Discovery of barley gene candidates for low temperature and drought tolerance via environmental association

Li Lei[1], Ana M. Poets[1], Chaochih Liu[1], Skylar R. Wyant[1], Paul J. Hoffman[1], Corey K. Carter[1], Richard M. Trantow[1], Brian G. Shaw[1], Xin Li[1], Gary J. Muehlbauer[1,2], Fumiaki Katagiri[2], Peter L. Morrell[1]

[1]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108
[2]Department of Plant and Microbial Biology, Microbial and Plant Genomics Institute, University of Minnesota, St. Paul, Minnesota 55108

Summary

- Barley is cultivated from the equator to the Arctic Circle. The wild progenitor species, *Hordeum vulgare ssp. spontaneum*, occupies a relatively narrow latitudinal range (~30 - 40° N) primarily at low elevation, < 1500 m. Adaptation to the range of cultivation has occurred over ~8,000 years. The genetic basis of this adaptation is amenable to study through environmental association.
- Using genotyping from 7,864 SNPs in 784 barley landraces, we perform mixed model association analysis relative to bioclimatic variables and analysis of allele frequency differentiation across multiple partitions of the data. Using resequencing data from a subset of the landraces, we test for linkage disequilibrium (LD) between SNPs queried in genotyping and SNPs in neighboring loci.
- We identify seven loci previously reported to contribute to adaptive differences to flowering time and abiotic stress in barley and four loci previously identified in other plant species. In many cases, patterns of LD are consistent with the causative variant occurring in the immediate vicinity of the queried SNP.
- The identification of barley orthologs to well characterized genes may provide new understanding of the nature of adaptive variation and could permit a more targeted use of potentially adaptive variants in barley breeding and germplasm improvement.

Introduction

Wild barley, the progenitor of cultivated barley, occurs across a limited latitudinal range, with the primary distribution between 30 - 40˚ N latitude (Harlan and Zohary, 1966). The geographic range of wild barley (*Hordeum vulgare* ssp. *spontaneum*) is bisected by the Zagros Mountains, with peaks up to 4,400 m, but wild barley is largely limited in distribution to sites at < 1,500 m (Zohary et al., 2012).

Barley was domesticated from its wild progenitor ~10,000 - 12,000 years ago. Domestication occurred at least twice (Morrell and Clegg, 2007) and involved genetic contributions from across the geographic range of wild barley (Poets et al., 2015a). The dissemination of cultivated barley beyond the initial centers of origin began roughly 2,000 years after domestication (Willcox, 2002). Thus barley landraces and modern cultivars are the result of adaptation to growing conditions in Eurasia, North Africa, and much more recently, to Australia and the New World. In Eurasia, the process occurred as humans adopted an agropastoral lifestyle and spread from the Fertile Crescent into a variety of geographic regions. This included cultivation in regions with cooler and more mesic climates in Europe (Pinhasi et al., 2005) and into drier climates in North Africa and Central Asia (Harris and Gosden, 1996). In East Africa, Asia, and Europe, barley is frequently produced at high elevations. Barley remains among the most important crops in Nepal and Tibet, where it is grown at elevations up to 4,700 m.

The adaptation of cereals such as barley and wheat to northern latitudes or dry climatic conditions involved changes in vernalization requirements, growth habit, and flowering time (Turner et al., 2005; Yan et al., 2006; Comadran et al., 2012; Zakhrabekova et al., 2012; Dawson et al., 2015). Wild species adapted to Mediterranean climates typically grow over winter and flower in the spring. This is known as winter growth habit. Under cultivation, winter annuals such as barley and wheat have been adapted to colder climates through spring planting, resulting in a summer harvest. This is known as spring growth habit. Spring planting can make cultivation possible at higher latitudes but also increases exposure to frost and freezing conditions (Visioni et al., 2013).

The genetic basis of vernalization and flowering time adaptation in barley has been explored extensively and multiple genes have been cloned (seeHansson et al., 2018). There are also numerous mapping studies and some functional studies that have identified regions of the genome associated with low temperature tolerance in barley (Hayes et al., 1993; Francia et al., 2004; Reinheimer et al., 2004; Skinner et al., 2006; Tondelli et al., 2006; Visioni et al., 2013). Only the CBF gene cluster is extensively characterized (Knox et al., 2010). Because collecting cold tolerance phenotypes that go beyond “winter survival” is challenging, it can be difficult to

explore new genes related to cold tolerance in barley (Visioni et al., 2013). There have been multiple studies of drought tolerance in barley, but the genetic basis of drought tolerance has been less extensively explored (Honsdorf et al., 2014). Previous research identified the *P5CS* gene (encoding delta-1-pyrroline-5-carboxylate synthase) as a primary drought-responsive gene in barley that also confers some level of drought resistance (Abu-Romman et al., 2011; Shaar-Moshe et al., 2015; Wehner et al., 2016; Xia et al., 2017).

Approaches for the identification of genetic variants contributing to environmental adaptation must differentiate adaptive variation from loci subject to neutral evolutionary processes. Demographic effects acting on populations, such as rates of migration or changes in population size impact the entire genome, whereas selection alters allele frequencies at individual loci (Cavalli-Sforza, 1966). Lewontin and Krakauer (1973) proposed an approach to identify variants subject to differential selection between populations based on allele frequency differentiation, measured by F-statistics, to identify variants subject to differential selection between populations. F-statistics based comparisons suffer from several weaknesses, including a high expected variance in F_{ST} values (Nei and Maruyama, 1975) and the often arbitrary nature of partitioning of populations (Lotterhos and Whitlock, 2014). If informative population partitions are defined, F_{ST} measures can potentially identify loci subject to strong differential selection (Beaumont and Balding, 2004). Another approach used to identify the genetic basis of environmental adaptation is mixed-model association analysis. This approach explicitly controls for population structure and treats bioclimatic variables such as average temperatures and rainfall as “phenotypes” (Eckert et al., 2010; Yoder et al., 2014). An association analysis of bioclimatic variables successfully identified candidate loci for environmental adaptation in soybean (Anderson et al., 2016).

Here, we present allele frequency (F_{ST}) outlier and mixed-model association analyses applied to 784 barley landraces genotyped with the barley 9K Illumina Infinium iSelect Custom Genotyping BeadChip (Comadran et al., 2012) to identify loci potentially involved in cold and drought tolerance. For the F_{ST} outlier analysis, we focus on partitions of the sample that differentiate growth conditions. These include, latitude, elevation, and spring versus winter growth habit. Previous studies have identified genetic differentiation in eastern and western landraces, associated with multiple domestications (Morrell and Clegg, 2007; Saisho and Purugganan, 2007; Poets et al., 2015a). Thus to identify factors that contribute most to allele frequency differentiation, we also calculate F_{ST} for a longitudinal comparison. We address the following questions: 1) Which of the comparisons explains the largest portion of allele frequency differentiation?, 2) How many previously reported cold temperature and drought tolerance-

related loci show evidence of contributing to climatic adaptation?, 3) Given the linkage disequilibrium expected in a self-fertilizing species, how frequently are SNPs identified in our analyses in the proximity of potentially causative loci? For the third question, we make use of exome capture resequencing from a sample of 62 landraces from the larger panel. This permits a direct estimate of linkage disequilibrium between SNPs identified in our broader panel of accessions and variants in a window of loci surrounding each locus.

We identify a total of seven barley loci previously reported to be involved in either cold or drought tolerance or in flowering time. Furthermore, our analyses identified four barley orthologs of genes characterized as contributing to these traits in other plant species. A slight relaxation of the empirical cutoff for outlier F_{ST} values identifies an additional three characterized barley genes and eight orthologs from other plants previously identified as contributing to potential adaptive phenotypes. Considering both allele frequency outlier and bioclimatic association analyses, we identify 339 barley loci not previously reported to be associated with environmental adaptation. Comparisons of linkage disequilibrium (LD) between SNPs in genotyping and resequencing suggest that because of relatively low gene density in barley, roughly a quarter of the genes we identify on the basis of SNP genotyping are the strongest candidates for association.

Materials and Methods

Plant Materials

We used the 803 barley landrace accessions selected from among 2,446 cultivated barley lines in the USDA-maintained core collection (Muñoz-Amatriaín et al., 2014). These lines also examined in a study of barley domestication history by Poets et al. (2015c). The 803 individuals were collected from Europe, Asia, and North Africa, covering the range of barley cultivation in human pre-history (Willcox, 2002; Pinhasi et al., 2005; Poets et al., 2015c). The reported geographic coordinates for each accession were manually confirmed to identify potentially inaccurate locations (Table S1). The elevations of the collection locations were inferred from the NASA Shuttle Radar Topographic Mission (SRTM) 90 m data (http://www.cgiar-csi.org/) on Oct 7, 2015 using the getData ‘raster’ R function (Hijmans et al., 2016).

Genotyping data

All samples were genotyped using the 7,864 SNPs on the 9k Illumina Infinium iSelect Custom Genotyping BeadChip (Comadran et al., 2012) genotyping platform (henceforth referred

to as 9K SNPs). For more detail on the SNP discovery panel see the description in Comadran et al. (2012; 2015c). The genotyped dataset was filtered for monomorphic SNPs, SNPs with > 10% missingness, or SNPs with > 10% observed heterozygosity.

