Winfield, Mark O., Paul A. Wilkinson, Alexandra M. Allen, Gary LA Barker, Jane A. Coghill, Amanda Burridge, Anthony Hall et al. “Targeted re-sequencing of the allohexaploid wheat exome.” *Plant biotechnology journal* 10, no. 6 (2012): 733-742.

Wolfenbarger, Laressa L., and Paul R. Phifer. "The ecological risks and benefits of genetically engineered plants." *Science* 290, no. 5499 (2000): 2088-2093.

Xia, Y., Z. Ning, G. Bai, R. Li, G. Yan *et al.*, 2012 Allelic variations of a light harvesting chlorophyll a/b-binding protein gene (*Lhcb1*) associated with agronomic traits in barley. PLoS One 7: e37573-e37573.

Xu, T. W. "Origin and evolution of cultivated barley in China." *I ch'uan hsueh pao= Acta genetica Sinica* (1982).

Xu, Xun, Xin Liu, Song Ge, Jeffrey D. Jensen, Fengyi Hu, Xin Li, Yang Dong et al. "Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes." *Nature biotechnology* 30, no. 1 (2012): 105-111.

Yan, Liuling, Artem Loukoianov, Ann Blechl, Gabriela Tranquilli, Wusirika Ramakrishna, Phillip SanMiguel, Jeffrey L. Bennetzen, Viviana Echenique, and Jorge Dubcovsky. "The wheat VRN2 gene is a flowering repressor down-regulated by vernalization." *Science* 303, no. 5664 (2004): 1640-1644.

Yan, Liuling, Artem Loukoianov, Gabriela Tranquilli, Marcelo Helguera, T. Fahima, and Jorge Dubcovsky. "Positional cloning of the wheat vernalization gene VRN1." *Proceedings of the National Academy of Sciences* 100, no. 10 (2003): 6263-6268.

Yan, Limei, Dong Fu, Chunlei Li, Amm Blechl, Gabriela Tranquilli, M. Bonafede, A. Sanchez, Miroslav Valarik, S. Yasuda, and Jorge Dubcovsky. "The wheat and barley vernalization gene VRN3 is an orthologue of FT." *Proceedings of the National Academy of Sciences* 103, no. 51 (2006): 19581-19586.

Yang, Ziheng, and Joseph P. Bielawski. “Statistical methods for detecting molecular adaptation.” *Trends in ecology & evolution* 15, no. 12 (2000): 496-503.

Zakhrabekova, Shakhira, Simon P. Gough, Ilka Braumann, André H. Müller, Joakim Lundqvist, Katharina Ahmann, Christoph Dockter et al. "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 109, no. 11 (2012): 4326-4331.

Zeng, K., Fu, Y. X., Shi, S., & Wu, C. I. “Statistical tests for detecting positive selection by utilizing high-frequency variants.” *Genetics*, *174*(3), (2006): 1431-1439.

Zeng, Kai, Suhua Shi, and Chung-I. Wu. "Compound tests for the detection of hitchhiking under positive selection." *Molecular biology and evolution* 24, no. 8 (2007): 1898-1908.

Zeng, Xingquan, Hai Long, Zhuo Wang, Shancen Zhao, Yawei Tang, Zhiyong Huang, Yulin Wang et al. "The draft genome of Tibetan hulless barley reveals adaptive patterns to the high stressful Tibetan Plateau." *Proceedings of the National Academy of Sciences* 112, no. 4 (2015): 1095-1100.

Zeven, Anton C. "Landraces: a review of definitions and classifications."*Euphytica* 104, no. 2 (1998): 127-139.

Zhang, Lin-Bin, Qihui Zhu, Zhi-Qiang Wu, Jeffrey Ross-Ibarra, Brandon S. Gaut, Song Ge, and Tao Sang. "Selection on grain shattering genes and rates of rice domestication." *New Phytologist* 184, no. 3 (2009): 708-720.

Zhou, Zhengkui, Yu Jiang, Zheng Wang, Zhiheng Gou, Jun Lyu, Weiyu Li, Yanjun Yu et al. "Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean." *Nature biotechnology* (2015).

Zhu, Danmeng, Alexander Maier, Jae-Hoon Lee, Sascha Laubinger, Yusuke Saijo, Haiyang Wang, Li-Jia Qu, Ute Hoecker, and Xing Wang Deng. "Biochemical characterization of Arabidopsis complexes containing CONSTITUTIVELY PHOTOMORPHOGENIC1 and SUPPRESSOR OF PHYA proteins in light control of plant development." *The Plant Cell* 20, no. 9 (2008): 2307-2323.

Zhu, Qihui, Xiaoming Zheng, Jingchu Luo, Brandon S. Gaut, and Song Ge. "Multilocus analysis of nucleotide variation of Oryza sativa and its wild relatives: severe bottleneck during domestication of rice." *Molecular Biology and Evolution* 24, no. 3 (2007): 875-888.

Zhu, Y. L., Q. J. Song, D. L. Hyten, C. P. Van Tassell, L. K. Matukumalli, D. R. Grimm, S. M. Hyatt, E. W. Fickus, N. D. Young, and P. B. Cregan. “Single-nucleotide polymorphisms in soybean.” *Genetics* 163, no. 3 (2003): 1123-1134.

Zohary, Daniel, and Maria Hopf. "Domestication of plants in the world. The origin and spread of cultivated plants in west Asia, Europe and the Nile valley." (1993).

Zohary, Daniel. "Monophyletic vs. polyphyletic origin of the crops on which agriculture was founded in the Near East." *Genetic Resources and Crop Evolution* 46, no. 2 (1999): 133-142.

Zohary, Daniel. "Unconscious selection and the evolution of domesticated plants." *Economic Botany* 58, no. 1 (2004): 5-10.

SUPPLEMENTARY MATERIAL - CHAPTER 1

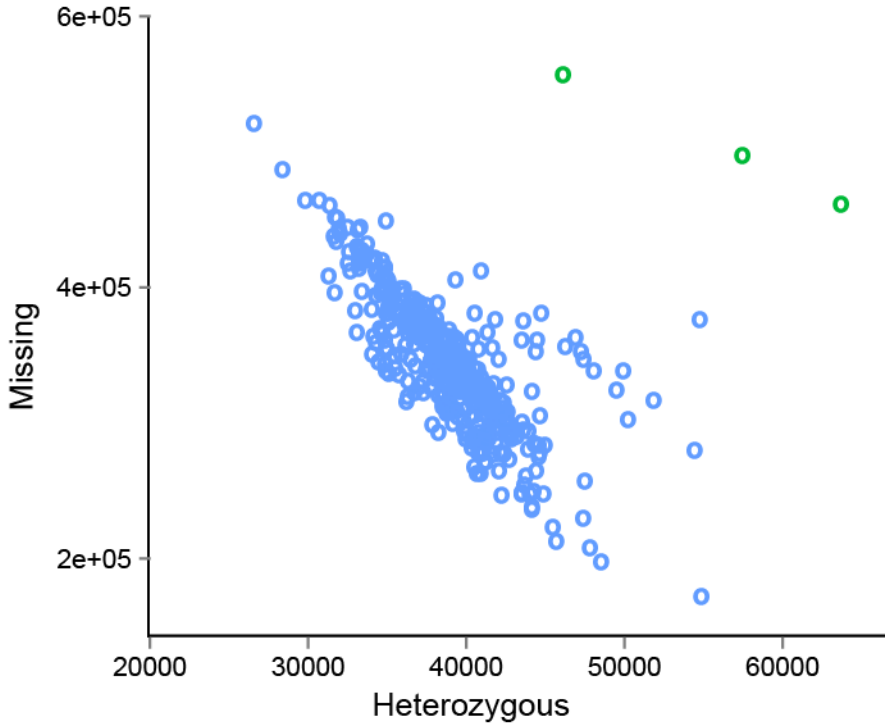


Figure S1. Distribution of heterozygous and missing genotypes in the SNP dataset.

The pooled and individually-sequenced samples are depicted as blue and green pictograms, respectively.

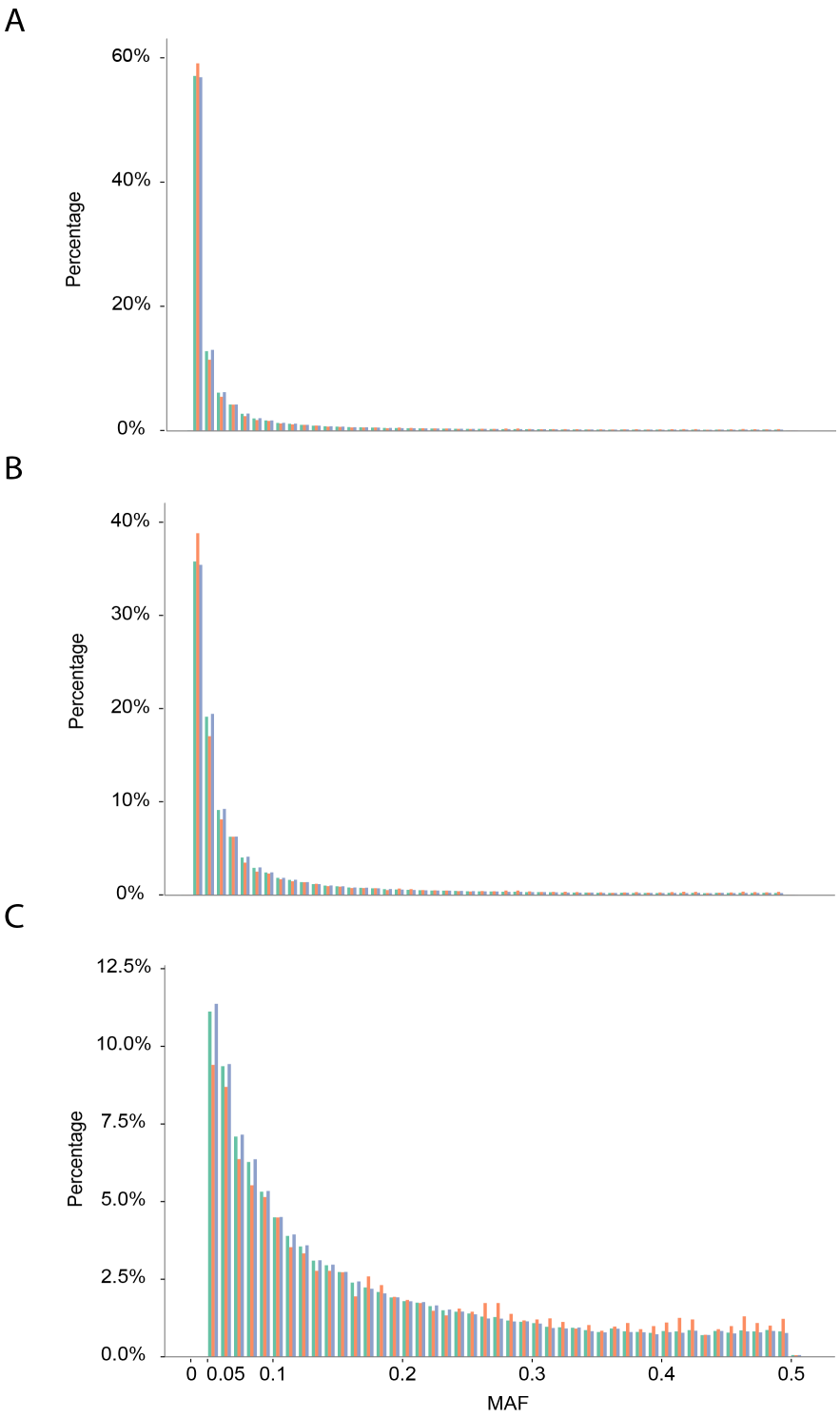


Figure S2. Distribution of minor allele frequency (MAF) of the SNPs.

The MAF spectra are shown for the complete set with singleton SNPs (A), without singleton SNPs (B) and for the dataset with MAF >= 0.05 (C). All the sub-sets contain the SNPs with < 0.5 missing data frequency. Cyan, red and blue bars represent the MAF distributions in all, coding and non-coding SNP subsets, respectively.

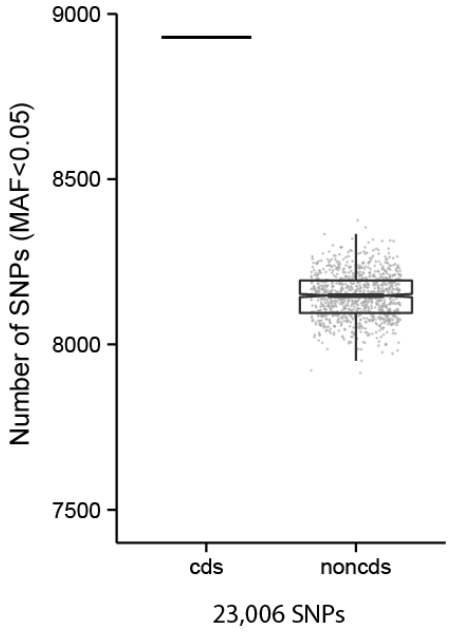


Figure S3. Robustness of the MAF spectra calculations in the subsets with unequal number of SNPs.

A set of 23,006 SNPs, equal to a total number of coding SNPs, was 1,000 times randomly drawn from a larger dataset of non-coding SNPs (193,185). The resulting numbers of rare SNPs (MAF < 0.05) are shown as gray dots and distribution statistics is summarized in a notched box plot.

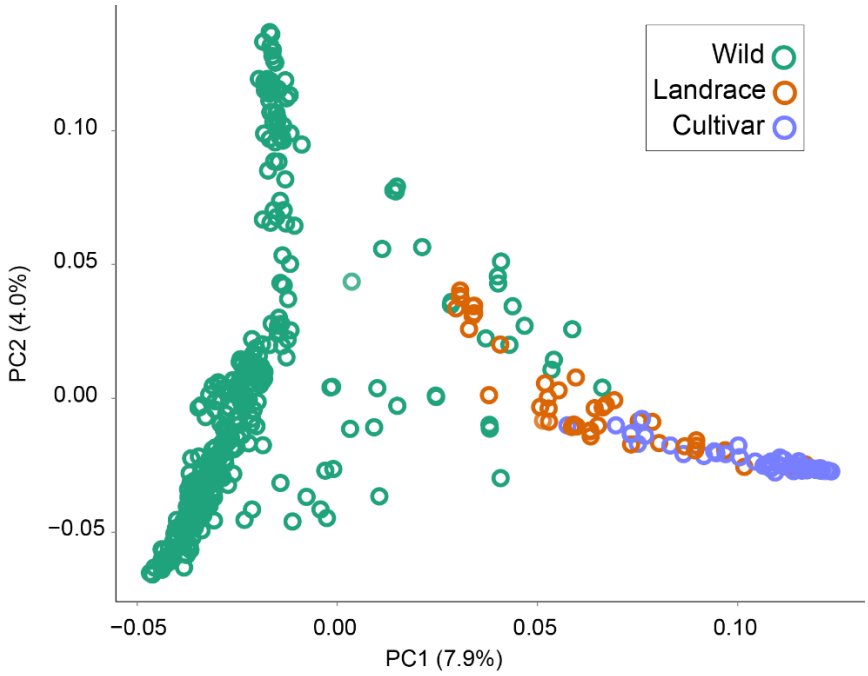


Figure S4. Decomposition of genetic variation of all barley genotypes by PCA.

Only two major PCs are shown. The percentages of variation explained by the PCs are shown in parentheses. Wild, landrace and cultivated genotypes are illustrated by green, orange and blue pictograms.

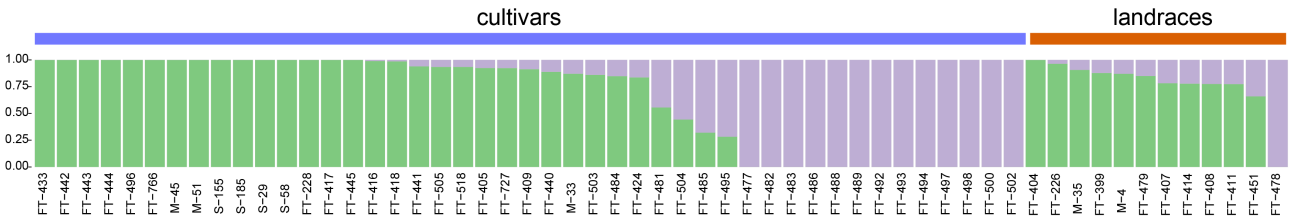


Figure S5. Population structure analysis of the cultivar and landrace genotypes (K=2).

Vertical bars correspond to individual genotypes and color indicates their membership in the subpopulations. The cultivar and landrace subgroups are labeled by blue and orange horizontal bars, respectively.

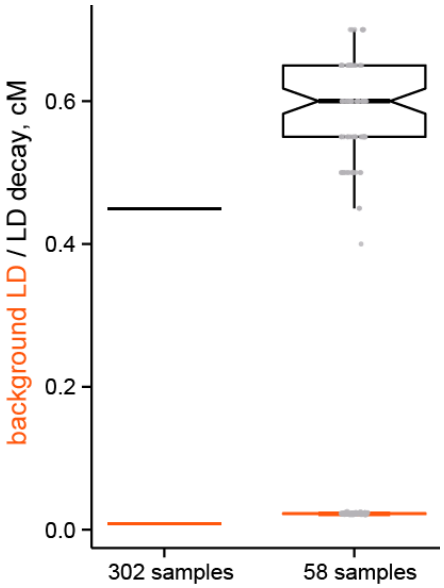


Figure S6. Variation of linkage disequilibrium (LD) parameters in the randomly sub-sampled wild genotypes.

58 genotypes were 1000 times randomly drawn from a total of 302 wild genotypes. The red horizontal line and box plot illustrate variation in the background LD, the black horizontal line and box plot in the LD decay. Gray dots correspond to the individual LD estimates for each random subsample.

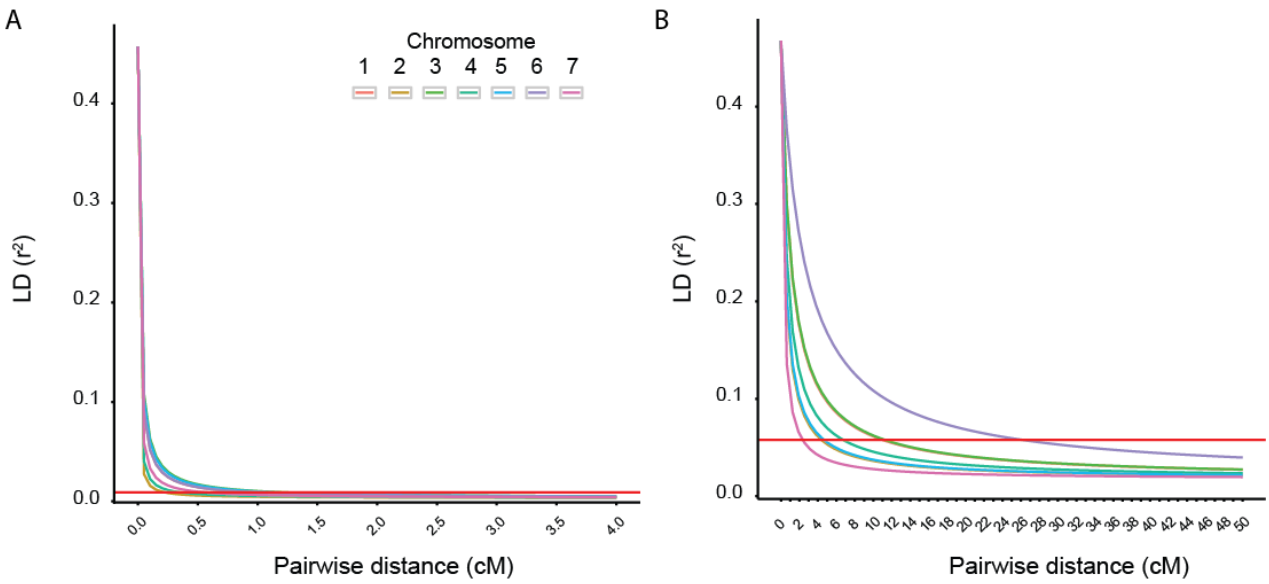


Figure S7. LD decay of individual chromosomes in wild (A) and domesticated (B) barley.

Red horizontal lines show the background LD. Colors of the regression line correspond to individual chromosomes according to the legend.

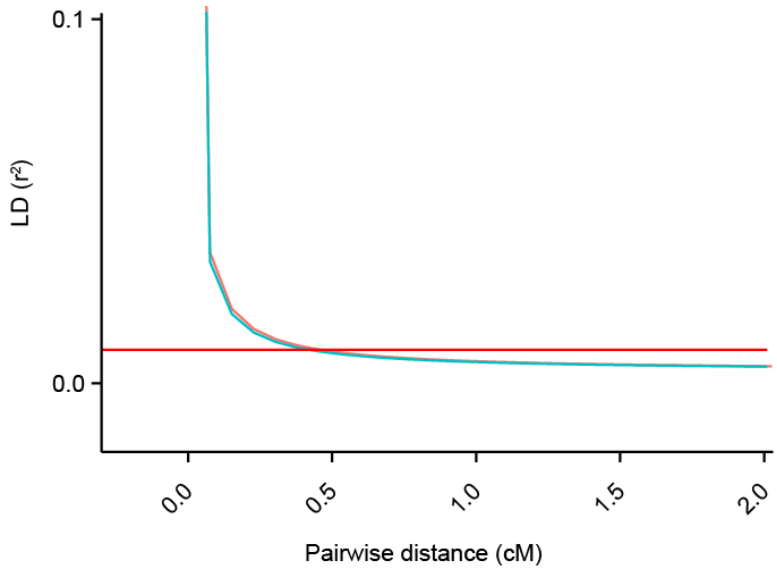


Figure S8. Comparison of LD decay rate estimated using all SNP pairs vs. the SNP pairs with matching minor allele frequency (MAF) in 0.1 bins.

Red horizontal line delineates the level of background LD. Red and cyan regression curves correspond to LD decay obtained from the SNP pairs matched by MAF and all SNP pairs, respectively.

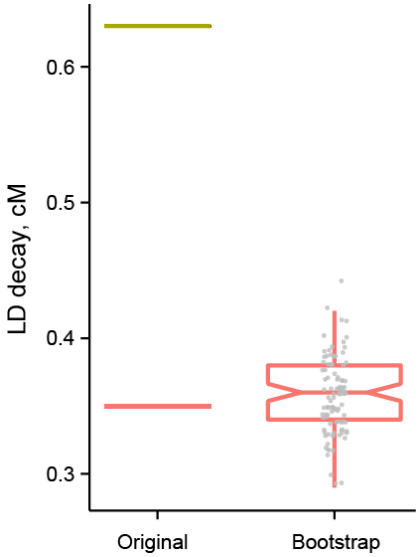


Figure S9. Effect of the varying numbers of frequency-matched SNP on the LD decay estimates in wild barley.

The original dataset contains 19,772 and 3,510 SNPs in the 0.1 (red) and 0.2 (green) MAF bins. The bootstrapped dataset contains 1000x random draws of the 3,510 SNPs from the 0.1 MAF bin.

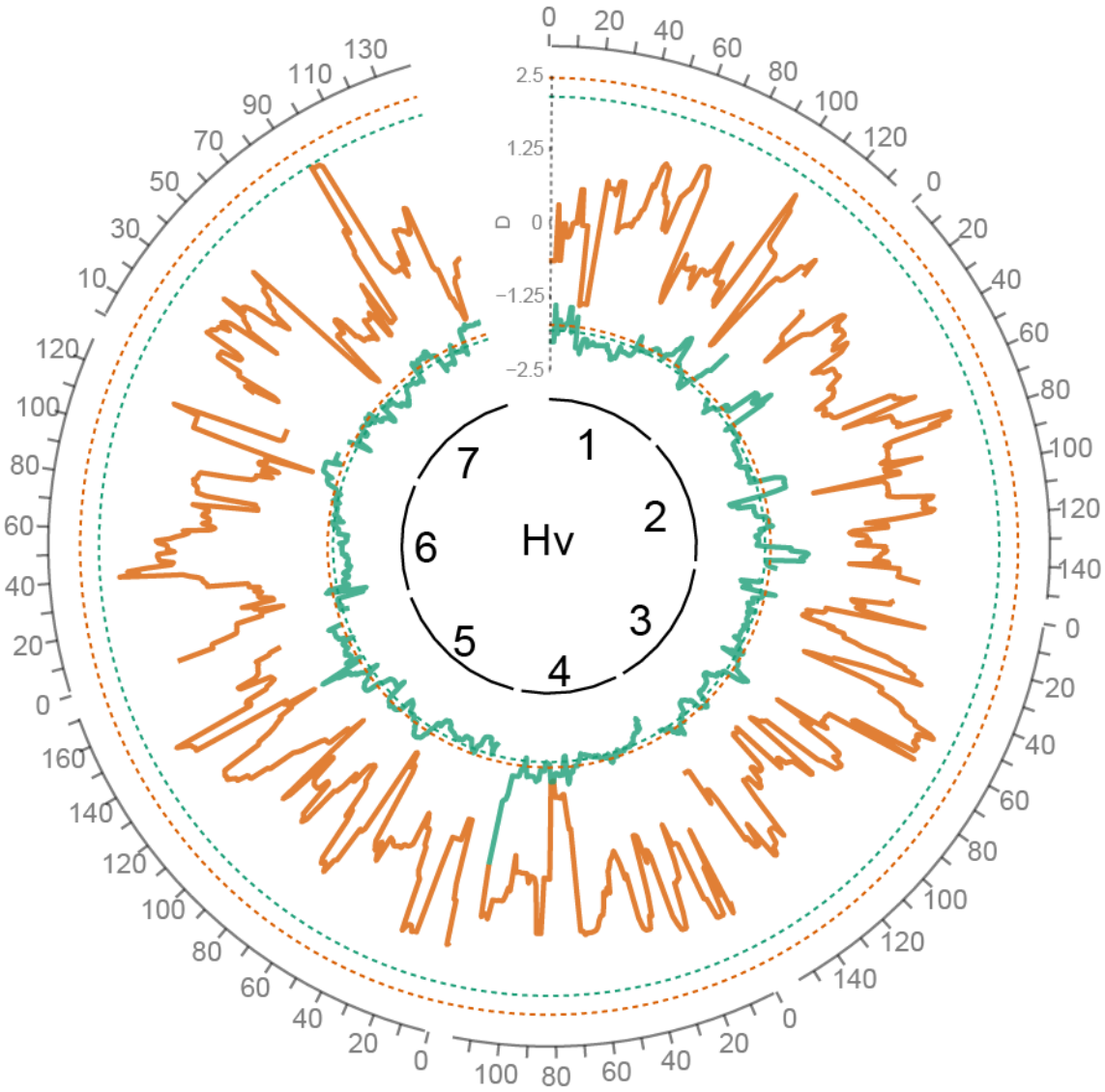


Figure S10. Distribution of Tajima's D values along the chromosomes.

The numbers inside the innermost circle indicate barley linkage groups and genetic distances in cM are shown on the outermost gray scale. Tajima's D values (sliding window 10 cM, 1-cM step) are shown for the wild (green) and domesticated (orange) barley subgroups. Thresholds of neutral D variation obtained by coalescent simulations are shown for wild and domesticated barley as green and orange dashed lines, respectively. The scale of the D values is shown at the '0' position of the linkage group 1.

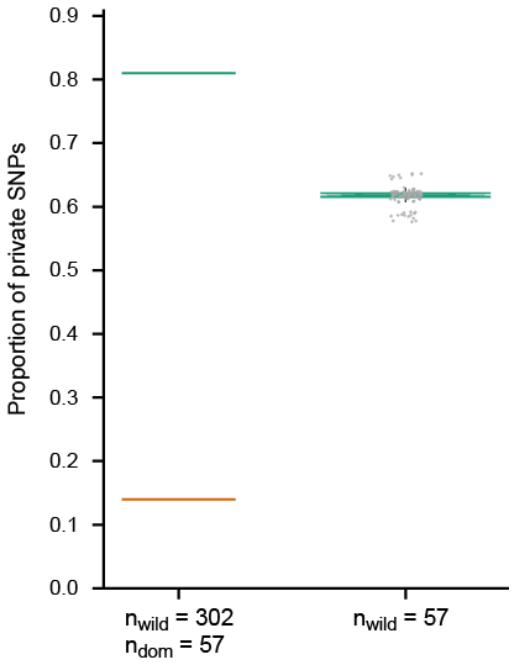


Figure S11. Effect of the unequal number of samples on estimates of private allele ratios in wild and domesticated barley.

The horizontal green and red lines show the ratios calculated from the original wild and domesticated barley datasets, respectively. The green boxplot and the gray dots illustrate variation of the ratios in the equalized wild barley dataset.

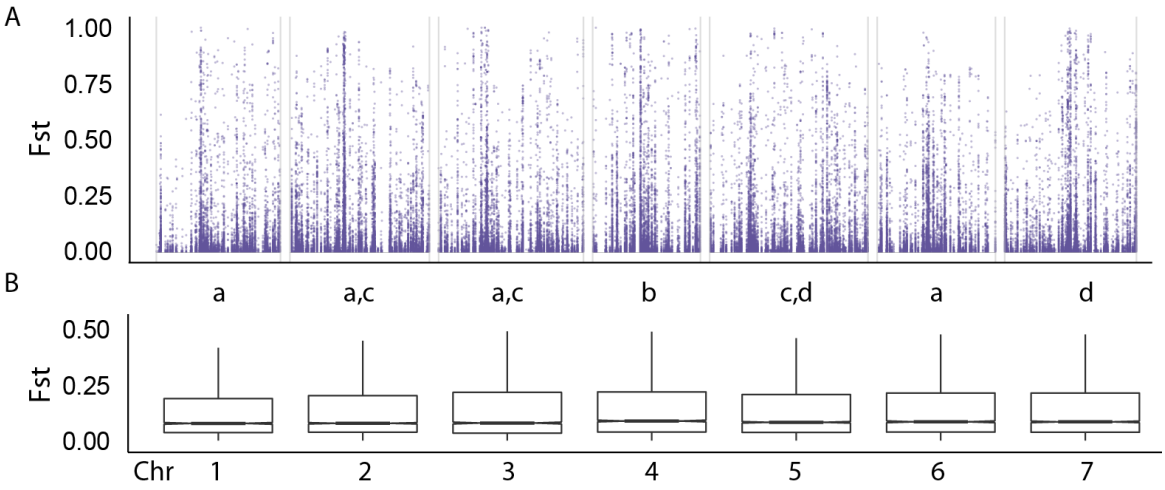


Figure S12. Fixation index (Fst) for all SNP sites between wild and domesticated barley.

(A) The individual Fst values plotted along the barley chromosomes.
(B) Summary statistics for distributions of the Fst values by chromosome. The letters indicate significantly different distributions ($p < 0.01$).

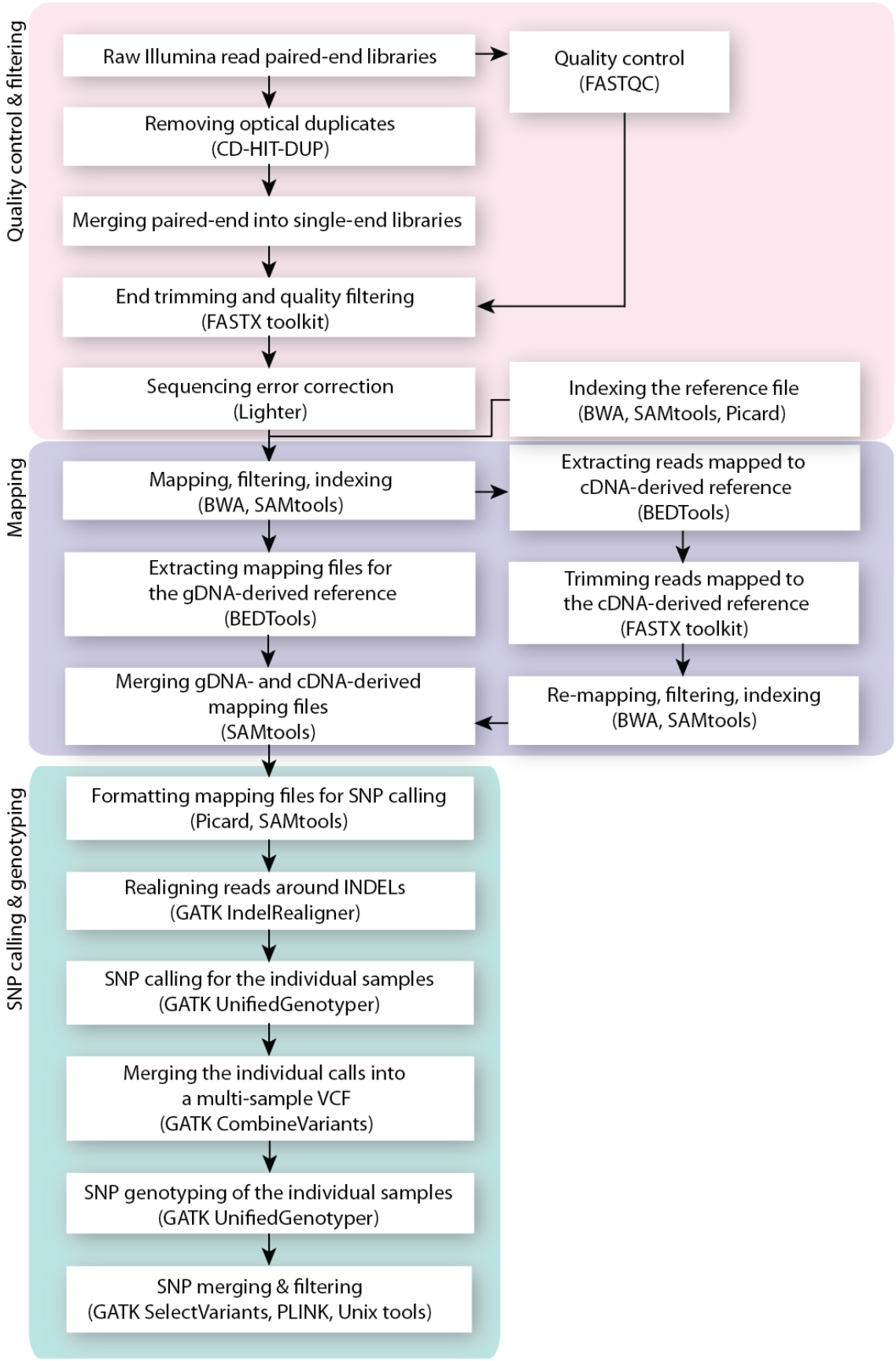


Figure S11. The data analysis pipeline - read filtering, mapping, SNP calling and genotyping.