Estimating crossover relative to physical distance

We identified the physical position of 9K SNPs based on mapping SNP contextual sequence to the barley reference genome (Mascher et al., 2017) (Methods S1). A crossover rate in cM/Mb was estimated based on physical positions of 9K SNPs and the consensus genetic map of Muñoz-Amatriaín et al. (2011) using a Python script.

Exome resequencing data

We generated exome resequencing from 62 landrace accessions from a randomly chosen subset of the 804 landraces, of which 37 are 6-row spring barley and the remaining 25 are 2-row spring accessions (Table S2). All sequences were submitted to the National Center for Biotechnology Information's Sequence Read Archive associated with BioProject numbers PRJNA473780 and PRJNA488050. The data were processed using publicly available software integrated with bash scripts in the "sequence_handling" workflow (Hoffman et al., 2018). Detailed methods are in Methods S1. Variant calling is similar to that reported by Kono et al. (2016). Parameters used here are in Methods S1. Variant annotation was performed using ANNOVAR (Wang et al., 2010) with gene models provided by Mascher et al. (2017).

Bioclimatic and geographic variables

WorldClim bioclimatic data at a resolution of 2.5 minutes were downloaded on Oct 7, 2015 using the getData 'raster' R function (Hijmans et al., 2016) in the R statistical language (Team, 2017). The latitude, longitude, elevation, and BIO1 to BIO19 values of the collection locations for 784 landraces are given in the phenotype data file (Supplemental data1). Environmental variables can be divided into three categories. The latitude, longitude, and elevation were classified as geographic factors, BIO1 to BIO11 were classified as temperature, and BIO12 to 19 were classified as precipitation. In order to determine if consolidation of components identifies novel variants, the top three Independent Components (ICs) calculated from BIO1 to BIO19 values after standardization for each BIO variable and treated additional phenotypes. This was performed using the icaimax 'ica' function (ICA1 to ICA3) from the ica-package in R (Bell and Sejnowski, 1995) (Supplemental data 1).

Environmental association mapping

Quality control for genotyping data was based on two criteria: (1) culling of SNPs with ≥ 20% missingness; and (2) culling SNPs to include only on copy of SNPs in complete LD (Supplemental dataset 2).-The analysis was performed using a compressed mixed linear model (Zhang et al., 2010) implemented in the Genome Association and Prediction Integrated Tool (GAPIT) R package (Lipka et al., 2012).

F_{ST} estimation

To compare allele frequency differentiation in partitions of the data we calculated F-statistics (Wright, 1949) for individual SNPs using the measure of Weir and Cockerham (1984) as implemented in the R package HierFstat (de Meeûs and Goudet, 2007). We calculate F_{ST} for five partitions of the data. The elevation comparison uses a cutoff of 3,000 m to delineate high elevation accessions. This includes accessions from the European Alps, the Caucasus, Himalayan, and Hindu Kush Mountain regions, and the Ethiopian Plateau. We compared allele frequencies at two latitudinal ranges: (1) within the wild range of the species (30˚ N – 40˚ N) versus landraces at latitudes higher than 40˚ N (high latitude), and (2) within the wild range of the species (30˚ N – 40˚ N) versus landraces at latitudes lower than 30˚ N (low latitude). High latitude includes roughly the northern extent of the range of wild barley and extends across Eurasia from roughly the Central Iberian Peninsula to Northern Japanese Archipelago. Low latitude includes roughly the southern extent of the range of wild barley. This extends from northwestern Africa to just south of the Japanese Archipelago. We compared allele frequency divergence between accessions with winter and spring growth habit. For comparison of the relative degree of differentiation among partitions of the sample, we also calculate F_{ST} for longitude, dividing the eastern and western portion of the range at the 48th meridian. To account for differences in sample size among partitions of the data (Bhatia et al., 2013), we used resampling with equalized sample numbers and 10 iterations without replacement. F_{ST} estimates for each SNP were averaged across 10 iterations and outliers were identified at the 99th percentile of the distribution. In order to calculate the p-value for each F_{ST} value we performed 1,000 permutations. The details can be found in Methods S1. Identification of previously reported loci related to cold, drought tolerance, and flowering time

A literature search was used to identify genes previously reported to contribute to flowering time and cold or drought tolerance (Table S3-5). Google Scholar searches were performed with the terms “cold tolerance” or “freezing tolerance” or "drought tolerance" or "flowering time" and “plant" or “barley.” The protein-coding sequence (CDS) of identified genes were used as the query or the subject in BLASTN against the barley high confident CDS in May 2016 on the IPK Barley BLAST Server (Mascher et al., 2017). Barley genes and their interval

information were extracted if combined "Score", "Identity", "Percentage", and "Expectation" produced the overall highest ranking and the "Query length" is >100 bp. In the event of identical scores, all highest ranked hits were extracted.

The bedops 'closest-features' function (Neph et al., 2012) was used to compare the locations of SNPs and gene intervals. Specifically, if the SNPs were located in the gene interval or 10 Kb up or downstream of the genes, we considered those SNPs as identifying the closest gene.

Linkage disequilibrium around SNPs

For each 9K SNP identified in environmental association analysis or among F_{ST} outliers, we calculated linkage disequilibrium (LD) with surrounding SNPs. We focused on 200 Kb windows, 100 Kb upstream and downstream of the focal SNP. We used both genotyping data and exome capture resequencing for the comparison. When the 9K SNP was also genotyped by exome capture, it was used for the LD analysis. If the 9K SNP was not present in exome capture, we performed the LD analysis with the SNP in closest proximity with a similar minor allele frequency. We filtered out SNPs using a minor allele frequency threshold of 1% for both environmental association analysis and F_{ST} outliers. For LD analysis, filtering of variants could be anti-conservative, thus for this analysis we removed SNPs with ≥ 50% missing data. We used the R package, LDheatmap (Shin et al., 2006), to calculate r^2 (Lewontin, 1988).

Inference of ancestral state

The ancestral state for each SNP from both 9K (Supplemental data3) and resequencing datasets (Supplemental data4) was inferred using whole genome resequencing data from *Hordeum murinum* ssp. *glaucum* (Kono et al., 2018) with the programs ANGSD and ANGSD-wrapper (Korneliussen et al., 2014; Durvasula et al., 2016). Methods are detailed in Kono et al. (2018). Both minor and derived allele frequency were calculated using a Python script.

Haplotype analysis for individual genes

To assess evidence for functional diversity in the immediate vicinity of SNPs identified in our analysis, we examined haplotype-level diversity in loci flanking associations. We used exome capture resequencing from the panel described above. Overlapping SNP genotyping was extracted from SNP calls in a variant call format (VCF) file using `vcf-intersect` from vcflib (https://github.com/vcflib/vcflib). Missing genotypes were imputed using PHASE (Stephens et al., 2001; Stephens and Scheet, 2005) with Plink 1.9 (Chang et al., 2015) to convert the VCF format into PHASE format. Homozygotes were treated as haploid and heterozygotes were treated as diploid for haplotype identification.

Data and code availability

Supporting data (Supplemental data 1-7), including a VCF file of exome capture SNPs and the full results for environmental association and F_{ST} are available at https://conservancy.umn.edu. Scripts for analysis and associated files are available on a project Github site at https://github.com/MorrellLAB/Env_Assoc.

Results and Discussion

Summary of genotyping and resequencing data

After quality filtering of the genotyping and resequencing data and exclusion of landrace accessions without discrete locality information, our SNP genotyping dataset includes 6,152 SNPs in 784 accessions (Fig. S1; Table S1). The exome resequencing includes 482,714 SNPs in 62 samples (Fig. S1; Table S2). The relative frequency of SNPs in both panels is shown in Fig. S2.

Environmental association and F_{ST} outliers

The environmental association with bioclimatic and geographic variables identified 158 SNPs in significant associations (with p-value $< 5 \times 10^{-4}$ threshold, 0.15th percentile) (Fig. S3; Table S6). F_{ST} comparisons for elevation, low and high latitude, and growth habit identified 241 outliers (with the 99th percentile of F_{ST} as a threshold) (Fig. 1; Fig. S3; Table S7). Considering both the environmental association and F_{ST} comparisons, we identify a total of 387 unique SNPs putatively associated with environmental adaptation in our genotyping panel (Fig. S3).

Environmental associations and F_{ST} outliers shared 12 SNPs in 11 annotated genes. The only characterize gene found in both analyses is the *HvPhyC* locus in barley. For details regarding overlapping results see Table S8.