Table S1 Target genes selected for enrichment.

https://www.dropbox.com/s/b06jo62e6oofgla/Table_S1.xlsx?dl=0

Table S2 Reference genome for read mapping

https://www.dropbox.com/s/9wq46d4qgcreova/Table_S2.xlsx?dl=0

Table S5 Arabidopsis homologs of flowering, meristem and inflorescence development genes.

https://www.dropbox.com/s/rpj4jjs5i5clcsp/Table_S5.xlsx?dl=0

Table S3 Characteristics of the germplasm used for re-sequencing.

Sample_ID	Line_ID	Source*	Country**	Status	Structure	Membership	
						Pop1	Pop2
FT227	B1K-70-01	IPK	ISR	cultivar	admixed	0.8329	0.1671
FT425	Steptoe	IPK	USA	cultivar	admixed	0.893	0.107
FT450	HOR1804	IPK	AFG	landrace	admixed	0.5571	0.4429
FT415	Chilean	IPK	AUS	landrace	admixed	0.5785	0.4215
FT403	Tibet	IPK	CHN	landrace	admixed	0.645	0.355
MK052	Paishapu 1	NBRP	CHN	landrace	admixed	0.5387	0.4613
MK053	Yanghsin 2	NBRP	CHN	landrace	admixed	0.5453	0.4547
MK054	Paoanchen 1	NBRP	CHN	landrace	admixed	0.5359	0.4641
FT402	G-401 I	IPK	DZA	landrace	admixed	0.8717	0.1283
FT437	L871	IPK	EGY	landrace	admixed	0.7419	0.2581
FT412	G-434 H	IPK	ETH	landrace	admixed	0.5852	0.4148
FT435	L521	IPK	ETH	landrace	admixed	0.8985	0.1015
FT436	L1043	IPK	IRN	landrace	admixed	0.5152	0.4848
FT434	L1897	IPK	JOR	landrace	admixed	0.715	0.285
MK003	Rum	MPIPZ	JOR	landrace	admixed	0.8384	0.1616
MK034	Acsad	MPIPZ	JOR	landrace	admixed	0.7933	0.2067
FT413	Nippon	IPK	JPN	landrace	admixed	0.5832	0.4168
FT400	G-399 H	IPK	MAR	landrace	admixed	0.827	0.173
FT447	BCC129	IPK	MAR	landrace	admixed	0.8594	0.1406
FT448	BCC131	IPK	MAR	landrace	admixed	0.8484	0.1516
FT410	G-427 G	IPK	PAK	landrace	admixed	0.5781	0.4219
FT421	G-1573 A	IPK	SYR	landrace	admixed	0.7441	0.2559
FT438	Tadmor	IPK	SYR	landrace	admixed	0.7532	0.2468
FT439	Arta	IPK	SYR	landrace	admixed	0.7498	0.2502
FT398	G-396 I	IPK	URY	landrace	admixed	0.8621	0.1379
MK001	B21	MPIPZ	n.d.	landrace	admixed	0.7882	0.2118
MK002	B25	MPIPZ	n.d.	landrace	admixed	0.6651	0.3349
MK005	L761	MPIPZ	n.d.	landrace	admixed	0.8718	0.1282
MK025	B24	MPIPZ	n.d.	landrace	admixed	0.8591	0.1409
MK028	B28	MPIPZ	n.d.	landrace	admixed	0.8755	0.1245
MK029	B29	MPIPZ	n.d.	landrace	admixed	0.8791	0.1209
MK030	B30	MPIPZ	n.d.	landrace	admixed	0.762	0.238
FT090	B1K-22-06	IPK	ISR	wild	admixed	0.391	0.609
FT139	B1K-32-09	IPK	ISR	wild	admixed	0.3414	0.6586
FT141	B1K-32-17	IPK	ISR	wild	admixed	0.2815	0.7185
FT184	B1K-42-17	IPK	ISR	wild	admixed	0.2494	0.7506
FT190	B1K-44-05	IPK	ISR	wild	admixed	0.3448	0.6552
FT230	HID-3	IPK	IRQ	wild	admixed	0.5676	0.4324
FT235	HID-20-1	IPK	IRN	wild	admixed	0.3994	0.6006
FT238	HID-24	IPK	IRN	wild	admixed	0.5164	0.4836
FT239	HID-28	IPK	IRN	wild	admixed	0.2378	0.7622
FT240	HID-30	IPK	IRN	wild	admixed	0.3388	0.6612

FT241	HID-32	IPK	IRN	wild	admixed	0.5217	0.4783
FT244	HID-44	IPK	IRN	wild	admixed	0.5424	0.4576
FT245	HID-45	IPK	IRN	wild	admixed	0.5254	0.4746
FT246	HID-46	IPK	IRN	wild	admixed	0.6945	0.3055
FT247	HID-52	IPK	IRN	wild	admixed	0.6775	0.3225
FT249	HID-54	IPK	TUR	wild	admixed	0.4572	0.5428
FT251	HID-56	IPK	TUR	wild	admixed	0.6634	0.3366
FT266	HID-90	IPK	TUR	wild	admixed	0.2468	0.7532
FT268	HID-99	IPK	SYR	wild	admixed	0.1131	0.8869
FT272	HID-107	IPK	JOR	wild	admixed	0.1006	0.8994
FT273	HID-109	IPK	SYR	wild	admixed	0.1023	0.8977
FT285	HID-143	IPK	IRN	wild	admixed	0.1951	0.8049
FT288	HID-146	IPK	ISR	wild	admixed	0.2429	0.7571
FT294	HID-166	IPK	ISR	wild	admixed	0.1385	0.8615
FT307	HID-193	IPK	ISR	wild	admixed	0.2311	0.7689
FT309	HID-196	IPK	ISR	wild	admixed	0.2333	0.7667
FT312	HID-209	IPK	ISR	wild	admixed	0.3711	0.6289
FT334	HID-257	IPK	ISR	wild	admixed	0.474	0.526
FT339	HID-271	IPK	IRN	wild	admixed	0.3364	0.6636
FT343	HID-280	IPK	TUR	wild	admixed	0.7728	0.2272
FT348	HID-301	IPK	IRN	wild	admixed	0.5229	0.4771
FT354	HID-309	IPK	IRN	wild	admixed	0.8299	0.1701
FT365	HID-328-1	IPK	CHN	wild	admixed	0.6116	0.3884
FT370	HID-339-1	IPK	TKM	wild	admixed	0.3296	0.6704
FT371	HID-342	IPK	ISR	wild	admixed	0.2241	0.7759
FT378	HID-366	IPK	IRN	wild	admixed	0.3415	0.6585
FT383	HID-373-1	IPK	ISR	wild	admixed	0.6395	0.3605
FT384	HID-375-1	IPK	ISR	wild	admixed	0.2595	0.7405
FT385	HID-376-2	IPK	ISR	wild	admixed	0.6369	0.3631
FT386	HID-377-1	IPK	ISR	wild	admixed	0.7559	0.2441
FT387	HID-377-2	IPK	ISR	wild	admixed	0.7599	0.2401
FT513	HID-113	IPK	SYR	wild	admixed	0.1111	0.889
FT228	B1K-70-02	IPK	ISR	cultivar	domesticated	1	0
FT405	Triumph	IPK	CZE	cultivar	domesticated	1	0
FT409	Lyallpur	IPK	PAK	cultivar	domesticated	0.9462	0.0538
FT416	Nudinka	IPK	DEU	cultivar	domesticated	1	0
FT417	Proctor	IPK	GBR	cultivar	domesticated	1	0
FT418	Igri	IPK	DEU	cultivar	domesticated	1	0
FT424	Morex	IPK	USA	cultivar	domesticated	0.9633	0.0367
FT433	Scarlett	IPK	DEU	cultivar	domesticated	1	0
FT440	Keel	IPK	AUS	cultivar	domesticated	0.9994	0.0006
FT441	Flagship	IPK	AUS	cultivar	domesticated	1	0
FT442	Auriga	IPK	DEU	cultivar	domesticated	1	0
FT443	Marthe	IPK	DEU	cultivar	domesticated	1	0
FT444	Barke	IPK	DEU	cultivar	domesticated	1	0
FT445	Optic	IPK	GBR	cultivar	domesticated	1	0
FT477	Anadolu 86	IPK	TUR	cultivar	domesticated	1	0

FT481	Aydanhanim	IPK	TUR	cultivar	domesticated	1	0
FT482	Efes 3	IPK	TUR	cultivar	domesticated	1	0
FT483	Efes 6	IPK	TUR	cultivar	domesticated	1	0
FT484	Cetin 2000	IPK	TUR	cultivar	domesticated	0.9581	0.0419
FT485	Efes-4	IPK	TUR	cultivar	domesticated	1	0
FT486	Efes 2	IPK	TUR	cultivar	domesticated	1	0
FT488	Angora	IPK	TUR	cultivar	domesticated	1	0
FT489	Bülbül 89	IPK	TUR	cultivar	domesticated	1	0
FT492	Tarm 92	IPK	TUR	cultivar	domesticated	1	0
FT493	Anadolu 98	IPK	TUR	cultivar	domesticated	1	0
FT494	Yesevi 93	IPK	TUR	cultivar	domesticated	1	0
FT495	Kalayci 97	IPK	TUR	cultivar	domesticated	1	0
FT496	Yercil 147	IPK	TUR	cultivar	domesticated	1	0
FT497	Efes 98	IPK	TUR	cultivar	domesticated	1	0
FT498	Obruk 86	IPK	TUR	cultivar	domesticated	1	0
FT500	Orza 96	IPK	TUR	cultivar	domesticated	1	0
FT502	Efes 1	IPK	TUR	cultivar	domesticated	1	0
FT503	Erginel 90	IPK	TUR	cultivar	domesticated	0.9595	0.0405
FT504	Hamidiye 85	IPK	TUR	cultivar	domesticated	1	0
FT505	Sladoran	IPK	TUR	cultivar	domesticated	1	0
FT518	Er/Apm	IPK	SYR	cultivar	domesticated	1	0
FT727	Bowman	IPK	USA	cultivar	domesticated	1	0
FT766	Triumph	IPK	DEU	cultivar	domesticated	1	0
MK033	Antonella	MPIPZ	DEU	cultivar	domesticated	0.9541	0.0459
MK045	Bonus	MPIPZ	SWE	cultivar	domesticated	1	0
MK051	Foma	MPIPZ	GBR	cultivar	domesticated	1	0
S029	Grace	MPIPZ	DEU	cultivar	domesticated	1	0
S058	Quench	MPIPZ	DEU	cultivar	domesticated	1	0
S155	Tatum	MPIPZ	AUT	cultivar	domesticated	1	0
S185	Beatrix	MPIPZ	GBR	cultivar	domesticated	1	0
FT226	B1K-55-06	IPK	ISR	landrace	domesticated	1	0
FT399	G-398 H	IPK	ETH	landrace	domesticated	0.9811	0.0189
FT404	G-408 I	IPK	KEN	landrace	domesticated	1	0
FT407	G-419 I	IPK	CHN	landrace	domesticated	0.9986	0.0014
FT408	G-423 H	IPK	ETH	landrace	domesticated	0.9802	0.0198
FT411	G-428 I	IPK	CHE	landrace	domesticated	1	0
FT414	Siglah	IPK	YEM	landrace	domesticated	0.9923	0.0077
FT451	HOR7985	IPK	TUR	landrace	domesticated	1	0
FT478	Tokak 157/37	IPK	TUR	landrace	domesticated	1	0
FT479	Kral 97	IPK	TUR	landrace	domesticated	0.9434	0.0566
MK004	Mutah	MPIPZ	JOR	landrace	domesticated	0.9623	0.0377
MK035	Yarmouk	MPIPZ	JOR	landrace	domesticated	0.9999	0.0001
FT001	B1K-02-02	IPK	ISR	wild	wild	0	1
FT002	B1K-02-05	IPK	ISR	wild	wild	0	1
FT003	B1K-02-08	IPK	ISR	wild	wild	0	1
FT004	B1K-02-10	IPK	ISR	wild	wild	0.0001	0.9999
FT005	B1K-02-18	IPK	ISR	wild	wild	0	1

FT006	B1K-03-04	IPK	ISR	wild	wild	0	1
FT007	B1K-03-07	IPK	ISR	wild	wild	0	1
FT008	B1K-03-09	IPK	ISR	wild	wild	0	1
FT009	B1K-04-04	IPK	ISR	wild	wild	0	1
FT010	B1K-04-08	IPK	ISR	wild	wild	0.0228	0.9772
FT012	B1K-04-13	IPK	ISR	wild	wild	0	1
FT013	B1K-05-11	IPK	ISR	wild	wild	0.0305	0.9695
FT014	B1K-05-13	IPK	ISR	wild	wild	0.0005	0.9995
FT016	B1K-05-18	IPK	ISR	wild	wild	0.0083	0.9917
FT017	B1K-06-07	IPK	ISR	wild	wild	0.0138	0.9862
FT018	B1K-06-09	IPK	ISR	wild	wild	0.0255	0.9745
FT020	B1K-06-20	IPK	ISR	wild	wild	0.019	0.981
FT021	B1K-07-02	IPK	ISR	wild	wild	0.0397	0.9603
FT023	B1K-07-16	IPK	ISR	wild	wild	0.0189	0.9811
FT024	B1K-07-17	IPK	ISR	wild	wild	0.0292	0.9708
FT025	B1K-08-06	IPK	ISR	wild	wild	0.0033	0.9967
FT026	B1K-08-07	IPK	ISR	wild	wild	0.0043	0.9957
FT027	B1K-08-10	IPK	ISR	wild	wild	0.0439	0.9561
FT028	B1K-08-13	IPK	ISR	wild	wild	0.0174	0.9826
FT029	B1K-08-18	IPK	ISR	wild	wild	0.0248	0.9752
FT030	B1K-09-03	IPK	ISR	wild	wild	0.0008	0.9992
FT031	B1K-09-07	IPK	ISR	wild	wild	0.0085	0.9915
FT032	B1K-09-10	IPK	ISR	wild	wild	0	1
FT033	B1K-09-16	IPK	ISR	wild	wild	0	1
FT034	B1K-09-18	IPK	ISR	wild	wild	0.0101	0.9899
FT036	B1K-10-05	IPK	ISR	wild	wild	0	1
FT037	B1K-10-10	IPK	ISR	wild	wild	0	1
FT038	B1K-10-13	IPK	ISR	wild	wild	0	1
FT039	B1K-10-18	IPK	ISR	wild	wild	0	1
FT041	B1K-11-05	IPK	ISR	wild	wild	0	1
FT042	B1K-11-11	IPK	ISR	wild	wild	0	1
FT044	B1K-12-01	IPK	ISR	wild	wild	0	1
FT045	B1K-12-06	IPK	ISR	wild	wild	0.0017	0.9983
FT046	B1K-12-10	IPK	ISR	wild	wild	0.0035	0.9965
FT047	B1K-12-17	IPK	ISR	wild	wild	0.0039	0.9961
FT048	B1K-13-01	IPK	ISR	wild	wild	0.0293	0.9707
FT049	B1K-13-06	IPK	ISR	wild	wild	0.0426	0.9574
FT050	B1K-13-10	IPK	ISR	wild	wild	0	1
FT051	B1K-13-14	IPK	ISR	wild	wild	0.0289	0.9711
FT052	B1K-13-20	IPK	ISR	wild	wild	0.0095	0.9905
FT053	B1K-14-04	IPK	ISR	wild	wild	0.0138	0.9862
FT054	B1K-14-06	IPK	ISR	wild	wild	0.0365	0.9635
FT056	B1K-14-13	IPK	ISR	wild	wild	0.0327	0.9673
FT057	B1K-15-07	IPK	ISR	wild	wild	0.0347	0.9653
FT058	B1K-15-12	IPK	ISR	wild	wild	0.0501	0.9499
FT059	B1K-15-15	IPK	ISR	wild	wild	0.0321	0.9679
FT060	B1K-15-19	IPK	ISR	wild	wild	0.0211	0.9789

FT061	B1K-16-01	IPK	ISR	wild	wild	0.0291	0.9709
FT063	B1K-16-10	IPK	ISR	wild	wild	0.0302	0.9698
FT064	B1K-16-15	IPK	ISR	wild	wild	0.0128	0.9872
FT065	B1K-16-19	IPK	ISR	wild	wild	0	1
FT066	B1K-17-03	IPK	ISR	wild	wild	0	1
FT067	B1K-17-07	IPK	ISR	wild	wild	0	1
FT068	B1K-17-10	IPK	ISR	wild	wild	0	1
FT069	B1K-17-13	IPK	ISR	wild	wild	0	1
FT070	B1K-17-17	IPK	ISR	wild	wild	0	1
FT071	B1K-18-03	IPK	ISR	wild	wild	0	1
FT072	B1K-18-05	IPK	ISR	wild	wild	0	1
FT073	B1K-18-09	IPK	ISR	wild	wild	0.0042	0.9958
FT074	B1K-18-16	IPK	ISR	wild	wild	0.013	0.987
FT075	B1K-18-17	IPK	ISR	wild	wild	0.0018	0.9982
FT076	B1K-19-01	IPK	ISR	wild	wild	0.0051	0.9949
FT077	B1K-19-08	IPK	ISR	wild	wild	0.0001	0.9999
FT078	B1K-19-12	IPK	ISR	wild	wild	0	1
FT079	B1K-19-15	IPK	ISR	wild	wild	0	1
FT080	B1K-19-20	IPK	ISR	wild	wild	0	1
FT081	B1K-20-02	IPK	ISR	wild	wild	0	1
FT082	B1K-20-10	IPK	ISR	wild	wild	0	1
FT083	B1K-20-13	IPK	ISR	wild	wild	0	1
FT084	B1K-20-17	IPK	ISR	wild	wild	0.0004	0.9996
FT085	B1K-21-02	IPK	ISR	wild	wild	0	1
FT086	B1K-21-08	IPK	ISR	wild	wild	0	1
FT087	B1K-21-11	IPK	ISR	wild	wild	0	1
FT088	B1K-21-20	IPK	ISR	wild	wild	0	1
FT089	B1K-22-03	IPK	ISR	wild	wild	0.0026	0.9974
FT091	B1K-22-10	IPK	ISR	wild	wild	0.098	0.902
FT092	B1K-22-16	IPK	ISR	wild	wild	0.0573	0.9427
FT093	B1K-22-18	IPK	ISR	wild	wild	0	1
FT094	B1K-23-02	IPK	ISR	wild	wild	0	1
FT095	B1K-23-06	IPK	ISR	wild	wild	0	1
FT097	B1K-23-14	IPK	ISR	wild	wild	0	1
FT098	B1K-24-02	IPK	ISR	wild	wild	0	1
FT099	B1K-24-05	IPK	ISR	wild	wild	0.0114	0.9886
FT100	B1K-24-10	IPK	ISR	wild	wild	0	1
FT102	B1K-24-18	IPK	ISR	wild	wild	0.0001	0.9999
FT103	B1K-25-01	IPK	ISR	wild	wild	0	1
FT104	B1K-25-05	IPK	ISR	wild	wild	0.0005	0.9995
FT105	B1K-25-09	IPK	ISR	wild	wild	0.0104	0.9896
FT108	B1K-26-04	IPK	ISR	wild	wild	0.0662	0.9338
FT110	B1K-26-11	IPK	ISR	wild	wild	0.0149	0.9851
FT112	B1K-26-17	IPK	ISR	wild	wild	0.0117	0.9883
FT113	B1K-27-02	IPK	ISR	wild	wild	0	1
FT114	B1K-27-08	IPK	ISR	wild	wild	0	1
FT115	B1K-27-11	IPK	ISR	wild	wild	0	1

FT116	B1K-27-15	IPK	ISR	wild	wild	0	1
FT117	B1K-27-19	IPK	ISR	wild	wild	0	1
FT118	B1K-28-04	IPK	ISR	wild	wild	0	1
FT119	B1K-28-06	IPK	ISR	wild	wild	0	1
FT120	B1K-28-10	IPK	ISR	wild	wild	0	1
FT121	B1K-28-15	IPK	ISR	wild	wild	0	1
FT122	B1K-28-17	IPK	ISR	wild	wild	0	1
FT123	B1K-29-04	IPK	ISR	wild	wild	0	1
FT124	B1K-29-13	IPK	ISR	wild	wild	0	1
FT125	B1K-29-20	IPK	ISR	wild	wild	0	1
FT126	B1K-30-03	IPK	ISR	wild	wild	0	1
FT127	B1K-30-07	IPK	ISR	wild	wild	0	1
FT128	B1K-30-09	IPK	ISR	wild	wild	0	1
FT129	B1K-30-11	IPK	ISR	wild	wild	0	1
FT130	B1K-30-13	IPK	ISR	wild	wild	0	1
FT131	B1K-30-20	IPK	ISR	wild	wild	0	1
FT132	B1K-31-01	IPK	ISR	wild	wild	0	1
FT133	B1K-31-05	IPK	ISR	wild	wild	0	1
FT134	B1K-31-11	IPK	ISR	wild	wild	0.0004	0.9996
FT135	B1K-31-15	IPK	ISR	wild	wild	0	1
FT136	B1K-31-19	IPK	ISR	wild	wild	0	1
FT138	B1K-32-08	IPK	ISR	wild	wild	0.0419	0.9581
FT140	B1K-32-13	IPK	ISR	wild	wild	0.0215	0.9785
FT142	B1K-33-03	IPK	ISR	wild	wild	0	1
FT145	B1K-33-19	IPK	ISR	wild	wild	0.0171	0.9829
FT146	B1K-34-04	IPK	ISR	wild	wild	0	1
FT148	B1K-34-09	IPK	ISR	wild	wild	0	1
FT149	B1K-34-10	IPK	ISR	wild	wild	0.0001	0.9999
FT150	B1K-34-14	IPK	ISR	wild	wild	0	1
FT151	B1K-34-20	IPK	ISR	wild	wild	0.0474	0.9526
FT152	B1K-35-04	IPK	ISR	wild	wild	0	1
FT153	B1K-35-08	IPK	ISR	wild	wild	0	1
FT154	B1K-35-11	IPK	ISR	wild	wild	0	1
FT155	B1K-35-16	IPK	ISR	wild	wild	0	1
FT156	B1K-35-19	IPK	ISR	wild	wild	0	1
FT158	B1K-36-05	IPK	ISR	wild	wild	0	1
FT159	B1K-36-15	IPK	ISR	wild	wild	0	1
FT160	B1K-36-20	IPK	ISR	wild	wild	0	1
FT161	B1K-37-03	IPK	ISR	wild	wild	0	1
FT163	B1K-37-11	IPK	ISR	wild	wild	0.0129	0.9871
FT164	B1K-37-15	IPK	ISR	wild	wild	0	1
FT166	B1K-38-01	IPK	ISR	wild	wild	0	1
FT167	B1K-38-12	IPK	ISR	wild	wild	0.0227	0.9773
FT168	B1K-38-14	IPK	ISR	wild	wild	0.0124	0.9876
FT169	B1K-39-02	IPK	ISR	wild	wild	0.0561	0.9439
FT170	B1K-39-20	IPK	ISR	wild	wild	0.0477	0.9523
FT171	B1K-40-02	IPK	ISR	wild	wild	0	1

FT172	B1K-40-08	IPK	ISR	wild	wild	0.0134	0.9866
FT173	B1K-40-16	IPK	ISR	wild	wild	0.0177	0.9823
FT174	B1K-41-03	IPK	ISR	wild	wild	0	1
FT175	B1K-41-08	IPK	ISR	wild	wild	0.0025	0.9975
FT176	B1K-41-11	IPK	ISR	wild	wild	0	1
FT177	B1K-41-13	IPK	ISR	wild	wild	0	1
FT178	B1K-41-18	IPK	ISR	wild	wild	0.0154	0.9846
FT179	B1K-42-01	IPK	ISR	wild	wild	0.0073	0.9927
FT181	B1K-42-07	IPK	ISR	wild	wild	0.0327	0.9673
FT182	B1K-42-10	IPK	ISR	wild	wild	0	1
FT183	B1K-42-16	IPK	ISR	wild	wild	0.0108	0.9892
FT185	B1K-43-01	IPK	ISR	wild	wild	0.0054	0.9946
FT187	B1K-43-14	IPK	ISR	wild	wild	0	1
FT188	B1K-43-19	IPK	ISR	wild	wild	0	1
FT189	B1K-44-03	IPK	ISR	wild	wild	0.0001	0.9999
FT191	B1K-44-07	IPK	ISR	wild	wild	0	1
FT194	B1K-44-20	IPK	ISR	wild	wild	0	1
FT195	B1K-45-02	IPK	ISR	wild	wild	0	1
FT196	B1K-45-08	IPK	ISR	wild	wild	0	1
FT198	B1K-45-14	IPK	ISR	wild	wild	0	1
FT199	B1K-46-03	IPK	ISR	wild	wild	0	1
FT200	B1K-46-07	IPK	ISR	wild	wild	0	1
FT202	B1K-46-15	IPK	ISR	wild	wild	0	1
FT203	B1K-46-20	IPK	ISR	wild	wild	0	1
FT204	B1K-47-04	IPK	ISR	wild	wild	0	1
FT205	B1K-47-08	IPK	ISR	wild	wild	0	1
FT207	B1K-47-13	IPK	ISR	wild	wild	0	1
FT208	B1K-47-20	IPK	ISR	wild	wild	0	1
FT209	B1K-48-03	IPK	ISR	wild	wild	0	1
FT211	B1K-48-11	IPK	ISR	wild	wild	0	1
FT212	B1K-48-16	IPK	ISR	wild	wild	0	1
FT213	B1K-48-19	IPK	ISR	wild	wild	0	1
FT214	B1K-49-04	IPK	ISR	wild	wild	0.0005	0.9995
FT215	B1K-49-08	IPK	ISR	wild	wild	0.03	0.97
FT217	B1K-49-13	IPK	ISR	wild	wild	0	1
FT218	B1K-49-18	IPK	ISR	wild	wild	0	1
FT219	B1K-50-08	IPK	ISR	wild	wild	0.0548	0.9452
FT220	B1K-50-14	IPK	ISR	wild	wild	0.0441	0.9559
FT221	B1K-51-01	IPK	ISR	wild	wild	0.0003	0.9997
FT222	B1K-51-07	IPK	ISR	wild	wild	0.019	0.981
FT229	HID-1	IPK	IRQ	wild	wild	0	1
FT231	HID-4	IPK	IRQ	wild	wild	0.013	0.987
FT232	HID-6	IPK	IRQ	wild	wild	0.0404	0.9596
FT233	HID-8-1	IPK	IRQ	wild	wild	0.013	0.987
FT234	HID-10	IPK	IRQ	wild	wild	0.0405	0.9595
FT236	HID-21	IPK	IRN	wild	wild	0.0002	0.9998
FT242	HID-39	IPK	IRN	wild	wild	0.054	0.946

FT243	HID-40	IPK	IRN	wild	wild	0.0555	0.9445
FT248	HID-53	IPK	IRN	wild	wild	0.0341	0.9659
FT250	HID-55	IPK	TUR	wild	wild	0	1
FT252	HID-57	IPK	TUR	wild	wild	0.0464	0.9536
FT254	HID-60	IPK	TUR	wild	wild	0.0238	0.9762
FT255	HID-61	IPK	TUR	wild	wild	0.0727	0.9273
FT256	HID-65	IPK	TUR	wild	wild	0.0203	0.9797
FT260	HID-70	IPK	TUR	wild	wild	0.0213	0.9787
FT262	HID-84	IPK	TUR	wild	wild	0.0841	0.9159
FT263	HID-85	IPK	TUR	wild	wild	0.0319	0.9681
FT267	HID-96	IPK	JOR	wild	wild	0.0839	0.9161
FT270	HID-101	IPK	SYR	wild	wild	0.0729	0.9271
FT271	HID-104	IPK	SYR	wild	wild	0.0246	0.9754
FT274	HID-114	IPK	LBN	wild	wild	0	1
FT275	HID-122	IPK	JOR	wild	wild	0.0778	0.9222
FT279	HID-133	IPK	AFG	wild	wild	0.0473	0.9527
FT280	HID-135	IPK	AFG	wild	wild	0.0479	0.9521
FT281	HID-137	IPK	TUR	wild	wild	0.0207	0.9793
FT282	HID-138	IPK	IRN	wild	wild	0.0185	0.9815
FT283	HID-140	IPK	IRQ	wild	wild	0.0254	0.9746
FT284	HID-142	IPK	IRN	wild	wild	0.0013	0.9987
FT286	HID-144	IPK	IRN	wild	wild	0.0298	0.9702
FT287	HID-145	IPK	ISR	wild	wild	0	1
FT289	HID-149	IPK	ISR	wild	wild	0.0184	0.9816
FT292	HID-157	IPK	ISR	wild	wild	0	1
FT293	HID-160	IPK	ISR	wild	wild	0.0458	0.9542
FT295	HID-168	IPK	ISR	wild	wild	0	1
FT296	HID-169	IPK	ISR	wild	wild	0.0747	0.9253
FT297	HID-174	IPK	ISR	wild	wild	0	1
FT298	HID-175	IPK	ISR	wild	wild	0	1
FT299	HID-176	IPK	ISR	wild	wild	0	1
FT300	HID-180	IPK	ISR	wild	wild	0	1
FT301	HID-182	IPK	ISR	wild	wild	0.0895	0.9105
FT302	HID-183	IPK	ISR	wild	wild	0.0006	0.9994
FT303	HID-184	IPK	ISR	wild	wild	0.0001	0.9999
FT310	HID-198	IPK	ISR	wild	wild	0	1
FT311	HID-202	IPK	ISR	wild	wild	0	1
FT313	HID-213	IPK	ISR	wild	wild	0	1
FT314	HID-215	IPK	ISR	wild	wild	0	1
FT315	HID-217	IPK	ISR	wild	wild	0	1
FT316	HID-218	IPK	ISR	wild	wild	0	1
FT317	HID-222	IPK	ISR	wild	wild	0.0095	0.9905
FT318	HID-224	IPK	ISR	wild	wild	0.0002	0.9998
FT320	HID-230	IPK	ISR	wild	wild	0	1
FT321	HID-233	IPK	ISR	wild	wild	0	1
FT322	HID-236	IPK	ISR	wild	wild	0	1
FT323	HID-237	IPK	ISR	wild	wild	0	1

FT324	HID-239	IPK	ISR	wild	wild	0.031	0.969
FT325	HID-242	IPK	ISR	wild	wild	0	1
FT326	HID-243	IPK	ISR	wild	wild	0.0262	0.9738
FT327	HID-244	IPK	ISR	wild	wild	0	1
FT328	HID-245	IPK	ISR	wild	wild	0	1
FT329	HID-246	IPK	ISR	wild	wild	0.0001	0.9999
FT331	HID-249	IPK	IRN	wild	wild	0.0493	0.9507
FT332	HID-250	IPK	IRN	wild	wild	0.0411	0.9589
FT335	HID-258	IPK	ISR	wild	wild	0.0099	0.9901
FT336	HID-262	IPK	ISR	wild	wild	0	1
FT337	HID-263	IPK	ISR	wild	wild	0	1
FT338	HID-264	IPK	ISR	wild	wild	0.0225	0.9775
FT340	HID-272	IPK	ISR	wild	wild	0	1
FT341	HID-273	IPK	ISR	wild	wild	0	1
FT342	HID-277	IPK	ISR	wild	wild	0.008	0.992
FT344	HID-290	IPK	IRN	wild	wild	0.0331	0.9669
FT345	HID-291	IPK	IRN	wild	wild	0.0354	0.9646
FT346	HID-297	IPK	IRN	wild	wild	0.041	0.959
FT347	HID-299	IPK	IRN	wild	wild	0.0178	0.9822
FT349	HID-302	IPK	IRN	wild	wild	0.0222	0.9778
FT350	HID-304	IPK	IRN	wild	wild	0.0473	0.9527
FT352	HID-307	IPK	IRN	wild	wild	0.0147	0.9853
FT353	HID-308	IPK	IRN	wild	wild	0.0395	0.9605
FT355	HID-310	IPK	IRN	wild	wild	0.0381	0.9619
FT372	HID-343	IPK	ISR	wild	wild	0.0054	0.9946
FT373	HID-355	IPK	TUR	wild	wild	0.0155	0.9845
FT376	HID-360	IPK	TKM	wild	wild	0.0698	0.9302
FT377	HID-361	IPK	ISR	wild	wild	0	1
FT381	HID-369	IPK	ISR	wild	wild	0.0537	0.9463
FT382	HID-372	IPK	ISR	wild	wild	0	1
FT389	HID-382	IPK	CHN	wild	wild	0.0346	0.9654
FT394	HID-386	IPK	ISR	wild	wild	0	1
FT446	HID-2	IPK	IRQ	wild	wild	0.012	0.988
FT452	HP-02-3	IPK	TUR	wild	wild	0.066	0.934
FT453	HP-03-2	IPK	TUR	wild	wild	0.0851	0.9149
FT454	HP-10-4	IPK	TUR	wild	wild	0.0512	0.9488
FT455	HP-10-5	IPK	TUR	wild	wild	0.0384	0.9616
FT456	HP-11-1	IPK	TUR	wild	wild	0.0714	0.9286
FT457	HP-13-2	IPK	TUR	wild	wild	0.0497	0.9503
FT458	HP-15-1	IPK	TUR	wild	wild	0.0611	0.9389
FT460	HP-15-3	IPK	TUR	wild	wild	0.0242	0.9758
FT461	HP-15-5	IPK	TUR	wild	wild	0.0609	0.9391
FT462	HP-20-1	IPK	TUR	wild	wild	0.0497	0.9503
FT464	HP-24-1	IPK	TUR	wild	wild	0.0239	0.9761
FT466	HP-26-1	IPK	TUR	wild	wild	0.0246	0.9754
FT468	HP-27-2	IPK	TUR	wild	wild	0.0309	0.9691
FT469	HP-28-6	IPK	TUR	wild	wild	0.024	0.976