Previously reported loci associated with environmental adaptation

We identified 12 SNPs located in seven previously reported loci, which were characterized as altering flowering time or being involved in cold or drought tolerance, and thus potentially contribute to geographic adaptive differences in barley. This includes four well characterized loci that contribute to flowering time including *HvPpd-H1* (Turner et al., 2005; Jones et al., 2008), *HvPhyC* (Nishida et al., 2013), *HvVrn-H1* (*HvAP1*) (Cockram et al., 2007), and *HvPRR1* (*HvTOC1*) (Ford et al., 2016) (Table 1). These loci are identified as the upper 99%

of F_{ST} values within a comparison. A 97.5% cutoff includes two additional barley flowering time loci, *HvPpd-H2* (*HvFT3*) (Casao et al., 2011) and *HvELF3 (Esp1L*/*eam8)* (Boden et al., 2014) (Table 1). Using a percentile cutoff for the p-value from the environmental association identifies four additional flowering time loci, *HvCMF4* (Cockram et al., 2012)*, HvCO1* (Campoli et al., 2012a)*, HvPRR95* (Cockram et al., 2012)*,* and *HvVrn-H3* (*HvFT1*) (Cockram et al., 2007). We also identified two loci previously reported as contributing to cold adaptation in barley, *HvCbf4B* (Stockinger et al., 2007) and *HvICE2* (Skinner et al., 2006). The 97.5% threshold for F_{ST} includes a characterized locus, *HvSS1* (Barrero-Sicilia et al., 2011). *HvCbf4B, HvPpd-H1,* and *HvVrn-H1* were reported as F_{ST} outliers in comparisons of North American breeding populations (Poets et al., 2015b)*. HvCbf4* was also found as F_{ST} outliers and association with temperature adaptation in comparisons of wild barley populations (Fang et al., 2014). A 1.5th percentile cutoff we identify two additional genes related to barley cold tolerance, *HvDhn13* (Choi et al., 1999) and *HvWRKY38* (Mare et al., 2004). We also found that *HvDhn7*, a locus previously reported as potentially associated with cold and drought tolerance (Choi et al., 1999), has large allele frequency difference in the longitudinal comparison, consistent with population structure reported in a resequencing panel (Morrell et al., 2005). A 1.5th percentile cutoff includes barley drought tolerance-related loci *HvDhn5* (Suprunova et al., 2004)and *HvWRKY38* (Mare et al., 2004)*.*

A further four variants identified in our barley panel had previously been associated with four potential adaptive loci in other plant species (Table 1). These include the flowering time-related loci characterized in *Arabidopsis thaliana, AtCOP1* (Xu et al., 2016) involved in cold stress resistance and the rice (*Oryza sativa*) locus *OsiSAP8* associated with cold, drought, and salt stress response (Kanneganti and Gupta, 2008). The 97.5% threshold for F_{ST} includes an additional eight loci previously characterized in other species (Table 1). A 1.5th percentile cutoff of p-value from environmental association identifies four additional cold tolerance-related loci in other plant species and eight loci related drought intolerance related loci (Table S9). The identification of multiple characterized loci between the 97.5th and 99th percentile for F_{ST} are indicative of the trade-off between false discovery and false negative rate in empirical scans for adaptive variation (see for exampleTeshima et al., 2006)

Relative Differentiation Among Partitions of the Sample

Of five F_{ST} comparisons, the samples stratified by longitude showed the greatest differentiation in allele frequency, with average genome-wide F_{ST} = 0.098 (± 0.11) (Supplemental data 5). A primary partitioning of barley populations by longitude, into eastern

and western populations has been reported previously (Morrell and Clegg, 2007; Saisho and Purugganan, 2007; Poets et al., 2015a). Samples above and below 3,000 m showed the second highest average F_{ST} = 0.069 (± 0.086) (Supplemental data 5). The high degree of allele frequency divergence may result in part from the relatively small sample size of high elevation accessions, with only 54 occurring above 3,000 meters (Fig. S4). The third largest average F_{ST} = 0.067 (± 0.086) was found for the low latitude comparison, followed by the high latitude comparison with average F_{ST} of 0.059 (± 0.073) (Supplemental data 5). Comparison between samples from spring and winter growth habit yields a lower average F_{ST} = 0.043 (± 0.057) (Supplemental data 5). Poets et al. 2015 applied a series of F_{ST} to North American barley breeding populations and also found limited differentiation between spring and winter growth habit.

F_{ST} outliers from geographic patterns and growth habit

For F_{ST} outlier we focus on comparisons most directly linked to climatic differentiation, high and low latitude, elevation, and growth habit. By taking the upper 1% tail of F_{ST} values from each comparison, each yield 62 outlier SNPs (Fig. 1; Fig. S3; Table S7) With the exception of high and low latitude, the comparisons tend to identify unique sets of variants. The two latitude comparisons share six of 124 SNPs (Fig. S5). There is no overlap in SNPs identified by growth habit with SNPs found in the high latitude, low latitude, or elevation comparisons (Fig. S5). It should be noted that winter barleys are less frequently grown at higher latitudes or elevations due to winter conditions, and indeed winter barleys from these locations relatively uncommon in the sample, thus constraining the comparisons (Table S1). The elevation comparison identified the largest number of previously characterized loci, including *HvPhyC*, *HvICE2*, *OSiSAP8* (Figure 1).

As expected, SNPs with the most extreme F_{ST} values for elevation, growth habit, and latitude comparisons formed very distinctive geographic patterns. For each comparison, the highest F_{ST} values occur in SNPs that fall within genes that are annotated but uncharacterized. The highest F_{ST} from the high latitude comparison occurred at SNP 12_30191 with the F_{ST} = 0.484 (p-value = 0). For this SNP, the ancestral allele dominates the wild barley distributed range, while the derived allele is more prevalent in higher latitude regions (Fig. 2a). This SNP hit the gene *HORVU7Hr1G084420* (RAS-related GTP-binding) (Table S7). The highest F_{ST} from the low latitude comparison is for the SNP SCRI_RS_170209 with the F_{ST} = 0.517 (p-value = 0). The ancestral state dominates the within the geographic range of wild barley and higher latitudes, while the derived allele is more prevalent in lower latitude regions (Fig. S6a). This SNP hit the gene *HORVU2Hr1G123480* (CCR4-NOT transcription complex subunit8) (Table

S7). The highest F_{ST} between samples from elevation comparison is for SNP (12_11529) with F_{ST} = 0.649 (p-value = 0). For this SNP the ancestral allele occurs at high elevations, such as the Himalaya Mountains, while the derived allele tends to occur at lower elevation (Fig. 2b). This SNP falls within the gene, *HORVU1Hr1G079140 (*Pectin lyase-like superfamily protein) (Table S7). The highest F_{ST} between samples from the growth habit comparison is for SNP SCRI_RS_148782 with F_{ST} = 0.395 (p-value = 0.395). For this SNP, the ancestral allele tends to occur in high elevation or latitude regions, while the derived allele is more widespread (Fig. S6b). This SNP falls within the gene *HORVU4Hr1G088440* (Methionyl-tRNA formyltransferase) (Table S7).

Environmental association to bioclimatic variables

We identified 159 unique SNPs significantly associated with at least one environmental factor with the threshold of p-value < 5 x 10^{-4} (Fig. S3; Table S6). All of the p-values and Benjamini-Hochberg FDR-values are reported in Supplemental data 6.

For individual environmental variables, we found 71 SNPs associated with precipitation (variables BIO12 to BIO19) and 43 with temperature (all variables from BIO1 to BIO11) (Fig. S3). We identified 54 SNPs associated with geographic variables (latitude, longitude, and elevation), and 26 associated with independent components (top three independent components calculated from BIO1 to BIO19 values after standardization for each BIO variable, called ICs) (Fig. S3). We also identified 47 cases where individual SNPs were associated with more than one environmental variable (Fig. S6). But more generally, as with the F_{ST} comparisons, the environmental variable tends to associate with a unique set of SNPs sets (Fig. S6). The largest proportion of unique SNPs were found for precipitation (30.38%), followed by geographic variables (29.75%), temperature variables (18.35%), and then ICs (5.7%) (Fig. S6). Perhaps as expected, the aggregated independent components generally did not identify novel variants.

Minor allele frequency of identified SNPs

The SNPs identified by environmental association and F_{ST} have relatively high minor allele frequency (MAF). For significant environmental associations, average MAF = 0.251 (with standard deviation of ± 0.133). For F_{ST} outliers average MAF = 0.288 (± 0.114). Average MAF = 0.262 (± 0.1396) for the full SNP data set. For association analysis, MAF limits the potential to associate genotype in phenotype. This contrasts with expectations that variants that are

adaptive for less frequently occupied habitats, such as high elevation sites of cultivation, should be relatively uncommon (average MAF=0.240 (± 0.091) of outliers from elevation comparison). This is true whether a variant that is adaptive to particular environment is conditionally neutral or exhibits antagonistic pleiotropy (Tiffin and Ross-Ibarra, 2014).

SNP density and LD near focal SNPs

As previously reported, SNP density is highest on chromosome arms and lower in pericentromeric regions (Muñoz-Amatriaín et al., 2015; Mascher et al., 2017). This trend is particularly evident for 9K SNPs (Figure 3b; and Figure S1), and is broadly consistent with lower SNP density in genomic regions with lower observed rates of crossing over (Muñoz-Amatriaín et al., 2011). Exome capture density is also reduced in pericentromeric regions, such that 51,567 SNPs are detected in 1.560 Gb in pericentromere regions (33 SNPs/Mb) versus 431,147 SNPs in 3.02 Gb (143 SNPs/Mb) on chromosome arms.