FT470	HP-31-2	IPK	TUR	wild	wild	0.039	0.961
FT471	HP-34-3	IPK	TUR	wild	wild	0.0832	0.9168
FT472	HP-37-2	IPK	TUR	wild	wild	0.0723	0.9277
FT473	HP-38-1	IPK	TUR	wild	wild	0.0609	0.9391
FT475	HP-41-2	IPK	TUR	wild	wild	0.0669	0.9331
FT506	HP-11-7	IPK	TUR	wild	wild	0.0908	0.9092
FT507	HID-7	IPK	IRQ	wild	wild	0.0168	0.9832
FT509	HID-35	IPK	IRN	wild	wild	0.0357	0.9643
FT510	HID-49	IPK	IRN	wild	wild	0.0621	0.9379
FT511	HID-95	IPK	JOR	wild	wild	0.0609	0.9391
FT512	HID-102	IPK	SYR	wild	wild	0.0746	0.9254
FT515	HID-132	IPK	IRQ	wild	wild	0.0184	0.9816
FT516	HID-134	IPK	AFG	wild	wild	0.0502	0.9498
FT517	HID-136	IPK	IRN	wild	wild	0	1
FT620	HID-64-1	IPK	TUR	wild	wild	0.0168	0.9832

* IPK – Gatersleben, MPIPZ – Cologne, NBRP – Okayama; ** - n.d., no data

*Table S4. Domestication-related candidate genes**

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
seq420	AK370055	1	38.031	-	-	-3.01	-	-	-	-	main
-	morex_contig_275920	1	46.176	-	-	-2.85	-	-	-	-	random
-	morex_contig_274229	1	47.663	-	-	-3.39	-	-	-	8.5	off_target
-	morex_contig_2561616	1	47.805	-	-	-	-	-	34.1	-	off_target
-	morex_contig_1561779	1	47.946	-	-	-3.11	-	-	-	0.9	off_target
seq162	morex_contig_71104_RC	1	47.946	-	-	-5.04	-	-	-	-	main
seq154	morex_contig_114046_RC	1	48.088	-	-	-3.42	-	-	-	2.2	main
seq262	morex_contig_136422_RC	1	48.088	-	-	-	-	-	91.8	-	main
-	morex_contig_146698	1	48.088	-	-	-	-	-	36.0	-	off_target
seq564	morex_contig_1558833	1	48.088	-	-	-4.84	-	-	-	6.5	main
seq231	bowman_contig_12278_RC	1	48.159	-	-	-	-	-	65.8	-	main
seq612	FJ951828	1	48.229	-	-	-	-	-	104.4	-	main
-	morex_contig_2554832	1	48.548	-	-	-	0.46	-	-	0.7	random
-	morex_contig_232695	1	50.850	-	-	-3.16	-	-	-	0.6	random
-	morex_contig_1627094	1	54.037	-	-	-	-	-	26.1	-	off_target
-	morex_contig_244506	1	54.391	-	-	-2.85	-	-	-	2.2	random
seq290	morex_contig_135896	1	54.887	-	-	-3.98	-	-	-	-	main
seq69	morex_contig_137029_RC	1	55.524	-	-	-3.82	-	-	-	-	main
-	morex_contig_47916	1	59.136	-	-	-2.78	-	-	-	-	off_target
-	morex_contig_50751	1	61.473	-	FWH_feat1_1	-	0.36	-	-	-	off_target
-	morex_contig_56703	1	61.827	-	FWH_feat1_1	-	0.18	-	-	-	off_target
-	morex_contig_1558870	1	62.323	-	FWH_feat1_1	-3.06	-	-	-	-	off_target

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
-	morex_contig_274252	1	62.323	-	FWH_feat1_1	-	-2.62	-	-	-	random
-	morex_contig_41303	1	62.689	-	FWH_feat1_1	-	-0.16	-	-	-	random
-	morex_contig_1590025	1	65.227	-	FWH_feat1_1	-3.82	-	-	-	-	random
-	morex_contig_46847	1	65.460	-	FWH_feat1_1	-	-1.83	-	-	-	random
-	morex_contig_215022	1	66.289	-	FWH_feat1_1	-	0.09	-	-	-	off_target
-	morex_contig_7000	1	67.918	dLD_feat1_1	FWH_feat1_1	-3.10	-	-	-	-	off_target
-	morex_contig_6243	1	70.538	dLD_feat1_1	FWH_feat1_1	-	0.46	-	-	-	random
-	morex_contig_2547611	1	71.105	dLD_feat1_1	FWH_feat1_1	-2.96	-	-	-	-	off_target
seq93	morex_contig_137689_RC	1	72.521	dLD_feat1_1	-	-	0.15	-	-	-	main
-	morex_contig_162791	1	81.020	dLD_feat1_1	-	-	0.34	NPI_feat2_1	-	-	off_target
-	morex_contig_39910	1	81.020	dLD_feat1_1	-	-4.16	-	NPI_feat2_1	-	-	off_target
-	morex_contig_36870	1	82.507	-	-	-	0.22	NPI_feat2_1	-	-	random
-	morex_contig_5702	1	85.637	-	-	-	0.15	NPI_feat2_1	-	-	random
seq322	morex_contig_132235_RC	1	87.875	-	-	-4.69	-	-	-	-	main
seq392	AK368116	1	90.297	-	-	-3.58	-	-	-	-	main
-	morex_contig_60897	1	109.424	dLD_feat2_1	-	-	0.62	-	-	-	off_target
seq133	AK248238	1	116.785	dLD_feat2_1	-	-	-	-	-	-	main
seq232	morex_contig_159925_RC	1	119.122	-	-	-	-	NPI_feat3_1	-	11.2	main
-	morex_contig_38275	1	121.955	-	-	-	-	NPI_feat3_1	-	12.5	off_target
seq108	morex_contig_37815	1	125.082	-	-	-	-	NPI_feat3_1	122.1	-	main
-	morex_contig_40090	1	126.133	-	-	-	-	NPI_feat3_1	-	3.3	random

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
seq620	morex_contig_46658_RC	1	127.089	-	-	-	-	NPI_feat3_1	-	2.4	main
-	morex_contig_57219	1	128.329	-	-	-	0.85	NPI_feat3_1	-	-	off_target
-	morex_contig_1577346	2	2.195	-	-	-3.07	-	-	-	-	off_target
-	morex_contig_42823	2	7.436	-	-	-3.61	-	-	-	-	random
seq358	morex_contig_50954	2	7.436	-	-	-3.45	-	-	-	-	main
-	morex_contig_47943	2	13.314	-	-	-3.37	-	-	-	-	random
-	morex_contig_160155	2	33.499	-	FWH_feat1_2	-4.85	-	-	-	-	random
-	morex_contig_50566	2	33.676	-	FWH_feat1_2	-	0.60	-	-	-	off_target
seq377	morex_contig_50132	2	37.448	-	FWH_feat1_2	-	-0.95	-	-	-	main
-	morex_contig_66114	2	38.102	-	FWH_feat1_2	-3.83	-	-	-	-	random
-	morex_contig_2551437	2	39.660	-	FWH_feat1_2	-	0.22	-	-	-	random
-	morex_contig_42659	2	50.354	-	-	-2.90	-	-	-	-	random
seq135	AK248677	2	50.921	-	-	-3.91	-	-	-	-	main
seq216	AK356047	2	52.904	-	-	-3.47	-	-	-	3.8	main
-	morex_contig_94758	2	52.904	-	-	-	0.09	-	32.9	-	random
-	morex_contig_136786	2	54.320	-	-	-3.03	-	-	-	7.2	random
-	morex_contig_49347	2	57.436	-	-	-2.91	-	-	-	-	off_target
-	morex_contig_45930	2	57.861	-	-	-4.85	-	-	-	-	random
seq273	AK360611	2	58.003	-	FWH_feat2_2	-3.37	-	-	-	-	main
seq617	morex_contig_135953	2	58.003	-	FWH_feat2_2	-	-2.53	-	-	-	main
-	morex_contig_303909	2	58.003	-	FWH_feat2_2	-	0.29	-	-	-	off_target
-	morex_contig_62507	2	58.003	-	FWH_feat2_2	-2.87	-	-	-	-	random
seq165	AK251264	2	58.052	-	FWH_feat2_	-	-	-	-	-	main

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
					2						
seq234	AK357417	2	58.052	-	FWH_feat2_2	-3.34	-	-	-	-	main
seq264	AK359655	2	58.052	-	FWH_feat2_2	-	-	-	-	0.8	main
seq471	AK373369	2	58.052	-	FWH_feat2_2	-	0.09	-	33.0	-	main
-	morex_contig_103341	2	58.052	-	FWH_feat2_2	-	-	-	-	3.3	off_target
-	morex_contig_115707	2	58.052	-	FWH_feat2_2	-	-	-	-	3.5	off_target
-	morex_contig_127738	2	58.052	-	FWH_feat2_2	-	-	-	31.6	-	off_target
-	morex_contig_1558082	2	58.052	-	FWH_feat2_2	-	-	-	-	4.9	off_target
seq626	morex_contig_1569988	2	58.052	-	FWH_feat2_2	-	0.87	-	-	-	main
-	morex_contig_1602950	2	58.052	-	FWH_feat2_2	-	-	-	-	5.3	off_target
-	morex_contig_1607359	2	58.052	-	FWH_feat2_2	-	-	-	43.2	-	off_target
-	morex_contig_2559923	2	58.052	-	FWH_feat2_2	-	-	-	-	2.1	off_target
-	morex_contig_50273	2	58.052	-	FWH_feat2_2	-	0.10	-	-	-	random
seq345	morex_contig_54264	2	58.052	-	FWH_feat2_2	-5.32	-	-	-	-	main
-	morex_contig_55735	2	58.052	-	FWH_feat2_2	-	-0.37	-	-	-	random
-	morex_contig_57606	2	58.052	-	FWH_feat2_2	-	-	-	-	8.2	random
-	morex_contig_60566	2	58.052	-	FWH_feat2_2	-	-	-	34.7	-	off_target
-	morex_contig_66012	2	58.052	-	FWH_feat2_2	-	-	-	30.3	-	off_target
-	morex_contig_7586	2	58.052	-	FWH_feat2_2	-3.30	-	-	-	3.5	off_target

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
seq404	bowman_contig_928173	2	58.074	-	FWH_feat2_2	-	-	-	-	-	main
seq404	morex_contig_1566721	2	58.074	-	FWH_feat2_2	-4.06	-	-	-	-	main
-	morex_contig_1576590	2	58.074	-	FWH_feat2_2	-	0.68	-	-	-	off_target
-	morex_contig_368529	2	58.074	-	FWH_feat2_2	-7.32	-	-	-	6.0	random
seq631	morex_contig_37282_RC	2	58.074	-	FWH_feat2_2	-5.03	-	-	34.2	-	main
-	morex_contig_40851	2	58.074	-	FWH_feat2_2	-	-	-	-	3.7	random
seq466	AK373114	2	58.640	-	FWH_feat2_2	-	-	-	-	-	main
seq551	BE216682	2	58.640	-	FWH_feat2_2	-	-	-	-	-	main
seq553	BE421563	2	58.640	-	FWH_feat2_2	-	-	-	-	-	main
-	morex_contig_44867	2	58.640	-	FWH_feat2_2	-	-2.18	-	-	-	random
-	morex_contig_44918	2	58.640	-	FWH_feat2_2	-4.85	-	-	-	-	random
seq148	morex_contig_45974	2	58.640	-	FWH_feat2_2	-	-2.11	-	-	-	main
seq150	morex_contig_134591_RC	2	58.782	-	FWH_feat2_2	-	-1.87	-	-	-	main
seq625	morex_contig_1564635	2	58.782	-	FWH_feat2_2	-	-	-	-	-	main
seq671	morex_contig_41142	2	58.782	-	FWH_feat2_2	-	0.27	-	-	-	main
seq132	morex_contig_41327	2	58.782	-	FWH_feat2_2	-	0.43	-	-	3.7	main
seq530	morex_contig_45976	2	58.782	-	FWH_feat2_2	-	0.09	-	92.8	-	main
-	morex_contig_46212	2	58.782	-	FWH_feat2_2	-	-	-	-	1.4	off_target
seq354	morex_contig_51348_RC	2	58.782	-	FWH_feat2_	-	0.64	-	-	-	main

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
					2						
-	morex_contig_63558	2	58.782	-	FWH_feat2_2	-4.80	-	-	-	-	off_target
-	morex_contig_74891	2	58.782	-	FWH_feat2_2	-4.99	-	-	-	2.4	off_target
-	morex_contig_78457	2	58.782	-	FWH_feat2_2	-	-	-	32.6	-	off_target
-	morex_contig_83966	2	58.782	-	FWH_feat2_2	-4.85	-	-	-	8.8	off_target
-	morex_contig_164937	2	58.853	-	FWH_feat2_2	-	-	-	38.7	-	off_target
seq450	barke_contig_268048	2	58.924	-	FWH_feat2_2	-	-0.05	-	-	1.2	main
seq555	BE422088	2	58.924	-	FWH_feat2_2	-	-	-	-	-	main
-	morex_contig_2549453	2	58.924	-	FWH_feat2_2	-	-0.53	-	-	-	random
-	morex_contig_38980	2	58.924	-	FWH_feat2_2	-3.11	-	-	-	-	off_target
-	morex_contig_68129	2	58.924	-	FWH_feat2_2	-3.58	-	-	-	-	off_target
seq52	DQ201152	2	59.065	-	FWH_feat2_2	-	-2.56	-	-	-	main
seq52	morex_contig_48345	2	59.065	-	FWH_feat2_2	-	-	-	-	-	main
seq12	morex_contig_58354_RC	2	59.207	-	FWH_feat2_2	-	0.48	-	-	-	main
seq174	morex_contig_244168_RC	2	59.348	-	FWH_feat2_2	-	0.18	-	-	8.7	main
seq39	morex_contig_1561605	2	59.419	-	FWH_feat2_2	-5.12	-	-	30.2	-	main
seq101	morex_contig_1567582	2	59.419	-	FWH_feat2_2	-	-	-	40.9	-	main
-	morex_contig_2521771	2	59.419	-	FWH_feat2_2	-	-	-	-	1.2	random
-	morex_contig_48955	2	64.448	-	FWH_feat2_2	-	0.38	-	-	-	off_target

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
seq11	AF490474	2	64.731	-	FWH_feat2_2	-3.82	-	-	-	-	main
-	morex_contig_120279	2	64.731	-	FWH_feat2_2	-3.82	-	-	-	-	off_target
seq11	morex_contig_161048	2	64.731	-	FWH_feat2_2	-	-	-	-	-	main
-	morex_contig_76778	2	67.351	-	FWH_feat2_2	-	0.18	-	-	-	random
-	morex_contig_42088	2	67.493	-	FWH_feat2_2	-4.10	-	-	-	-	off_target
-	morex_contig_1563347	2	70.822	-	FWH_feat2_2	-	-0.07	-	-	-	off_target
-	morex_contig_324710	2	70.822	-	FWH_feat2_2	-4.73	-	-	-	-	off_target
-	morex_contig_42643	2	70.822	-	FWH_feat2_2	-	0.09	-	-	3.1	random
-	morex_contig_64376	2	70.822	-	FWH_feat2_2	-	-	-	29.3	-	random
-	morex_contig_1616200	2	73.725	-	-	-4.04	-	-	-	3.8	off_target
seq223	morex_contig_113731	2	74.150	-	-	-4.14	-	-	-	-	main
-	morex_contig_1561549	2	80.170	dLD_feat1_2	-	-	-1.12	-	-	-	random
seq334	morex_contig_38448	2	81.516	dLD_feat1_2	-	-	-	-	-	-	main
seq513	AK375953	2	81.799	dLD_feat1_2	-	-	-	-	-	-	main
seq493	AK374546	2	82.649	dLD_feat1_2	-	-	-	-	-	-	main
-	morex_contig_41671	2	86.756	dLD_feat1_2	-	-	-	-	-	1.0	random
-	morex_contig_66757	2	98.700	-	-	-	-	-	24.4	-	random
-	morex_contig_59485	2	132.153	-	-	-3.25	-	-	-	3.3	off_target
-	morex_contig_1561585	2	132.578	-	-	-3.24	-	-	-	-	random
seq519	bowman_contig_289632	2	142.635	-	-	-2.92	-	-	-	-	main

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
-	morex_contig_136342	3	20.397	-	-	-3.12	-	-	-	-	off_target
seq294	morex_contig_40951	3	20.397	-	-	-3.16	-	-	-	-	main
-	morex_contig_38184	3	20.574	-	-	-4.85	-	-	-	-	random
-	morex_contig_2547961	3	37.110	-	FWH_feat1_3	-3.00	-	NPI_feat1_3	-	-	random
seq268	morex_contig_122474	3	37.394	-	FWH_feat1_3	-	0.46	NPI_feat1_3	-	6.1	main
seq26	AY740524	3	45.397	-	FWH_feat1_3	-	-0.60	NPI_feat1_3	-	3.0	main
seq55	DQ297407	3	45.550	-	FWH_feat1_3	-	0.18	NPI_feat1_3	-	13.3	main
seq508	morex_contig_137886_RC	3	45.822	-	FWH_feat1_3	-	-1.94	NPI_feat1_3	-	11.7	main
seq518	morex_contig_53805_RC	3	45.822	-	FWH_feat1_3	-	-2.58	NPI_feat1_3	-	16.0	main
seq26	morex_contig_58270	3	45.822	-	FWH_feat1_3	-	-	NPI_feat1_3	-	-	main
-	morex_contig_140607	3	46.034	-	FWH_feat1_3	-	-	NPI_feat1_3	-	11.4	random
seq373	morex_contig_1558090_RC	3	46.034	-	FWH_feat1_3	-4.34	-	NPI_feat1_3	27.6	-	main
-	morex_contig_39604	3	46.034	-	FWH_feat1_3	-	-	NPI_feat1_3	-	1.4	off_target
-	morex_contig_57926	3	46.034	-	FWH_feat1_3	-	-	NPI_feat1_3	24.7	-	off_target
-	morex_contig_67526	3	46.034	-	FWH_feat1_3	-	-	NPI_feat1_3	-	5.2	off_target
seq484	AK373991	3	46.176	-	FWH_feat1_3	-	0.09	NPI_feat1_3	52.3	-	main
-	morex_contig_136090	3	46.176	-	FWH_feat1_3	-	-	NPI_feat1_3	-	2.5	off_target
-	morex_contig_82681	3	46.176	-	FWH_feat1_3	-4.58	-	NPI_feat1_3	-	2.4	random
-	morex_contig_160729	3	46.669	-	FWH_feat1_3	-	-	NPI_feat1_3	46.4	-	off_target
seq357	morex_contig_37290_RC	3	47.096	-	FWH_feat2_	-	0.39	NPI_feat1_3	-	8.8	main

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
					3						
seq664	morex_contig_135958	3	49.292	-	FWH_feat2_3	-	-	NPI_feat1_3	-	2.2	main
-	morex_contig_48795	3	49.646	-	FWH_feat2_3	-	-	NPI_feat1_3	24.7	-	off_target
-	morex_contig_55734	3	49.681	-	FWH_feat2_3	-	-	NPI_feat1_3	-	3.9	off_target
-	morex_contig_1565672	3	49.717	-	FWH_feat2_3	-5.12	-	NPI_feat1_3	-	-	random
-	morex_contig_53487	3	49.717	-	FWH_feat2_3	-	0.39	NPI_feat1_3	-	-	random
seq588	morex_contig_81307	3	50.496	-	FWH_feat2_3	-	-2.45	NPI_feat1_3	-	2.3	main
-	morex_contig_58662	3	50.708	-	FWH_feat2_3	-	-	NPI_feat1_3	-	6.0	random
seq92	morex_contig_136208_RC	3	50.714	-	FWH_feat2_3	-	0.18	NPI_feat1_3	72.2	-	main
-	morex_contig_39346	3	50.921	-	FWH_feat2_3	-	-	NPI_feat1_3	-	1.8	off_target
seq178	morex_contig_135610_RC	3	52.030	-	-	-	-	-	113.2	-	main
-	morex_contig_47947	3	52.030	-	-	-	-	-	63.8	-	off_target
seq159	AK250812	3	53.258	-	-	-	-	-	33.6	-	main
-	morex_contig_2547149	3	56.728	-	-	-4.85	-	-	-	13.9	off_target
seq157	AK250647	3	57.082	-	-	-3.03	-	-	-	-	main
seq107	morex_contig_2550522_RC	3	64.164	dLD_feat1_3	-	-	-	-	-	-	main
-	morex_contig_157281	3	83.924	-	-	-3.20	-	-	-	-	random
seq306	morex_contig_40562	3	90.156	-	-	-4.83	-	-	-	-	main
-	morex_contig_47583	3	98.654	dLD_feat2_3	FWH_feat3_3	-	0.40	NPI_feat2_3	-	-	off_target
seq36	morex_contig_6163	3	98.654	dLD_feat2_3	FWH_feat3_3	-4.35	-	NPI_feat2_3	-	3.3	main
-	morex_contig_2547024	3	98.867	dLD_feat2_3	FWH_feat3_3	-	0.18	NPI_feat2_3	-	-	off_target

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
-	morex_contig_275148	3	104.320	dLD_feat2_3	FWH_feat3_3	-6.51	-	NPI_feat2_3	-	-	random
-	morex_contig_40481	3	105.949	dLD_feat2_3	FWH_feat3_3	-	-2.48	NPI_feat2_3	-	-	random
seq75	morex_contig_51490	3	106.020	dLD_feat3_3	FWH_feat3_3	-	-	NPI_feat3_3	-	3.5	main
seq478	HVVMRX83KhA0070N18_c4_RC	3	108.428	dLD_feat3_3	FWH_feat3_3	-	-0.58	NPI_feat3_3	-	1.6	main
-	morex_contig_1558045	3	108.428	dLD_feat3_3	FWH_feat3_3	-5.12	-	NPI_feat3_3	26.1	-	random
seq618	morex_contig_41521_RC	3	108.428	dLD_feat3_3	FWH_feat3_3	-5.12	-	NPI_feat3_3	-	21.1	main
seq478p	morex_contig_61827	3	108.428	dLD_feat3_3	FWH_feat3_3	-	-	NPI_feat3_3	-	1.5	main
-	morex_contig_59267	3	109.844	dLD_feat4_3	FWH_feat3_3	-	-	NPI_feat3_3	-	3.4	off_target
-	morex_contig_54404	3	112.869	dLD_feat4_3	FWH_feat3_3	-	-	NPI_feat3_3	24.3	-	off_target
seq249	AK358592	3	116.117	-	-	-	-	NPI_feat3_3	-	5.0	main
seq509	morex_contig_56871_RC	3	116.147	-	-	-	-	NPI_feat3_3	-	3.5	main
-	morex_contig_61812	3	116.360	-	-	-	-	NPI_feat3_3	-	2.9	off_target
-	morex_contig_1567942	3	116.638	-	-	-	-	NPI_feat3_3	27.0	-	random
-	morex_contig_60958	3	116.997	-	-	-	-	NPI_feat3_3	-	1.2	random
-	morex_contig_1560000	3	132.932	-	-	-3.12	-	-	-	3.7	off_target
-	morex_contig_2548035	3	137.606	-	-	-	0.09	-	33.0	-	off_target
-	morex_contig_45283	3	145.538	-	-	-	-	NPI_feat4_3	-	1.6	random
seq235	morex_contig_2548416	3	149.002	-	-	-	-	NPI_feat4_3	-	-	main
-	morex_contig_44798	3	149.002	-	-	-	-	NPI_feat4_3	-	1.1	random
seq331	AK364000	3	154.150	-	-	-	-	NPI_feat4_3	36.6	-	main
-	morex_contig_133654	3	154.150	-	-	-	-	NPI_feat4_3	-	6.5	off_target
-	morex_contig_189894	3	154.887	-	-	-	-	NPI_feat4_3	-	8.2	off_target

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
seq390	morex_contig_243551	3	154.887	-	-	-	-	NPI_feat4_3	23.8	-	main
-	morex_contig_43692	3	154.887	-	-	-	-	NPI_feat4_3	-	20.4	random
seq175	morex_contig_40927_RC	4	1.021	dLD_feat1_4	-	-	-	-	-	-	main
seq196	AK354101	4	3.470	dLD_feat1_4	-	-	-	NPI_feat1_4	-	-	main
-	morex_contig_89622	4	4.158	dLD_feat1_4	-	-	-	NPI_feat1_4	-	1.8	off_target
seq197	bowman_contig_15343_RC	4	20.892	-	-	-	-	-	30.4	-	main
-	morex_contig_111505	4	27.479	dLD_feat2_4	-	-	-	-	-	13.3	off_target
seq50	DQ201140	4	34.561	dLD_feat2_4	-	-3.01	-	NPI_feat2_4	-	1.8	main
-	morex_contig_157518	4	35.907	dLD_feat2_4	-	-	0.21	NPI_feat2_4	-	-	off_target
seq583	morex_contig_2547536	4	36.348	dLD_feat2_4	-	-	-	NPI_feat2_4	-	4.1	main
seq338	morex_contig_40563_RC	4	36.348	dLD_feat2_4	-	-	-	NPI_feat2_4	-	-	main
seq520	AK376473	4	39.802	dLD_feat2_4	-	-	-	NPI_feat2_4	-	3.0	main
-	morex_contig_49161	4	40.014	dLD_feat2_4	-	-	-	NPI_feat2_4	-	1.9	random
-	morex_contig_87304	4	41.631	dLD_feat2_4	-	-	-	NPI_feat2_4	59.0	-	off_target
-	morex_contig_88464	4	43.343	-	-	-3.34	-	-	-	1.1	random
seq409	morex_contig_45399	4	43.484	-	-	-4.57	-	-	-	-	main
-	morex_contig_6818	4	43.626	-	-	-4.79	-	-	-	3.7	random
-	morex_contig_60335	4	51.133	-	-	-	-	-	25.1	-	off_target
seq477	AK373558	4	51.275	-	-	-	-	-	24.9	-	main
-	morex_contig_136214	4	51.275	-	-	-	-	-	55.0	-	off_target
-	morex_contig_81640	4	51.275	-	-	-	-	-	92.0	-	off_target
seq406	bowman_contig_847295_RC	4	51.346	-	-	-	-	-	63.3	-	main

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
seq352	AK365283	4	51.404	-	-	-	-	-	32.0	-	main
-	morex_contig_38066	4	51.404	-	-	-	-	-	24.7	-	random
seq660	morex_contig_38516_RC	4	51.404	-	-	-	-	-	66.5	-	main
-	morex_contig_41686	4	51.404	-	-	-	-	-	89.0	-	off_target
seq644	morex_contig_49753	4	51.404	-	-	-4.20	-	-	-	4.0	main
-	morex_contig_55041	4	51.404	-	-	-	-	-	67.3	-	off_target
seq662	morex_contig_55803_RC	4	51.404	-	-	-3.02	-	-	37.4	-	main
seq406	morex_contig_72000	4	51.404	-	-	-	-	-	46.3	-	main
-	morex_contig_97471	4	51.404	-	-	-	-	-	38.5	-	off_target
-	morex_contig_9579	4	51.487	-	-	-4.58	-	-	-	4.5	random
seq283	morex_contig_45564	4	52.195	-	-	-3.14	-	-	-	9.0	main
seq110	morex_contig_38586	4	52.479	-	-	-	0.09	-	62.6	-	main
-	morex_contig_40811	4	58.357	-	-	-4.85	-	-	-	10.2	random
seq147	morex_contig_202661_RC	4	59.703	-	-	-	-	-	48.2	-	main
seq479	AK373704	4	60.907	-	-	-	-	-	27.5	-	main
seq327	morex_contig_137153_RC	4	63.244	-	-	-3.26	-	-	-	2.5	main
-	morex_contig_43347	4	73.654	-	-	-	-	-	43.1	-	off_target
-	morex_contig_55433	4	76.275	-	-	-	-	-	39.2	-	off_target
-	morex_contig_56234	4	80.014	-	-	-	-	NPI_feat3_4	-	2.2	random
seq95	morex_contig_138830_RC	4	80.453	-	-	-	-	NPI_feat3_4	-	3.1	main
seq523	AK376664	4	81.303	-	-	-	0.30	NPI_feat3_4	-	3.5	main
seq155	AK250506	4	81.569	-	-	-2.84	-	NPI_feat3_4	-	3.1	main
-	morex_contig_1562421	4	81.569	-	-	-	0.09	NPI_feat3_4	-	-	off_target
-	morex_contig_340974	4	81.569	-	-	-	0.09	NPI_feat3_4	-	-	random
seq193	morex_contig_60056	4	81.569	-	-	-6.51	-	NPI_feat3_4	-	19.1	main
-	morex_contig_1570633	4	83.640	-	-	-	0.18	NPI_feat3_4	-	-	random

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
-	morex_contig_141143	4	85.127	-	-	-	-1.63	NPI_feat3_4	-	-	off_target
seq667	morex_contig_1565866	4	98.938	-	-	-3.59	-	-	-	-	main
-	morex_contig_262412	4	99.079	-	-	-4.85	-	-	-	-	off_target
seq601	EX586552_MLOC_10435.2	5	34.236	-	-	-3.28	-	-	-	-	main
-	morex_contig_52160	5	34.236	-	-	-4.92	-	-	-	2.6	off_target
seq34	morex_contig_38686	5	43.760	-	-	-	0.53	-	63.6	-	main
-	morex_contig_1563184	5	43.958	-	-	-3.24	-	-	-	1.6	random
-	morex_contig_42673	5	44.097	-	-	-	-	-	73.7	-	random
-	morex_contig_69853	5	44.097	-	-	-2.86	-	-	-	2.4	off_target
seq434	morex_contig_136756	5	44.132	-	-	-	-	-	27.7	-	main
seq311	AK362848	5	48.376	-	-	-2.80	-	-	-	1.7	main
-	morex_contig_54778	5	59.722	-	-	-3.61	-	-	-	-	off_target
-	morex_contig_369156	5	68.297	-	-	-	0.87	NPI_feat1_5	-	1.5	random
seq637	morex_contig_42363_RC	5	69.306	-	-	-2.89	-	NPI_feat1_5	23.9	-	main
-	morex_contig_38623	5	70.139	-	-	-	-	NPI_feat1_5	-	1.1	random
seq194	AK354063	5	72.500	-	-	-	-	NPI_feat1_5	-	1.5	main
seq139	morex_contig_46394_RC	5	75.903	-	-	-	-	NPI_feat1_5	-	1.2	main
-	morex_contig_2555844	5	76.042	-	-	-	0.35	NPI_feat1_5	-	-	random
seq242	morex_contig_7095	5	79.711	-	-	-4.17	-	-	-	-	main
seq64	EU331986	5	95.139	-	-	-3.36	-	-	-	0.7	main
seq192	AK353826	5	99.934	-	-	-	0.09	-	53.8	-	main
-	morex_contig_60535	5	99.934	-	-	-3.08	-	-	-	1.0	off_target
-	morex_contig_138324	5	107.569	dLD_feat1_5	-	-	-2.11	-	-	-	random
-	morex_contig_40909	5	109.861	dLD_feat1_5	-	-	0.28	-	-	-	random
-	morex_contig_167966	5	119.236	-	-	-2.89	-	-	-	-	off_target
seq74	morex_contig_40406_RC	5	128.194	-	FWH_feat1_	-	-1.48	-	-	-	main

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
					5						
seq427	AK370656	5	129.931	-	FWH_feat1_5	-	0.18	-	-	-	main
seq248	AK358516	5	130.694	-	FWH_feat1_5	-	-0.24	-	-	-	main
seq340	AK364247	5	135.208	-	FWH_feat1_5	-	0.09	-	-	1.5	main
seq438	bowman_contig_1989279	5	135.208	-	FWH_feat1_5	-	-0.80	-	37.2	-	main
-	morex_contig_1569081	5	135.208	-	FWH_feat1_5	-	-	-	-	1.7	off_target
-	morex_contig_76908	5	136.389	-	FWH_feat1_5	-	0.09	-	-	-	random
seq442	bowman_contig_142733	5	136.606	-	FWH_feat1_5	-5.52	-	-	-	-	main
seq275	bowman_contig_16180_RC	5	136.806	-	FWH_feat1_5	-3.81	-	-	-	-	main
-	morex_contig_1568661	5	138.958	-	FWH_feat1_5	-	-1.47	-	-	-	random
-	morex_contig_1559810	5	145.694	-	FWH_feat1_5	-	-1.84	-	-	-	random
-	morex_contig_39896	5	155.556	-	-	-3.09	-	-	-	-	off_target
seq517	morex_contig_37036	5	166.319	-	-	-4.58	-	-	-	-	main
seq104	morex_contig_1572260_RC	5	168.889	-	-	-3.11	-	-	-	3.1	main
-	morex_contig_136762	6	49.504	-	-	-	-	-	30.7	-	random
-	morex_contig_1621407	6	52.620	-	-	-	-	-	31.0	-	off_target
seq378	morex_contig_2550116	6	52.620	-	-	-	-	-	35.6	-	main
seq229	AK357081	6	52.762	-	-	-5.10	-	-	-	19.9	main
-	morex_contig_48441	6	52.797	-	-	-	-	-	32.3	-	off_target
seq369	morex_contig_1560606_RC	6	53.602	-	-	-	0.18	-	22.8	-	main
-	morex_contig_1574297	6	53.602	-	-	-4.72	-	-	-	6.5	off_target
-	morex_contig_1565239	6	60.198	-	-	-3.25	-	-	-	-	off_target