We compared LD at focal SNPs to the surrounding region for 387 SNPs identified by environmental association analysis or as F_{ST} outliers. For 102 (25.5%) LD with an $r^2 > 0.45$ (90th percentile) was limited to SNPs within the same gene (Fig. S8a & b; Table 4). Detectable LD with flanking loci is limited to pericentromeric regions, because the locus tested is often the only annotated gene within the 200 kb window (Fig. 3c). For the remaining 257 SNPs, LD extends well beyond the locus where the initial association was identified (Fig. S8 c & d). The potential to identify individual loci likely to contribute to adaptive phenotypes is clearly limited by both gene density and recombination rate variation (Figure 3b, Fig. S1).

SNP density is a limitation in our study. With ~40,000 annotated genes in the barley genome (Mascher et al., 2017), roughly one in six genes was directly genotyped. Based on our own analysis of LD relative to genotyped SNPs, for roughly 1/4 of SNPs, there is limited LD with nearby loci. There are likely several contributing factors. First, in regions of the genome with higher crossing over rates and higher gene density, LD can be limited beyond the locus containing the genotyped SNP (Fig. 3c and Fig. S1). Second in regions with limited crossing over, gene density is low. LD would typically have to extend hundred of kilobases between genotyped SNPs and a causative variant at another locus to create an association. Third, the MAF of genotyped variants is relatively high (Fig. S2). High frequency variants have had more time to be subject to recombination and thus can be closer to linkage equilibrium (Nordborg and Tavaré, 2002).

Loci not identified by environmental association or as F_{ST} outliers

Among our comparisons, we identified many of the loci previously reported to contribute to cold or drought tolerance and flowering time. Among loci not detected is *HvCbf3* which shows limited evidence for a contribution to cold tolerance (Choi et al., 2002; Stockinger et al., 2007). Also, although we do identify the *HvVrn-H1 (HvAP1)* (Cockram et al., 2007) locus associated with vernalization and growth habit, we do not identify *HvVrn-H2* (*HvZCCT-Ha/b/c*) (Karsai et al., 2005). The *HvCbf3* gene was genotyped by two SNPs and *HvVrn-H2* by one SNP, but no evidence of association with adaptive phenotypes was found. Prior evidence of the contribution of *HvCbf3* to cold tolerance is equivocal (Choi et al., 2002; Stockinger et al., 2007). *HvVrn-H2* would most likely be identified as difference in growth habit, but F_{ST} values do not exceed 99th percentile threshold (0.002, -0.006, -0.003, and 0.0009). More generally, a challenge in both our forms of analysis is that even if they occur in the vicinity of a causative variant, individual SNPs may not occur on the same haplotype associated with a particular adaptive phenotype (see Nordborg and Tavaré, 2002).

Putative structural variation

For the elevation comparisons, 15 outliers among the 9K SNPs were identified with F_{ST} of ~0.40. All occur on chromosome 5H at the consensus genetic map position of 663.25 cM(Muñoz-Amatriaín et al., 2011). These SNPs span a physical distance of ~98 Mb (Table S6). As expected based on F_{ST} values, the minor allele frequencies of these SNPs are very similar (0.3544 - 0.3619), with minor alleles occurring in the same individuals in almost all cases. All 12 SNPs are in complete LD. This region of 5H, between 167.3 Mb and 260.0 Mb, overlaps with a region identified as a putative chromosomal inversion in wild barley (Fang et al., 2014). The SNPs that Fang et al. (2014) associated with the putative inversion occur between 126.7 Mb and 305.5 Mb. Evidence for an inversion in wild barley was based on elevated F_{ST} values, extended LD, and enrichment for environmental associations. Fang et al. (2014) reported a similar pattern on chromosome 2H in wild barley at positions that correspond to 267.3 Mb to 508.7 Mb. On 2H, we find less evidence of allele frequency differentiation than in wild barley; observing two SNPs, spanning ~494 Kb with F_{ST} = 0.33 in our landrace sample (Table S10). Chromosomal inversions can serve to maintain an association between variants at multiple loci that contribute to local adaptation (Dobzhansky and Sturtevant, 1938; Kirkpatrick and Barton, 2006). The putative inversion on chromosome 5H in barley appears to contribute adaptation to high elevation. Recent studies have identified putative chromosomal inversions that contribute to elevation and temperature gradients in teosinte and maize (Fang et al., 2012; Hufford et al., 2012; Pyhäjärvi et al., 2013), rainfall regime and annual versus perennial growth habit in

Mimulus guttattus (Sweigart and Willis, 2003; Lowry and Willis, 2010), and temperature and precipitation differences in wild barley (Fang et al., 2014).

Haplotype analysis at individual genes

Environmental association results identified a SNP SCRI_RS_137464 assignificantly associated with BIO8 (Mean temperature of Wettest Quarter) (p-value = 1.059×10^{-4}), which is in the *HvPRR1/HvTOC1* gene (Fig. 4a). *TOC1* is an important component of the circadian clock in *Arabidopsis* with a crucial function in the integration of light signals to control circadian and morphogenic responses, which is closely related to flowering time (Más et al., 2003). *HvPRR1/HvTOC1* is the ortholog of *TOC1* in *Arabidopsis thaliana*, and has a high level of sequence similarity and conservation of diurnal and circadian expression patterns, when compared to *TOC1* in *Arabidopsis* (Campoli et al., 2012b). Based on exome capture sequencing from 61 landraces, there were 48 SNPs identified, including SCRI_RS_137464. Four of five nonsynonymous SNPs are in the last exon of the gene (Fig. 4b & c). Five SNPs within *HvPRR1/HvTOC1* have relative strong LD with SCRI_RS_137464 (r^2 >0.45) located in the same gene (Fig. 4 a &b). Resequencing identified 20 haplotypes with the two variants at the SCRI_RS_137464 each occurring in 10 haplotypes (Fig. 4c) with no obvious geographic pattern.

Environmental association of both "minimum temperature of coldest month (BIO6) and "mean temperature of the coldest quarter" (BIO11) identified an association on chromosome 3H with SNP 11_10380 (p-value = 1.77×10^{-5}). The SNP is in barley gene *HORVU3Hr1G030150.1*, which is an ortholog of wheat gene *WCI16* (Wheat Cold Induced 16) (Sasaki et al., 2013) (Fig. S9). The derived alleles for genotyped SNP at this locus are much more common in landrace barley than in wild lines. In published genotyping wild barley [Fang et al. 2014], the minor allele at 11_10380 occurs in four accessions with geographic provenance information. Those accession occur at an average of 1460 m, near the upper end of the elevational range for wild barley. Estimated derived allele frequencies differ considerably in wild and landraces, with at 0.0072 and 0.13 respectively. The 200 Kb window surrounding the SNP contains one genes in addition to *WCI16* (Fig. S9a). *WCI16* encodes a putative transcription factor involved in stomata development. It represents a novel class of late embryogenesis abundant (LEA) proteins in response to cellular dehydration and is involved in freezing tolerance (Sasaki et al., 2013). When transformed into *Arabidopsis thaliana*, *WCI16* improves freezing tolerance in wheat and increases cold temperature tolerance (Sasaki et al., 2013). Based on exome capture sequencing from 61 landraces, there were six SNPs identified, including 11_10380. Three of six SNPs, including 11_10380, are in noncoding sequence (Fig. S9b). Of the three SNPs observed

in coding regions, one is a nonsynonymous change at nucleotide position of 119 a change among amino acids with similar properties (V to L). There is no evidence of LD between this SNP and others within a 200 Kb window (Fig. S 9a). Exome capture resequencing identified eight haplotypes, with three occurring at high frequency, seven haplotypes predominate at lower elevation and lower latitude, with two of those occurring most frequently. The haplotype that includes the genotyped SNP is associated with low elevation and low latitude, but the single nonsynonymous SNP found in resequencing is not clearly associated with any geographic pattern (Fig. S 9c).

Environmental association analysis suggested that the SNP SCRI_RS_235243 (P-value=3.28418×10^{-4}) significant associated with BIO14 (Precipitation of Driest Months), hit the barley gene *HORVU1Hr1G008120.1*, the ortholog gene of which in *Arabidopsis thaliana* and bread wheat is dehydroascorbate reductase (*DHAR*; EC 1.8.5.1), one of two important enzymes functioning in the regeneration of ascorbate (AsA) and thus plays a role in protection against oxidative stress (Eltayeb et al., 2006; Osipova et al., 2011). The 200 Kb window surrounding the genotyped SNP SCRI_RS_235243 contains four genes in additional to *DHAR*, including two with exome capture sequence coverage (Fig. S10a). Previous results suggested that over expression of *DHAR* can protect plants against drought, salt, and polyethylene glycol-induced stress in tobacco and bread wheat (Eltayeb et al., 2006; Osipova et al., 2011). Resequencing identified 53 SNPs in our panel, including SCRI_RS_235243. This included 28 SNPs in noncoding regions, 14 synonymous, and 11 nonsynonymous (Fig. S10 b & c). SCRI_RS_235243 is one of nine nonsynonymous SNPs in the first exon of *DHAR* gene (Fig. S10 b &c). Six SNPs are in high LD with SCRI_RS_235243 ($r^2 > 0.45$), all are noncoding variants within *DHAR* (Fig. 10b & c). The derived variant at SCRI_RS_235243 occurs on two haplotypes (Fig. S10c) that occur in high latitude regions.