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
seq9	AF490469	6	68.201	-	-	-3.95	-	-	-	-	main
seq607	morex_contig_1563686	6	71.671	-	-	-3.72	-	-	-	-	main
seq222	morex_contig_46313_RC	6	87.606	-	-	-3.47	-	-	-	-	main
seq277	barke_contig_298691	6	99.150	-	FWH_feat1_6	-4.29	-	-	-	-	main
-	morex_contig_53594	6	100.850	-	FWH_feat1_6	-	0.11	-	-	-	off_target
seq124	AB063178	6	105.099	-	FWH_feat1_6	-	0.03	-	-	-	main
seq332	morex_contig_135949_RC	6	108.428	-	FWH_feat1_6	-	-2.48	-	-	-	main
-	morex_contig_49997	6	113.244	-	FWH_feat1_6	-	-1.20	-	-	-	random
-	morex_contig_72622	6	115.935	-	FWH_feat1_6	-3.55	-	-	-	-	off_target
-	morex_contig_43201	6	117.918	-	-	-3.39	-	-	-	-	random
seq72	morex_contig_136432_RC	7	1.629	-	-	-3.25	-	-	-	1.0	main
-	morex_contig_6527	7	6.758	-	-	-	-	-	29.3	-	off_target
-	morex_contig_159612	7	13.881	-	-	-	-	-	-	1.1	off_target
-	morex_contig_48495	7	32.790	-	-	-	0.18	-	28.7	-	random
-	morex_contig_41044	7	40.652	-	-	-4.29	-	-	-	0.5	random
-	morex_contig_40281	7	42.493	-	-	-	-	-	23.1	-	random
-	morex_contig_135561	7	48.938	-	-	-	-	-	23.5	-	random
-	morex_contig_177369	7	54.391	-	-	-2.98	-	-	-	0.8	off_target
seq224	AK356695	7	61.756	-	FWH_feat1_7	-	-	-	-	-	main
seq142	morex_contig_43373	7	61.756	-	FWH_feat1_7	-	-0.91	-	-	-	main
seq318	AK363330	7	61.969	-	FWH_feat1_7	-	0.18	-	-	-	main
-	morex_contig_1563042	7	62.110	-	FWH_feat1_7	-	-1.70	-	-	-	off_target

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
-	morex_contig_42237	7	62.110	-	FWH_feat1_7	-3.11	-	-	-	-	random
seq413	morex_contig_44546	7	62.181	-	FWH_feat1_7	-	-0.76	-	-	-	main
seq532	AV835441_U35_28764_MLOC_14463.1_PR	7	62.288	-	FWH_feat1_7	-	-1.38	-	-	-	main
seq217	morex_contig_1566169	7	65.120	-	FWH_feat1_7	-3.47	-	-	-	-	main
-	morex_contig_66819	7	65.439	-	FWH_feat1_7	-	-0.21	-	-	-	random
seq443	AK371674	7	65.758	-	FWH_feat1_7	-	-1.89	-	-	-	main
seq44	bowman_contig_849794_RC	7	66.942	-	FWH_feat1_7	-	-1.47	-	-	-	main
seq400	morex_contig_1560804	7	67.280	-	FWH_feat1_7	-6.10	-	-	-	-	main
seq455	AK372518	7	67.635	-	FWH_feat1_7	-	-	-	-	16.8	main
seq567	BG369393	7	67.635	-	FWH_feat1_7	-	-	-	22.4	-	main
-	morex_contig_1566103	7	67.635	-	FWH_feat1_7	-	-	-	-	0.8	off_target
-	morex_contig_1579932	7	67.635	-	FWH_feat1_7	-	-	-	39.1	-	off_target
-	morex_contig_49903	7	67.635	-	FWH_feat1_7	-	-	-	-	0.4	off_target
seq88	morex_contig_140238	7	67.776	-	FWH_feat1_7	-	-0.48	-	-	-	main
seq186	morex_contig_45996	7	67.776	-	FWH_feat1_7	-	-	-	-	-	main
seq8	AF490467	7	67.918	-	FWH_feat1_7	-	-0.54	-	-	-	main
seq458	AK372654	7	67.918	-	FWH_feat1_7	-	0.00	-	-	-	main
seq576	BI952411	7	67.918	-	FWH_feat1_7	-	-	-	-	-	main
seq8	morex_contig_138334	7	67.918	-	FWH_feat1_	-	-	-	-	-	main

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
					7						
seq540	morex_contig_61751	7	69.263	-	FWH_feat1_7	-	-1.09	-	-	1.9	main
-	morex_contig_38891	7	69.325	-	FWH_feat1_7	-	-	-	22.0	-	off_target
-	morex_contig_1559680	7	69.559	-	FWH_feat1_7	-	-	-	-	13.5	off_target
-	morex_contig_49081	7	69.559	-	FWH_feat1_7	-	-	-	-	1.0	off_target
seq202	morex_contig_53730	7	69.559	-	FWH_feat1_7	-	0.33	-	24.2	-	main
seq131	AJ249147	7	70.125	-	FWH_feat1_7	-	-	-	107.7	-	main
-	morex_contig_1581454	7	70.125	-	FWH_feat1_7	-3.05	-	-	-	4.7	off_target
-	morex_contig_41934	7	70.125	-	FWH_feat1_7	-	-	-	-	4.3	off_target
-	morex_contig_68766	7	70.125	-	FWH_feat1_7	-	-	-	32.2	-	off_target
seq185	AK253094	7	70.538	-	FWH_feat1_7	-3.34	-	-	-	2.9	main
seq214	AK355971	7	70.538	-	FWH_feat1_7	-	-	-	-	-	main
seq312	AK363115	7	70.538	-	FWH_feat1_7	-	-	-	-	-	main
seq500	AK375083	7	70.538	-	FWH_feat1_7	-	0.09	-	-	-	main
seq30	bowman_contig_10723	7	70.538	-	FWH_feat1_7	-	0.20	-	-	-	main
-	morex_contig_112354	7	70.538	-	FWH_feat1_7	-	-	-	-	2.2	off_target
-	morex_contig_135487	7	70.538	-	FWH_feat1_7	-	-	-	44.4	-	off_target
-	morex_contig_138069	7	70.538	-	FWH_feat1_7	-	-	-	-	2.5	off_target
seq225	morex_contig_2550672	7	70.538	-	FWH_feat1_7	-	-2.18	-	-	3.8	main

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
seq307	morex_contig_367934	7	70.538	-	FWH_feat1_7	-	0.09	-	29.4	-	main
-	morex_contig_40236	7	70.538	-	FWH_feat1_7	-	-	-	-	2.5	random
seq570	morex_contig_43517	7	70.538	-	FWH_feat1_7	-	-	-	-	-	main
-	morex_contig_7417	7	70.538	-	FWH_feat1_7	-	-1.24	-	-	-	off_target
-	morex_contig_89173	7	70.538	-	FWH_feat1_7	-5.06	-	-	-	-	off_target
seq565	morex_contig_9154_RC	7	70.538	-	FWH_feat1_7	-4.49	-	-	-	-	main
-	morex_contig_133539	7	70.609	-	FWH_feat1_7	-	0.41	-	-	-	off_target
seq359	morex_contig_1566622_RC	7	70.609	-	FWH_feat1_7	-	-2.59	-	-	-	main
-	morex_contig_1619666	7	70.609	-	FWH_feat1_7	-	0.18	-	-	-	off_target
seq545	morex_contig_2548236	7	70.609	-	FWH_feat1_7	-5.59	-	-	-	-	main
seq423	morex_contig_38788	7	70.609	-	FWH_feat1_7	-2.78	-	-	-	-	main
seq149	morex_contig_49566	7	70.609	-	FWH_feat1_7	-	-	-	-	-	main
seq137	morex_contig_54219_RCmod	7	70.609	-	FWH_feat1_7	-	-1.70	-	-	-	main
-	morex_contig_159669	7	70.680	-	FWH_feat1_7	-3.30	-	-	-	-	off_target
-	morex_contig_54812	7	70.680	-	FWH_feat1_7	-	0.15	-	-	-	off_target
seq15	morex_contig_1565670	7	70.803	-	FWH_feat1_7	-	-	-	-	-	main
-	morex_contig_274546	7	70.803	-	FWH_feat1_7	-5.82	-	-	-	-	random
-	morex_contig_98446	7	70.803	-	FWH_feat1_7	-	0.09	-	-	-	random
seq100	morex_contig_1567295	7	70.822	-	FWH_feat1_7	-6.88	-	-	-	-	main
-	morex_contig_53371	7	71.459	-	FWH_feat1_7	-	0.24	-	-	-	off_target
-	morex_contig_50098	7	75.071	-	FWH_feat1_7	-	-	NPI_feat1_7	-	7.8	off_target
-	morex_contig_51577	7	75.071	-	FWH_feat1_7	-	-	NPI_feat1_7	44.6	-	random
seq105	morex_contig_158492	7	75.212	-	FWH_feat1_7	-5.32	-	NPI_feat1_7	60.2	-	main
-	morex_contig_42736	7	75.212	-	FWH_feat1_7	-	-	NPI_feat1_7	26.6	-	off_target
-	morex_contig_50632	7	75.212	-	FWH_feat1_7	-	-	NPI_feat1_7	-	1.8	off_target
-	morex_contig_194982	7	75.814	-	FWH_feat1_7	-	-	NPI_feat1_7	35.0	-	off_target

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
seq278	morex_contig_1558911_RC	7	76.416	-	FWH_feat1_7	-3.22	-	NPI_feat1_7	-	5.1	main
-	morex_contig_1570092	7	76.416	-	FWH_feat1_7	-	0.09	NPI_feat1_7	-	-	off_target
seq286	morex_contig_271159_RC	7	76.416	-	FWH_feat1_7	-	0.23	NPI_feat1_7	-	11.3	main
seq488	morex_contig_40188	7	76.416	-	FWH_feat1_7	-	-0.89	NPI_feat1_7	-	2.8	main
-	morex_contig_50005	7	76.416	-	FWH_feat1_7	-	0.09	NPI_feat1_7	-	-	off_target
-	morex_contig_370071	7	76.466	-	FWH_feat1_7	-4.66	-	NPI_feat1_7	-	-	random
-	morex_contig_37333	7	76.466	-	FWH_feat1_7	-	-	NPI_feat1_7	-	0.5	random
seq669	morex_contig_40601_RC	7	76.466	-	FWH_feat1_7	-	0.25	NPI_feat1_7	27.0	-	main
-	morex_contig_66811	7	76.466	-	FWH_feat1_7	-	-	NPI_feat1_7	39.7	-	off_target
-	morex_contig_9593	7	76.466	-	FWH_feat1_7	-	-	NPI_feat1_7	-	9.8	off_target
-	morex_contig_160133	7	76.558	-	FWH_feat1_7	-	-	NPI_feat1_7	-	4.4	off_target
seq627	morex_contig_2547662_RC	7	76.558	-	FWH_feat1_7	-3.38	-	NPI_feat1_7	27.3	-	main
-	morex_contig_2548958	7	76.558	-	FWH_feat1_7	-4.97	-	NPI_feat1_7	-	16.8	random
-	morex_contig_135463	7	76.629	-	FWH_feat1_7	-3.58	-	NPI_feat1_7	-	1.5	off_target
-	morex_contig_1568495	7	76.700	-	FWH_feat1_7	-	-	NPI_feat1_7	65.6	-	off_target
-	morex_contig_43146	7	76.700	-	FWH_feat1_7	-	-	NPI_feat1_7	-	1.0	off_target
seq85	morex_contig_55549	7	76.700	-	FWH_feat1_7	-	-0.62	NPI_feat1_7	-	1.1	main
seq270	bowman_contig_77397_RC	7	77.266	-	FWH_feat1_7	-	-1.56	NPI_feat1_7	-	3.8	main
seq342	AK364492	7	77.535	-	FWH_feat1_7	-	-	NPI_feat1_7	-	7.0	main
seq614	morex_contig_39696_RC	7	77.535	-	FWH_feat1_7	-	0.09	NPI_feat1_7	40.7	-	main
-	morex_contig_1565248	7	77.904	-	FWH_feat1_7	-	0.46	NPI_feat1_7	-	1.1	off_target
-	morex_contig_361657	7	77.904	-	FWH_feat1_7	-3.10	-	NPI_feat1_7	-	-	off_target
seq542	AY541065	7	79.241	-	FWH_feat1_7	-	-	NPI_feat1_7	-	-	main
-	morex_contig_8411	7	84.455	-	FWH_feat1_7	-4.58	-	NPI_feat1_7	-	-	off_target
-	morex_contig_2547188	7	84.561	-	FWH_feat1_7	-4.58	-	NPI_feat1_7	-	-	off_target
seq3	AB678347	7	89.136	-	FWH_feat1_7	-	-0.32	NPI_feat3_7	-	3.1	main

Locus	Reference	Chromosome	Position, cM	LD ratio feature	Fay & Wu's Hnorm			π wild / π dom			
					Feature	Outlier	Non-outlier	Feature	Outlier	Non-outlier	Target type
seq301	AK362526	7	89.136	-	FWH_feat1_7	-	0.20	NPI_feat3_7	-	0.9	main
-	morex_contig_274295	7	89.136	-	FWH_feat1_7	-6.15	-	NPI_feat3_7	-	3.8	off_target
-	morex_contig_47363	7	89.136	-	FWH_feat1_7	-	-	NPI_feat3_7	25.2	-	off_target
-	morex_contig_133728	7	89.518	-	FWH_feat1_7	-	-	NPI_feat3_7	-	8.2	off_target
seq120	morex_contig_53102	7	89.518	-	FWH_feat1_7	-4.59	-	NPI_feat3_7	-	4.5	main
seq363	AK366098	7	91.785	-	FWH_feat1_7	-	0.68	NPI_feat3_7	-	1.3	main
-	morex_contig_37018	7	92.068	-	-	-	-	NPI_feat3_7	-	11.7	off_target
seq171	morex_contig_442040	7	92.210	-	-	-	-	NPI_feat3_7	27.7	-	main
-	morex_contig_7874	7	92.210	-	-	-	-	NPI_feat3_7	-	4.8	off_target
-	morex_contig_37799	7	92.302	-	-	-	0.25	NPI_feat3_7	-	-	off_target
seq538	AW983439_MLOC_58330.2	7	108.782	-	-	-2.84	-	-	-	-	main
-	morex_contig_274814	7	109.731	-	-	-4.85	-	-	-	-	random
-	morex_contig_51275	7	131.870	-	-	-2.85	-	-	-	-	off_target
-	morex_contig_1570083	7	140.864	-	-	-3.65	-	-	-	-	off_target
seq241	HVVMRXALLhA0790N05_c3_RC	-	-	-	-	-3.59	-	-	-	-	main
-	morex_contig_100416	-	-	-	-	-6.07	-	-	-	-	off_target
seq177	morex_contig_1560516	7	-	-	-	-5.13	-	-	-	-	main
seq379	morex_contig_161307	7	-	-	-	-3.52	-	-	-	-	main
seq663	morex_contig_236164_RC	7	-	-	-	-2.72	-	-	-	-	main
seq456	morex_contig_244364_RC	3	-	-	-	-3.19	-	-	-	-	main
-	morex_contig_475398	-	-	-	-	-3.62	-	-	-	-	off_target
seq449	morex_contig_39739	3	-	-	-	-	-	-	108.8	-	main

* - presented are the outlier genes carrying signatures of selection and the flanking non-outlier genes.

SUPPLEMENTARY MATERIAL - CHAPTER 2

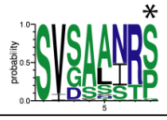
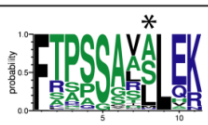
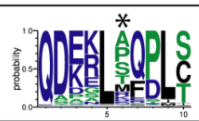
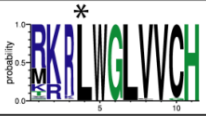
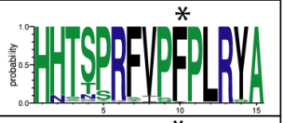
Table S1 Fifteen polymorphic haplotypes of HvPHYC.

Haplotype	Position (bp) relative to the start codon of the reference Morex *HvPHYC* gene (DQ238106)																					No. of samples
	149	**175**	**206**	**209**	**210**	249	**415**	906	**920**	1008	**1027**	**1091**	**1139**	1239	**1403**	1491	1659	1722	1764	**1921**	**1945**	
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	70
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	14
3	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	T	-	-	-	4
4	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	11
5	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	A	2
6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	2
7	-	-	-	-	-	-	-	-	-	-	-	C	C	-	-	-	-	-	-	-	-	1
8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	1
9	-	-	-	A	C	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	A	1
10	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
11	-	-	-	-	-	T	-	-	T	A	-	-	-	T	-	-	-	-	-	A	A	1
12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	A	-	-	-	-	1
13	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
14	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
15	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	-	-	-	-	1

n = 114

Non-synonymous single nucleotide polymorphisms (nsSNP) are highlighted in red. Position of the nsSNP found in *HvPHYC* from Bowman(*eam5*) is underlined.

Table S2 Non-synonymous SNPs in the exon 1 of HvPHYC and conservation of the corresponding amino-acid (a.a.) residues.

SNP ID[1]	A.a. substitution[2]	# of sequences in the logo	Motif logo[3]
149	S50Y	72	
175	S59T	231	
209-210	Y70F	249	
415	V139I	966	
920	S307L	4298	
1027	I343V	4331	
1091	L364D	4340	
1139	F380S	4353	
1403	K468R	4107	
1921	V641I	1610	
1945	G649S	1610	

1 - SNP positions given from the first nucleotide of the *HvPHYC* start codon;
2 - Deleterious a.a. substitutions as predicted by PROVEAN (cut-off score -2,5) are highlighted in red;
3 - substituted a.a. residues marked within the motifs with asterisks; letters in black,
green and blue refer to hydrophobic, neutral and hydrophilic a.a. residues, respectively.

Table S3 Accessions used for the HvPHYC re-sequencing and haplotype analysis.

***Hordeum* species**	**Genotype***	**Status**	**Growth habit**	**Origin**	***HvPHYC* haplotype**
vulgare subsp. *vulgare*	Arta	cultivar	winter	Syria	1
	Asahi 5	cultivar	n.d.**	Japan	4
	Azumamugi	cultivar	n.d.	Japan	7
	B.E. 22 (ASA)	cultivar	n.d.	Pakistan	2
	B1K-70-01	cultivar	spring	Israel	1
	B1K-70-02	cultivar	spring	Israel	1
	Bowman	cultivar	n.d.	USA	14
	Bowman(*eam5*)	introgression line	spring	n.d.	4
	Dicktoo	cultivar	n.d.	USA	15
	Erectoides 16	cultivar	n.d.	Sweden	1
	G-391 F	cultivar	spring	Italy	2
	G-413 I	cultivar	n.d.	Czech Republic	1
	Ghara 1 (1609)	cultivar	n.d.	Nepal	3
	Hamidiye 85	cultivar	spring	Turkey	1
	Haruna Nijo	cultivar	spring	Japan	4
	Hayachinemugi	cultivar	n.d.	Japan	4
	Hayakiso 2	cultivar	winter	Japan	4
	Hayakiso 3	cultivar	n.d.	Japan	1
	Indian dwarf	cultivar	n.d.	n.d.	1
	Indo Omugi	cultivar	spring	Taiwan	1
	Ishuku Shirazu	cultivar	spring	Japan	4
	Kagoshima Gold	cultivar	n.d.	Japan	1
	Kaikei 84	cultivar	n.d.	Japan	4
	Kanto Nijo 3	cultivar	n.d.	Japan	4
	Kawasaigoku	cultivar	spring	Japan	4
	Keel	cultivar	spring	Australia	1
	Kinai 5	cultivar	n.d.	Japan	1
	Kindoku	cultivar	n.d.	Sweden	1
	Kompolti	cultivar	n.d.	Hungary	2
	L871	cultivar	spring	Egypt	1
	Mari	cultivar	n.d.	Sweden	1
	Marthe	cultivar	spring	Germany	1
	Morex	cultivar	spring	USA	1
	Mota 4 (1-24-13)	cultivar	n.d.	Ethiopia	1
	Mota 6 (1-24-15)	cultivar	n.d.	Ethiopia	1
	Omugi 15	cultivar	n.d.	Japan	4
	Rum	cultivar	n.d.	Jordand	1
	Saikai Kawa 24	cultivar	n.d.	Japan	4
	Shiga Hayakiso 1	cultivar	n.d.	Japan	1
	Sladoran	cultivar	winter	Turkey	2
	Tadmor	cultivar	winter	Syria	1
	Tainan 2	cultivar	n.d.	Taiwan	1
	Turkey 759	cultivar	n.d.	Turkey	1
	Yarmouk	cultivar	n.d.	Lebanon	2
	Yercil 147	cultivar	winter	Turkey	2

***Hordeum* species**	**Genotype***	**Status**	**Growth habit**	**Origin**	***HvPHYC* haplotype**
	Zairai 1	cultivar	n.d.	Taiwan	1
	Zairai 2	cultivar	n.d.	Taiwan	1
	B1K-55-01	landrace	n.d.	Israel	2
	B1K-55-02	landrace	n.d.	Israel	1
	B1K-55-06	landrace	spring	Israel	2
	G-1573 A	landrace	n.d.	Syria	1
	G-398 H	landrace	n.d.	Ethiopia	2
	G-400 H	landrace	n.d.	Egypt	1
	G-404 H	landrace	spring	Tibet	3
	G-423 H	landrace	n.d.	Ethiopia	1
	G-434 H	landrace	n.d.	Ethiopia	1
	G-439 H	landrace	spring	Yemen	1
	G-440 E	landrace	n.d.	n.d.	1
	LR 1887	landrace	n.d.	n.d.	1
vulgare subsp. *spontaneum*	B1K-02-18	wild	n.d.	Israel	9
	B1K-03-07	wild	n.d.	Israel	1
	B1K-05-13	wild	n.d.	Israel	5
	B1K-08-13	wild	n.d.	Israel	5
	B1K-08-18	wild	n.d.	Israel	10
	B1K-13-01	wild	n.d.	Israel	2
	B1K-17-13	wild	n.d.	Israel	1
	B1K-21-02	wild	n.d.	Israel	1
	B1K-22-06	wild	n.d.	Israel	1
	B1K-22-10	wild	n.d.	Israel	1
	B1K-33-13	wild	n.d.	Israel	1
	HID-1	wild	winter	Iraq	1
	HID-10	wild	n.d.	Iraq	1
	HID-101	wild	winter	Syria	1
	HID-104	wild	n.d.	Syria	13
	HID-107	wild	winter	Jordan	2
	HID-109	wild	winter	Syria	12
	HID-114	wild	n.d.	Lebanon	1
	HID-122	wild	n.d.	Jordan	11
	HID-136	wild	winter	Iran	1
	HID-137	wild	n.d.	Turkey	13
	HID-140	wild	winter	Iraq	13
	HID-145	wild	n.d.	Israel	2
	HID-2	wild	n.d.	Iraq	1
	HID-24	wild	n.d.	Iran	3
	HID-257	wild	n.d.	Israel	1
	HID-301	wild	n.d.	Iran	1
	HID-309	wild	n.d.	Iran	2
	HID-330-1	wild	n.d.	n.d.	1
	HID-366	wild	n.d.	Iran	1
	HID-377-1	wild	n.d.	Israel	1
	HID-377-2	wild	n.d.	Israel	1
	HID-44	wild	n.d.	Iran	1

***Hordeum* species**	**Genotype***	**Status**	**Growth habit**	**Origin**	***HvPHYC* haplotype**
	HID-46	wild	n.d.	Iran	1
	HID-54	wild	n.d.	Turkey	2
	HID-55	wild	n.d.	Turkey	1
	HID-56	wild	n.d.	Turkey	1
	HID-70	wild	n.d.	Turkey	1
	HID-85	wild	n.d.	Turkey	1
	HID-96	wild	n.d.	Jordan	1
	HID-99	wild	winter	Syria	1
	HP-02-3	wild	n.d.	Turkey	1
	HP-03-2	wild	n.d.	Turkey	1
	HP-10-4	wild	n.d.	Turkey	1
	HP-10-5	wild	n.d.	Turkey	1
	HP-11-1	wild	n.d.	Turkey	1
	HP-13-2	wild	n.d.	Turkey	8
	HP-15-3	wild	winter	Turkey	6
	HP-15-5	wild	winter	Turkey	6
	HP-24-1	wild	winter	Turkey	1
	HP-26-1	wild	winter	Turkey	1
	HP-27-2	wild	winter	Turkey	1
agriocrithon	B1K-52-01	wild	n.d.	Israel	1
	HID-383-1	wild	winter	China	3
	HID-383-3	wild	n.d.	China	1

* - genotypes carrying the wild-type HvVRN1 allele are highlighted in bold; allelic composition at HvVRN-H1 in other genotypes is unknown; ** - n.d., no data.

Table S4 SCAR and CAPS markers for genotyping, sequencing and real-time experiments: PCR primers and amplification regimes.

	Primer	Sequence (5'-3')	Fragment size (bp)*	Ta**	Cycle ***	Restric-tion enzyme	Digested fragment size*
Allele-specific SCAR and CAPS markers	PIFcaps_f	GAGCAGTACGCGCACTTC	301	58	A	*Hg*aI	30+270/30+80+190
	PIFcaps_r	CTTTTGTTTGGGTGTATCGC					
	PHYCcaps_f	GGTCCTAATGCAAGGCATGT	880	62	A	*Btg*I	650+230/650+160+70
	PHYCcaps_r	CTCTTGCTGTTGAGCTGTGC					
	CK2Acaps_f	GTTTGTCTGCGCATGCGTG	410	60	B	*Acu*I	300+110/410
	CK2Acaps_r	ATGTTGGACAGAACATTCACAC					
	VIP4.1f	TGCTGGGATGTTATCCATG	200/350	57	B	none	
	VIP4.1r	GTGAATTGTAACAGCTCGC					
	VIP4.2f	CATGGGTGTTGGAATAATTG	180/200	57	B	none	
	VIP4.2r	ACCAAATGTCATTACGATCTC					
	VRN-H1f	AATACGACTCACTATAGGGGAAAACTTGAACAACACCAGAACC	320/360	50	C	none	
	VRN-H1r	TTCTGCATAAGAGTAGCGCTCAT					
VIP4 sequencing	seqVIP4.1f	ATGCAACTACTGATTGGCG	450	61	A	none	
	seqVIP4.1r	CTCAATCTCTTCGTTTGG					
	seqVIP4.2f	ATGCATCCAACAAGTCCC	900	63	A	none	
	seqVIP4.2r	ACTCACTCGATCAGGTTG					
HvPHYC exon 1 sequencing	Ex1seq_1f	CCCGTCCTTCTCCACAAAAG	1100	62	A	none	
	Ex1seq_1r	GAGCCACAGAGGCTGATAGG					
	Ex1seq_2f	ACTACCCGGCAACTGACATC	1200	62	A	none	
	Ex1seq_2r	ACAGAATCACCCTCCACGAG					

* Expected allele size is given in the following format: Bowman / Bowman(*eam5*).

** (A) – 98ºC for 2 m; 4 touchdown cycles of 98ºC – 30 s, (Ta+4) – 30 s (-1ºC/cycle), 72ºC – 1 m; 31 cycles of 98ºC – 30 s, Ta – 30 s, 72ºC – 1 m; final extension 72ºC – 10 m; (B) – 94ºC for 2 m; 4 touchdown cycles of 94ºC – 30 s, (Ta+4) – 30 s (-1ºC/cycle), 72ºC – 30 s; 31 cycles of 94ºC – 30 s, Ta – 30 s, 72ºC – 30 s; final extension 72ºC – 10 m; (C) – 94ºC for 2 m; 9 touchdown cycles of 94ºC – 30 s, (Ta+9) – 30 s (-1ºC/cycle), 72ºC – 30 s; 30 cycles of 94ºC – 30 s, Ta – 30 s, 72ºC – 30 s; (D) – 95oC for 5 m; 45 cycles of 95oC – 10 s, Ta – 10 s, 72oC – 10 s, 82oC – 10 s.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my supervisor Professor Maria von Korff. I am indebted to you for your continuous support, your trust and confidence and for the stress-free atmosphere you managed to create in your group.

I would like to thank Professor George Coupland for kindly admitting me to the graduate school in his department, examining my dissertation and, more importantly, for being a professional role model for me.

I am thankful to Professor Juliette de Meaux and Professor Maria Albani for their indispensable roles as my second examiner and the chair of the defense committee.

I am grateful to my second supervisor, José M Jiménez-Gómez, and the managers of the IMPRS program, Olof Persson and Johanna Spandl for keeping the process so smooth.

Thanks to all the current members and alumni of our group - Aman, Chiara, Jarod, Ermias, Xiaojing, Cristina, Lukas, Elisabeth, Kerstin, Andrea, Teresa, Mahwish, Wilma, Agatha and Filipa.

Thanks to my friends, who has been shaping my personality during these years - Benny for sharing your deepest thoughts and for your kind help with everything, Dima a fellow countryman, Tingting, Sven for your Zen spirit, Xiaojing for endless chats and the crash course in Chinese history and delicacies, Oliver for the craziest times we've had in Cologne, Ganga for Shanti and being a traveling companion. To my Moscow friends, George and Alex, special thanks for inspiring me, being physically distant, but mentally so close.

My dear Mum, thank you for your unconditional love and support. I am sorry for not being able to be closer for so long.

ERKLÄRUNG

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegt worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Jun.-Prof. Dr. Maria von Korff Schmising und Prof. Dr. George Coupland betreut worden.

Ich versichere, dass ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe und verpflichte mich, jedmögliche, die obigen Angaben betreffenden Veränderungen, dem Dekanat unverzüglich mitzuteilen.

Köln, den _____ _____

Artem Pankin

Teilpublikationen:

Das zweite Kapitel dieser Arbeit wurde publiziert in:

Artem Pankin*, Chiara Campoli*****, Xue Dong, Benjamin Kilian, Rajiv Sharma, Axel Himmelbach, Reena Saini, Seth J Davis, Nils Stein, Korbinian Schneeberger, Maria von Korff: *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 (2014); 198:383-396. * - equal contribution

ARTEM PANKIN, Kandidat der Wissenschaften (Biologie)

Adresse: Brüsseler Str. 77
50672 Köln
Telefon: +49 176 61386859
Email : artem.pankin@yahoo.com; pankin@mpipz.mpg.de

Geburtsdatum: 21. Dezember 1982
Geburtsort: Chimki, Moskau, Russland

BERUFSERFAHRUNG

02/2010-10/2011 **Wissenschaftlischer Mitarbeiter**, USDA-ARS, Beltsville, MD

11/2004-03/2011 **Wissenschaftlicher Mitarbeiter**, Russischen wissenschaftlichen Forschungsinstituts für landwirtschaftliche Biotechnologie

AUSBILDUNG

seit 11/2011 **Doktorand** am Max-Planck Institut für Pflanzenzüchtungsforschung, Köln mit Promotionsstipendium der International Max-Planck Research School; Betreuer: Prof. Dr. George Coupland und Jun. Prof. Dr. Maria von Korff
Thema: " Diversitätsanalyse bei Gerste (Hordeum vulgare)"

11/2004–11/2008 **Aspirantur** am Russischen wissenschaftlichen Forschungsinstituts für landwirtschaftliche Biotechnologie
Betreuer: Prof. Dr. Emil Khavkin

09/1999–06/2004 **Diplomarbeit** am „K.A. Timirjasew – Landwirtschaftsakademie", Moskau

06/1999 **Abitur** am Gymnasium №16, Chimki, Moskau, Russland

PUBLIKATIONEN

09/2015 Benedikt Digel, **Artem Pankin**, Maria von Korff: *Global Transcriptome Profiling of Developing Leaf and Shoot Apices Reveals Distinct Genetic and Environmental Control of Floral Transition and Inflorescence Development in Barley.* The Plant Cell; 27:2318-2334

09/2014 **Artem Pankin***, Chiara Campoli*, Xue Dong, Benjamin Kilian, Rajiv Sharma, Axel Himmelbach, Reena Saini, Seth J Davis, Nils Stein, Korbinian Schneeberger, Maria von Korff: *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. * - equal contribution

01/2014 Liangyu Liu, Jessika Adrian, **Artem Pankin**, Jinyong Hu, Xue Dong, Maria von Korff, Franziska Turck: *Induced and natural variation of promoter length modulates the photoperiodic response of FLOWERING LOCUS T.* Nature Communications; 5:4558.

06/2013 Chiara Campoli*, **Artem Pankin***, Benedikt Drosse, Cristina M Casao, Seth J Davis, Maria von Korff: *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. * - equal contribution

03/2011 **Artem Pankin**, Emil Khavkin: *Genome-specific SCAR markers help solve taxonomy issues: a case study with Sinapis arvensis (Brassiceae, Brassicaceae).* American Journal of Botany; 98:e54-7.

01/2011 **Artem Pankin**, Ekaterina Sokolova, Elena Rogozina, Maria Kuznetsova, Kenneth Deahl, Richard Jones, Emil Khavkin: *Allele mining in the gene pool of wild Solanum species for homologues of late blight resistance gene RB/Rpi-blb1*. Plant Genetic Resources; 9:305-308.