At all three of the loci, a putative causative variants is not immediately apparent. However, as two of the loci noted above, *HvPRR1/HvTOC1* and *DHAR,* demonstrate barley landraces are frequently segregating for a abundance of potentially functional variants.

Conclusions

Based on genotyping from 6,152 SNPs in 784 barley landraces from across the pre-historic range of the species in the Old World, we detect environmental associations to bioclimatic variables or identify allele frequency differentiation at 11 loci with prior evidence of contribution to climatic adaption in plants (Table 1). This includes well characterized loci that contribute to

flowering time or drought or cold adaptation in barley including *HvPpd-H1(HvPRR37)* (Turner et al., 2005; Jones et al., 2008), *HvPhyC* (Nishida et al., 2013), and *HvCbf4B*(Stockinger et al., 2007), and *HvICE2* (Skinner et al., 2006). We also identify 357 loci in environmental association or as allele frequency differentiation outliers that were not previously reported as associated with adaptive phenotypes in other plant species.

Our study benefits from sample size. Russell et al. (2016) performed environmental association with 1,688,807 SNPs from exome capture resequencing in 137 cultivated samples. While the analysis identified 12 loci associated with flowering time, many other previously reported genes went undetected, prompting the authors to suggest a lack of power owing to small sample size (Russell et al., 2016). Despite limited SNPs density and the sampling of relatively common variants, our comparative analysis identified a number of previously identified barley loci and many plausible candidate loci from other plant species. Environmental association is a promising approach to explore the adaptive variants for low temperature and drought stress in barley. Better coverage of gene space through exome capture or whole genome resequencing in a relatively deep panel of accessions would likely uncover a much more comprehensive set of variants contributing to environmental adaptation. This could contribute to targeted use of variants for adaptation to environmental and climatic conditions for barley breeding and germplasm improvement, with the potential to improve the understanding of loci that contribute to adaptation in wheat and other cereals.

Acknowledgements

We would like to thank A Proulx for annotation of physical positions of 9K SNPs. E Vonderharr assisted with SRA submissions. This study was supported by the U.S. NSF Plant Genome Program (IOS-1339393) to PLM and USDA Triticeae Coordinated Agricultural Project 2011-68002-30029 to GJM and PLM, and NSF (MCB-1518058) to FK. This research was carried out with hardware and software support provided by the Minnesota Supercomputing Institute (MSI) at the University of Minnesota.

Author contributions

PLM, AMP, FK, and LL designed the project. GJM generated the exome-capture resequencing data. RMT and BGS assisted with genotyping data quality control. LL, FK, CL, SRW, CC, XL, and AMP analyzed the data. LL, CL, and PLM wrote the manuscript.

References

Abu-Romman SM, Ammari TG, Irshaid LA, Salameh NM, Hasan MK, Hasan HS 2011. Cloning and expression patterns of the *HvP5CS* gene from barley (*Hordeum vulgare*). *Journal of Food, Agriculture and Environment* **9**: 279–284.

Anderson JE, Kono TJY, Stupar RM, Kantar MB, Morrell PL 2016. Environmental association analyses identify candidates for abiotic stress tolerance in *Glycine soja*, the wild progenitor of cultivated soybeans. *G3: Genes, Genomes, Genetics* **6**: 835–843.

Barrero-Sicilia C, Hernando-Amado S, González-Melendi P, Carbonero P 2011. Structure, expression profile and subcellular localisation of four different sucrose synthase genes from barley. *Planta* **234**: 391–403.

Beaumont MA, Balding DJ 2004. Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**: 969–980.

Bell AJ, Sejnowski TJ 1995. An information-maximization approach to blind separation and blind deconvolution. *Neural computation* **7**: 1129–1159.

Bhatia G, Patterson N, Sankararaman S, Price AL 2013. Estimating and interpreting F_{ST}: the impact of rare variants. *Genome Research* **23**: 1514–1521.

Boden SA, Weiss D, Ross JJ, Davies NW, Trevaskis B, Chandler PM, Swain SM 2014. *EARLY FLOWERING3* regulates flowering in spring barley by mediating gibberellin production and *FLOWERING LOCUS T* expression. *The Plant Cell* tpc. 114.123794.

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

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

Casao MC, Karsai I, Igartua E, Gracia MP, Veisz O, Casas AM 2011. Adaptation of barley to mild winters: a role for *PPDH2*. *BMC Plant Biology* **11**: 164.

Cavalli-Sforza LL 1966. Population structure and human evolution. *Proceedings of the Royal Society of London Series B: Biological Sciences* **164**: 362–379.

Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**: 7.

Choi D-W, Zhu B, Close TJ 1999. The barley (*Hordeum vulgare* L.) dehydrin multigene family: sequences, allele types, chromosome assignments, and expression characteristics of 11 *Dhn* genes of cv Dicktoo. *Theoretical and Applied Genetics* **98**: 1234–1247.

Choi D-W, Rodriguez EM, Close TJ 2002. Barley *Cbf3* gene identification, expression pattern, and map location. *Plant Physiology* **129**: 1781–1787.

Cockram J, Chiapparino E, Taylor SA, Stamati K, Donini P, Laurie DA, O'Sullivan DM 2007. 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**: 993–1001.

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

Comadran J, Kilian B, Russell J, Ramsay L, Stein N, Ganal M, Shaw P, Bayer M, Thomas W *et al.* 2012. Natural variation in a homolog of *Antirrhinum CENTRORADIALIS* contributed to

spring growth habit and environmental adaptation in cultivated barley. *Nature Genetics* **44**: 1388–1392.

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

de Meeûs T, Goudet J 2007. A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. *Infection, Genetics and Evolution* **7**: 731–735.

Dobzhansky T, Sturtevant AH 1938. Inversions in the chromosomes of *Drosophila pseudoobscura*. *Genetics* **23**: 28.

Durvasula A, Hoffman PJ, Kent TV, Liu C, Kono TJY, Morrell PL, Ross-Ibarra J 2016. ANGSD-wrapper: utilities for analysing next-generation sequencing data. *Molecular Ecology Resources* **16**: 1449–1454.

Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, González-Martínez SC, Neale DB 2010. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* **185**: 969–982.

Eltayeb AE, Kawano N, Badawi GH, Kaminaka H, Sanekata T, Morishima I, Shibahara T, Inanaga S, Tanaka K 2006. Enhanced tolerance to ozone and drought stresses in transgenic tobacco overexpressing dehydroascorbate reductase in cytosol. *Physiologia Plantarum* **127**: 57–65.

Fang Z, Gonzales AM, Clegg MT, Smith KP, Muehlbauer GJ, Steffenson BJ, Morrell PL 2014. Two genomic regions contribute disproportionately to geographic differentiation in wild barley. *G3: Genes, Genomes, Genetics* **4**: 1193–1203.

Fang Z, Pyhäjärvi T, Weber AL, Dawe RK, Glaubitz JC, González JDJS, Ross-Ibarra C, Doebley J, Morrell PL *et al.* 2012. Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics* **191**: 883–894.

Ford B, Deng W, Clausen J, Oliver S, Boden S, Hemming M, Trevaskis B 2016. Barley (*Hordeum vulgare*) circadian clock genes can respond rapidly to temperature in an *EARLY FLOWERING 3*-dependent manner. *Journal of Experimental Botany* **67**: 5517–5528.

Francia E, Rizza F, Cattivelli L, Stanca AM, Galiba G, Toth B, Hayes PM, Skinner JS, Pecchioni N 2004. Two loci on chromosome 5H determine low-temperature tolerance in a 'Nure'(winter)×'Tremois'(spring) barley map. *Theoretical and Applied Genetics* **108**: 670–680.

Hansson M, Komatsuda T, Stein N, Muehlbauer GJ 2018. Molecular mapping and cloning of genes and QTLs. eds. *The Barley Genome*. Springer, 139–154.

Harlan JR, Zohary D 1966. Distribution of wild wheats and barley. *Science* **153**: 1074–1080.

Harris DR, Gosden C 1996. The beginnings of agriculture in western Central Asia. *The origins and spread of agriculture and pastoralism in Eurasia* 370–389.

Hayes PM, Blake T, Chen THH, Tragoonrung S, Chen F, Pan A, Liu B 1993. Quantitative trait loci on barley (*Hordeum vulgare* L.) chromosome 7 associated with components of winterhardiness. *Genome* **36**: 66–71.

Hijmans RJ, Etten JV, Cheng J, Mattiuzzi M, Sumner M, Greenberg JA, Lamigueiro OP, Bevan A, Racine EB *et al.* 2016. raster: Geographic Data Analysis and Modeling.

Hoffman PJ, Wyant SR, Kono TJY, Morrell PL 2018. MorrellLab/sequence_handling: Release v2.0: SNP calling with GATK 3.8.

Honsdorf N, March TJ, Berger B, Tester M, Pillen K 2014. High-throughput phenotyping to detect drought tolerance QTL in wild barley introgression lines. *PLoS One* **9**: e97047.

Hufford MB, Xu X, Van Heerwaarden J, Pyhäjärvi T, Chia J-M, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE *et al.* 2012. Comparative population genomics of maize domestication and improvement. *Nature Genetics* **44**: 808.

Jones H, Leigh FJ, Mackay I, Bower MA, Smith LMJ, Charles MP, Jones G, Jones MK, Brown TA *et al.* 2008. Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the Fertile Crescent. *Molecular Biology and Evolution* **25**: 2211–2219.

Kanneganti V, Gupta AK 2008. Overexpression of *OsiSAP8*, a member of stress associated protein (SAP) gene family of rice confers tolerance to salt, drought and cold stress in transgenic tobacco and rice. *Plant Molecular Biology* **66**: 445–462.

Karsai I, Szűcs P, Mészáros K, Filichkina T, Hayes PM, Skinner JS, Láng L, Bedő Z 2005. 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**: 1458–1466.

Kirkpatrick M, Barton N 2006. Chromosome inversions, local adaptation and speciation. *Genetics* **173**: 419–434.

Knox AK, Dhillon T, Cheng H, Tondelli A, Pecchioni N, Stockinger EJ 2010. *CBF* gene copy number variation at *Frost Resistance-2* is associated with levels of freezing tolerance in temperate-climate cereals. *Theoretical and Applied Genetics* **121**: 21–35.

Kono TJ, Fu F, Mohammadi M, Hoffman PJ, Liu C, Stupar RM, Smith KP, Tiffin P, Fay JC *et al.* 2016. The role of deleterious substitutions in crop genomes. *Molecular Biology and Evolution* **33**: 2307–2317.

Kono TJY, Liu C, Vonderharr EE, Koenig D, Fay JC, Smith KP, Morrell PL 2018. Estimating the relative contribution of deleterious and neutral SNPs to agronomic phenotypes. *in preparation*

Korneliussen TS, Albrechtsen A, Nielsen R 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**: 356.

Lewontin RC 1988. On measures of gametic disequilibrium. *Genetics* **120**: 849–852.

Lewontin RC, Krakauer J 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**: 175–195.

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z 2012. GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**: 2397–2399.

Lotterhos KE, Whitlock MC 2014. Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology* **23**: 2178–2192.

Lowry DB, Willis JH 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS biology* **8**: e1000500.

Mare C, Mazzucotelli E, Crosatti C, Francia E, Cattivelli L 2004. *Hv-WRKY38*: a new transcription factor involved in cold-and drought-response in barley. *Plant Molecular Biology* **55**: 399–416.

Más P, Alabadí D, Yanovsky MJ, Oyama T, Kay SA 2003. Dual role of *TOC1* in the control of circadian and photomorphogenic responses in *Arabidopsis*. *The Plant Cell* **15**: 223–236.

Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE *et al.* 2017. A chromosome conformation capture ordered sequence of the barley genome. *Nature* **544**: 427–433.

Morrell PL, Clegg MT 2007. Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proceedings of the National Academy of Sciences* **104**: 3289–3294.

Morrell PL, Toleno DM, Lundy KE, Clegg MT 2005. Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. *Proceedings of the National Academy of Sciences* **102**: 2442–2447.

Muñoz-Amatriaín M, Cuesta-Marcos A, Endelman JB, Comadran J, Bonman JM, Bockelman HE, Chao S, Russell J, Waugh R *et al.* 2014. The USDA barley core collection: genetic diversity, population structure, and potential for genome-wide association studies. *PLOS ONE* **9**: e94688.

Muñoz-Amatriaín M, Moscou MJ, Bhat PR, Svensson JT, Bartoš J, Suchánková P, Šimková H, Endo TR, Fenton RD *et al.* 2011. An improved consensus linkage map of barley based on flow-sorted chromosomes and single nucleotide polymorphism markers. *The Plant Genome* **4**: 238–249.

Muñoz-Amatriaín M, Lonardi S, Luo M, Madishetty K, Svensson JT, Moscou MJ, Wanamaker S, Jiang T, Kleinhofs A *et al.* 2015. Sequencing of 15 622 gene-bearing BAC s clarifies the gene-dense regions of the barley genome. *The Plant Journal* **84**: 216–227.

Nei M, Maruyama T 1975. Letters to the editors: Lewontin-Krakauer test for neutral genes. *Genetics* **80**: 395.

Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J *et al.* 2012. BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**: 1919–1920.

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

Nordborg M, Tavaré S 2002. Linkage disequilibrium: what history has to tell us. *Trends in Genetics* **18**: 83–90.

Osipova SV, Permyakov AV, Permyakova MD, Pshenichnikova TA, Börner A 2011. Leaf dehydroascorbate reductase and catalase activity is associated with soil drought tolerance in bread wheat. *Acta Physiologiae Plantarum* **33**: 2169–2177.

Pinhasi R, Fort J, Ammerman AJ 2005. Tracing the origin and spread of agriculture in Europe. *PLoS Biol* **3**: e410.

Poets AM, Fang Z, Clegg MT, Morrell PL 2015a. Barley landraces are characterized by geographically heterogeneous genomic origins. *Genome Biology* **16**: 173.

Poets AM, Mohammadi M, Seth K, Wang H, Kono TJY, Fang Z, Muehlbauer GJ, Smith KP, Morrell PL 2015b. The effects of both recent and long-term selection and genetic drift are readily evident in North American barley breeding populations. *G3: Genes, Genomes, Genetics* g3. 115.024349.

Poets AM, Fang Z, Clegg MT, Morrell PL 2015c. Barley landraces are characterized by geographically heterogeneous genomic origins. *Genome Biology* **16**: 173.

Pyhäjärvi T, Hufford MB, Mezmouk S, Ross-Ibarra J 2013. Complex patterns of local adaptation in teosinte. *Genome Biology and Evolution* **5**: 1594–1609.

Reinheimer JL, Barr AR, Eglinton JK 2004. QTL mapping of chromosomal regions conferring reproductive frost tolerance in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* **109**: 1267–1274.

Russell J, Mascher M, Dawson IK, Kyriakidis S, Calixto C, Freund F, Bayer M, Milne I, Marshall-Griffiths T *et al.* 2016. Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. *Nature Genetics* **48**: 1024.

Saisho D, Purugganan M 2007. Molecular phylogeography of domesticated barley traces expansion of agriculture in the Old World. *Genetics*

Sasaki K, Christov NK, Tsuda S, Imai R 2013. Identification of a novel LEA protein involved in freezing tolerance in wheat. *Plant and Cell Physiology* **55**: 136–147.

Shaar-Moshe L, Hübner S, Peleg Z 2015. Identification of conserved drought-adaptive genes using a cross-species meta-analysis approach. *BMC Plant Biology* **15**: 111.

Shin J-H, Blay S, McNeney B, Graham J 2006. LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *Journal of Statistical Software* **16**: 1–10.

Skinner JS, Szűcs P, von Zitzewitz J, Marquez-Cedillo L, Filichkin T, Stockinger EJ, Thomashow MF, Chen THH, Hayes PM 2006. Mapping of barley homologs to genes that regulate low temperature tolerance in *Arabidopsis*. *Theoretical and Applied Genetics* **112**: 832–842.

Stephens M, Scheet P 2005. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *The American Journal of Human Genetics* **76**: 449–462.

Stephens M, Smith NJ, Donnelly P 2001. A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics* **68**: 978–989.

Stockinger EJ, Skinner JS, Gardner KG, Francia E, Pecchioni N 2007. Expression levels of barley *Cbf* genes at the *Frost resistance-H2* locus are dependent upon alleles at *Fr-H1* and *Fr-H2*. *The Plant Journal* **51**: 308–321.

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

Sweigart AL, Willis JH 2003. Patterns of nucleotide diversity in two species of *Mimulus* are affected by mating system and asymmetric introgression. *Evolution* **57**: 2490–2506.

Team RC 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016.

Teshima KM, Coop G, Przeworski M 2006. How reliable are empirical genomic scans for selective sweeps. *Genome Research* **16**: 702–712.

Tiffin P, Ross-Ibarra J 2014. Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution* **29**: 673–680.

Tondelli A, Francia E, Barabaschi D, Aprile A, Skinner JS, Stockinger EJ, Stanca AM, Pecchioni N 2006. Mapping regulatory genes as candidates for cold and drought stress tolerance in barley. *Theoretical and Applied Genetics* **112**: 445–454.

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

Visioni A, Tondelli A, Francia E, Pswarayi A, Malosetti M, Russell J, Thomas W, Waugh R, Pecchioni N *et al.* 2013. Genome-wide association mapping of frost tolerance in barley (*Hordeum vulgare* L.). *BMC Genomics* **14**: 424.

Wang K, Li M, Hakonarson H 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* **38**: e164–e164.

Wehner G, Balko C, Humbeck K, Zyprian E, Ordon F 2016. Expression profiling of genes involved in drought stress and leaf senescence in juvenile barley. *BMC Plant Biology* **16**: 3.
Weir BS, Cockerham CC 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
Willcox G 2002. Geographical variation in major cereal components and evidence for independent domestication events in Western Asia. *The dawn of farming in the Near East* **6**: 133–140.
Wright S 1949. The genetical structure of populations. *Annals of Eugenics* **15**: 323–354.
Xia Y, Li R, Bai G, Siddique KHM, Varshney RK, Baum M, Yan G, Guo P 2017. Genetic variations of *HvP5CS1* and their association with drought tolerance related traits in barley (*Hordeum vulgare* L.). *Scientific reports* **7**: 7870.
Xu D, Zhu D, Deng XW 2016. The role of *COP1* in repression of photoperiodic flowering. *F1000Research* **5**:
Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S *et al.* 2006. The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. *Proceedings of the National Academy of Sciences* **103**: 19581–19586.
Yoder JB, Stanton-Geddes J, Zhou P, Briskine R, Young ND, Tiffin P 2014. Genomic signature of adaptation to climate in *Medicago truncatula*. *Genetics* **196**: 1263–1275.
Zakhrabekova S, Gough SP, Braumann I, Müller AH, Lundqvist J, Ahmann K, Dockter C, Matyszczak I, Kurowska M *et al.* 2012. Induced mutations in circadian clock regulator *Mat-a* facilitated short-season adaptation and range extension in cultivated barley. *Proceedings of the National Academy of Sciences* **109**: 4326–4331.
Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK *et al.* 2010. Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* **42**: 355.
Zohary D, Hopf M, Weiss E 2012. Domestication of plants in the Old World.

Figures and Tables:

Figures:

Figure 1: F_{ST} for 9K SNPs in samples from comparisons of high latitude (a), high latitude (b), elevation (c), and growth habit (d).

Figure 2: The geographic distribution of the SNPs with high F_{ST}. (a & c) The geographic distribution of allelic types of 9K SNP 12_30191 with highest F_{ST} = 0.4839. The F_{ST} was from the high latitude (HL) comparison. (b & d) The geographic distribution of allelic types of 9K SNP 12_11529 with highest F_{ST} value of 0.6493. The F_{ST} was from elevation (E) comparison. The

color bar indicates the elevations in meters. The filled pink circles indicate the derived allele, while the blue open circles indicated the ancestral allele.

Figure 3: (a) The genomic distribution of outlier SNPs identified according to the F_{ST} comparisons of elevation (below 3,000 m vs above 3,000 m), low latitude (below 30˚N vs 30-40˚N), high latitude (30-40˚N vs above 40˚N), growth habit (winter vs spring) and association analysis of 21 bioclimatic variables, which are categorized into three classes (precipitation, temperature, and geographic variables) on chromosome 3H; (B) exome capture target density (dark blue line), cross over rate in cM/2 Mb (purple line), the genomic distribution of SNPs identified in the 62 barley landraces (vertical light blue lines), and 9K SNPs (red triangle) on chromosome 3H; (C) LD plots for SNPs significant associated with at least one bioclimatic variables (bottom) on chromosome 3H. Each plot shows +/- 100 Kb of the SNP. The vertical dotted lines in the upper panel indicate that those outlier SNPs are the same in different traits. For the LD plots, genotyped SNPs are at location 0, and positions upstream and downstream are listed as negative and positive values. The light blue bars are genes in the 200 Kb windows surround the genotyped SNPs. The SNPs from the left to right are: 11_20742, SCRI_RS_229693, 11_20162, 11_10380, SCRI_RS_178900, 12_31393, SCRI_RS_6793, 11_11021, SCRI_RS_201224, SCRI_RS_204462, and SCRI_RS_144559.

Figure 4: (A) The Linkage disequilibrium (LD) analysis of genotyped SNP SCRI_RS_137464 significantly associated with mean temperature of wettest quarter (BIO8). The blue bars indicate genes in the 200 Kb window surrounding SCRI_RS_137464, the red arrow indicates the *HvPRR1/HvTOC1* (flowering time-related gene) that includes SNP SCRI_RS_137464 (B) The gene structure of *HvPRR1/HvTOC1* and the functional annotation of SNPs in this gene. (C) Haplotype structures of *HvPRR1/HvTOC1* based on the SNPs in this gene. L: low; H: high.

Tables:

Table1: Loci identified in environmental association or F_{ST} comparisons that are previously reported to contribute to flowering time, cold and drought tolerance. Gene names are preceded by two letter prefix with the genus and specific epithet for the species where the gene was identified. This include, *At - Arabidopsis thaliana, Ta - Triticum aestivum, Os - Oryza sativa, Br - Brassica napus,* and *Sod - Saccharum officinarum.* The F_{ST} comparisons involve the following

comparisons: E: elevation; LL: Low Latitude; GH: growth habit; HL: high latitude. The * indicates SNPs included at the 97.5% threshold but the 99% threshold.

	Barley	Other plants	Bioclimatic variables	F_{ST}
Cold		*TaWCI 16*	6,11,IC1	
		OsiSAP8		E
	HvICE2			E
	HvCBF4B			LL
	**HvSS1*			E
	HvDhn7		Longitude	
Drought	*HvDhn7*		Longitude	
		TaEXPA2	IC1	
		**OsNAC52*		LL
		**SodERF3*		HL
		**OsCDPK7*		E
		**AtACBP2*		LL
		**AtERECTA*		HL
		**BrERF4*		GH
		**OsNAC5*		LL
Flowering time	*HvVrn-H1/HvAP1*		9	
	HvPhyC		7, 14,17, & IC2	E
	HvPRR1/HvTOC1		8	
	**HvPpd-H2/HvFT3*			GH
	**HvELF3/Esp1L/eam8*			GH
	HvPpd-H1/HvPRR37			LL, HL
		AtCOP1		GH
		**AtFVE*		HL

Note:
1 = Annual Mean Temperature; 2 = Mean Diurnal Range (Mean of monthly (max temp - min temp)); 3 = Isothermality (2/7) (* 100); 4 = Temperature Seasonality (standard deviation *100); 5 = Max Temperature

of Warmest Month; 6 = Min Temperature of Coldest Month; 7 = Temperature Annual Range (5-6); 8 = Mean Temperature of Wettest Quarter; 9 = Mean Temperature of Driest Quarter; 10 = Mean Temperature of Warmest Quarter; 11 = Mean Temperature of Coldest Quarter; 12 = Annual Precipitation; 13 = Precipitation of Wettest Month; 14 = Precipitation of Driest Month; 15 = Precipitation Seasonality (Coefficient of Variation); 16 = Precipitation of Wettest Quarter; 17 = Precipitation of Driest Quarter; 18 = Precipitation of Warmest Quarter; 19 = Precipitation of Coldest Quarter; IC = Independent Component.

Table 2: The number of SNPs identified by F_{ST} outlier approaches and number of previously reported genes they identify. HE: High elevation; LL: Low Latitude; GH: growth habit; HL: high latitude

Outliers	F_{ST}	Flowering time		Cold tolerance		Drought tolerance	
		Outliers	Genes	Outliers	Genes	Outliers	Genes
62	HL	6	1	0	0	0	0
62	LL	6	1	1	1	0	0
62	E	1	1	2	2	0	0
62	GH	1	1	0	0	0	0

Table 3: The number of SNPs significantly associated with climatic factors and known genes they hit.

Categories	Categories	Number	Climatic factors
Flowering time	**SNPs**	3	7,8,9,14, and 17
	Genes	3	
Cold tolerant	**SNPs**	2	6,11, and IC1
	Genes	2	
Drought tolerant	**SNPs**	2	IC1 and longitude
	Genes	2	

Note:
1 = Annual Mean Temperature; 2 = Mean Diurnal Range (Mean of monthly (max temp - min temp)); 3 = Isothermality (2/7) (* 100); 4 = Temperature Seasonality (standard deviation *100); 5 = Max Temperature of Warmest Month; 6 = Min Temperature of Coldest Month; 7 = Temperature Annual Range (5-6); 8 = Mean Temperature of Wettest Quarter; 9 = Mean Temperature of Driest Quarter; 10 = Mean Temperature of Warmest Quarter; 11 = Mean Temperature of Coldest Quarter; 12 = Annual Precipitation; 13 =

Precipitation of Wettest Month; 14 = Precipitation of Driest Month; 15 = Precipitation Seasonality (Coefficient of Variation); 16 = Precipitation of Wettest Quarter; 17 = Precipitation of Driest Quarter; 18 = Precipitation of Warmest Quarter; 19 = Precipitation of Coldest Quarter; IC = Independent Component.

Table 4: Linkage disequilibrium for all SNPs associated with environmental variables or identified as F_{ST} outliers.

	LD within a gene	Extended LD	No LD	Missing
F_{ST}	56	159	23	3
Association	46	98	8	7

Supplemental figures:

Methods S1

Fig. S1: Exome capture target density (dark blue line), recombination rate in cM/2 Mb (purple line), the genomic distribution of SNPs identified in the 62 barley landraces (vertical light blue lines), and 9K iSelect SNPs (red triangle) for seven chromosomes. The crossing over rate were calculated using 9K SNPs. The genetic position of each SNP were inferred from Muñoz-Amatriaín et al. (2011).

Fig. S2: (a) The derived site frequency spectrumfor 2,806 SNPs inferred ancestral state from the 9K iSelect array genotyped in the 784 landraces and (b) 340,260 SNPs with exome-capture resequencing data in 62 landraces. Ancestral state was based on majority state from *H. murinum spp. glaucum* resequencing mapped to the Morex assembly. (c) Minor allele frequency, for 6152 SNPs with 9K iSelect array in the 784 landraces and (d) for 482,714 SNPs with exome-capture resequencing data in 62 landraces.

Fig. S3: The genomic distribution of outlier SNPs identified according to the F_{ST} comparisons of elevation (below 3,000 m vs above 3,000 m), low latitude (below 30˚N vs 30-40˚N), high latitude (30-40˚N vs above 40˚N), growth habit (winter vs spring) and association analysis of 21 bioclimatic variables, which are categorized into three classes (precipitation, temperature, and geographic variables) on all chromosomes;

Fig. S4: The geographic distribution of barley landraces used in this study. Blue triangles indicate samples with elevation >3,000 m. The gray dot indicate samples with elevation ≤3,000 m.

Fig. S5: Venn diagram for the F_{ST} outliers from the comparisons of elevation, high and low latitude, and growth habit.

Fig. S6: The geographic distribution of the SNPs with high F_{ST}. (a) The geographic distribution of allelic types of 9K SNP SCRI_RS_170209 with highest F_{ST} = 0.5171. The F_{ST} was from the low latitude (LL) comparison. (b) The geographic distribution of allelic types of 9K SCRI_RS_148782
with highest F_{ST} = 0.3948. The F_{ST} was from growth habit (GH) comparison. The color bar indicates the elevations in meters. The filled pink circles indicate the derived allele, while the blue open circles indicated the ancestral allele.

Fig. S7: Venn diagram for the candidate SNPs significant associated with three categories of environmental variables: precipitation, temperature, and geographic factors.

Fig. S8: LD decay plot for 200 Kb window around the significant SNPs associated with environmental variables. The blue bars underneath the x-axis are the annotated genes in the 200 Kb windows. The vertical dash lines are the candidate SNPs' locations, and minors and plus means upstream and downstream of the candidate SNPs' locations.

Fig. S9: (a) The Linkage disequilibrium (LD) analysis of candidate SNP SCRI_RS_137464 significant associated with "min temperature of coldest month (BIO6)" and "mean temperature of the coldest quarter" (BIO6 and 11). The blue bars indicate genes in the 200 Kb window surrounding 11_10380, the red arrow indicates the WCI 16 (cold tolerant-related gene) hit by 11_10380, (b) The gene structure of WCI 16 and the functional annotation of SNPs in this gene. (c) Haplotype structures of WCI 16 based on the SNPs in this gene. L: low; H: high.

Fig. S10: (a) The Linkage disequilibrium (LD) analysis of candidate SNP SCRI_RS_235243 significant associated with "precipitation of driest months" (BIO14). The blue bars indicate genes in the 200 Kb window surrounding SCRI_RS_235243, the red arrow indicates the DHAR (drought tolerant-related gene) hit by SCRI_RS_235243, (b) The gene structure of DHAR and

the functional annotation of SNPs in this gene. (c) Haplotype structures of DHAR based on the SNPs in this gene. L: low; H: high.

Supplemental tables:

Table S1: 784 barley landraces after removing 19 accessions from the 803 Poets et al. 2015 panel.

Table S2: The detail information of the exome capture data from 62 landraces.

Table S3: Known flower time-related genes list.

Table S4: Known cold tolerance-related genes list.

Table S5: Known drought tolerance-related genes list.

Table S6: 159 unique candidate SNPs significantly associated with at least one environmental variable.

Table S7: F_{ST} outliers from elevation, high and low latitude, and growth habit comparisons.

Table S8: The overlapped SNPs identified by F_{ST} outliers and association analysis approaches.

Table S9: Previous reported genes identified by taking bottom 1.5th percentile threshold of p-value from environmental associations.

Table S10: F_{ST} outliers in putative inverted region reported by Fang et al. 2014

Supplemental datasets:

Supplemental data 1: Phenotype matrix with 25 geographic and climatic variables for environmental association.

Supplemental data 2: Genotype matrix with 5,800 SNPs for environmental association

Supplemental data 3: Inferred ancestral status for each 9K SNP.

Supplemental data 4: Inferred ancestral status for each exam resequencing SNP.

Supplemental data 5: All p-values and FST from elevation, low and high latitude, longitude, and growth habit.

Supplemental data 6: All P-values and Benjamini-Hochberg FDR-values from the environmental associations for 25 variables.

Supplemental data 7: The annotations for SNPs called from 62 landraces exome data

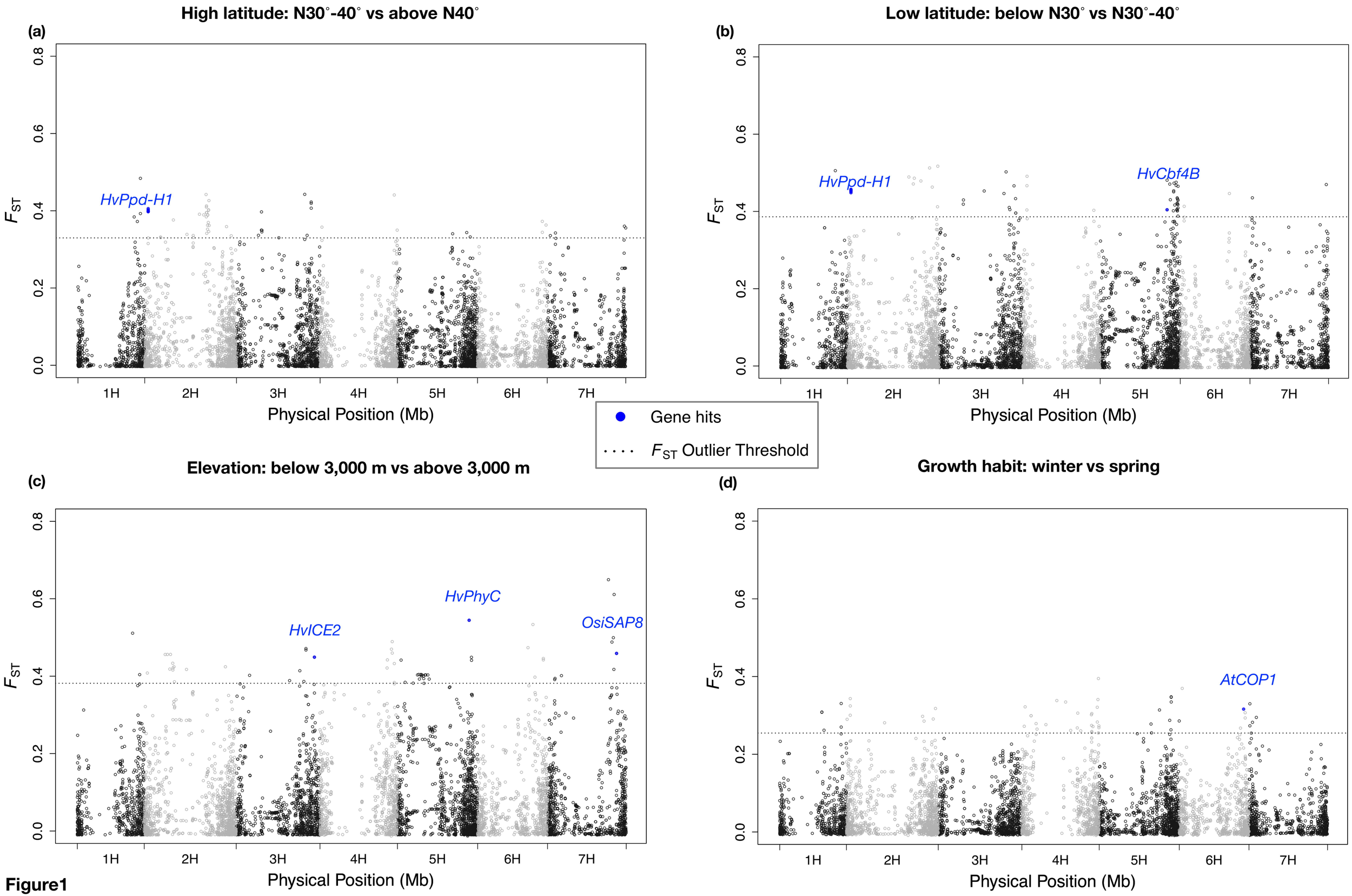

High latitude: N30°-40° vs above N40°
(a)
HvPpd-H1
Low latitude: below N30° vs N30°-40°
(b)
HvPpd-H1
HvCbf4B
Gene hits
F_{ST} Outlier Threshold
Elevation: below 3,000 m vs above 3,000 m
(c)
HvICE2
HvPhyC
OsiSAP8
Growth habit: winter vs spring
(d)
AtCOP1
F_{ST}
0.0
0.2
0.4
0.6
0.8
1H
2H
3H
4H
5H
6H
7H
Physical Position (Mb)

Figure1

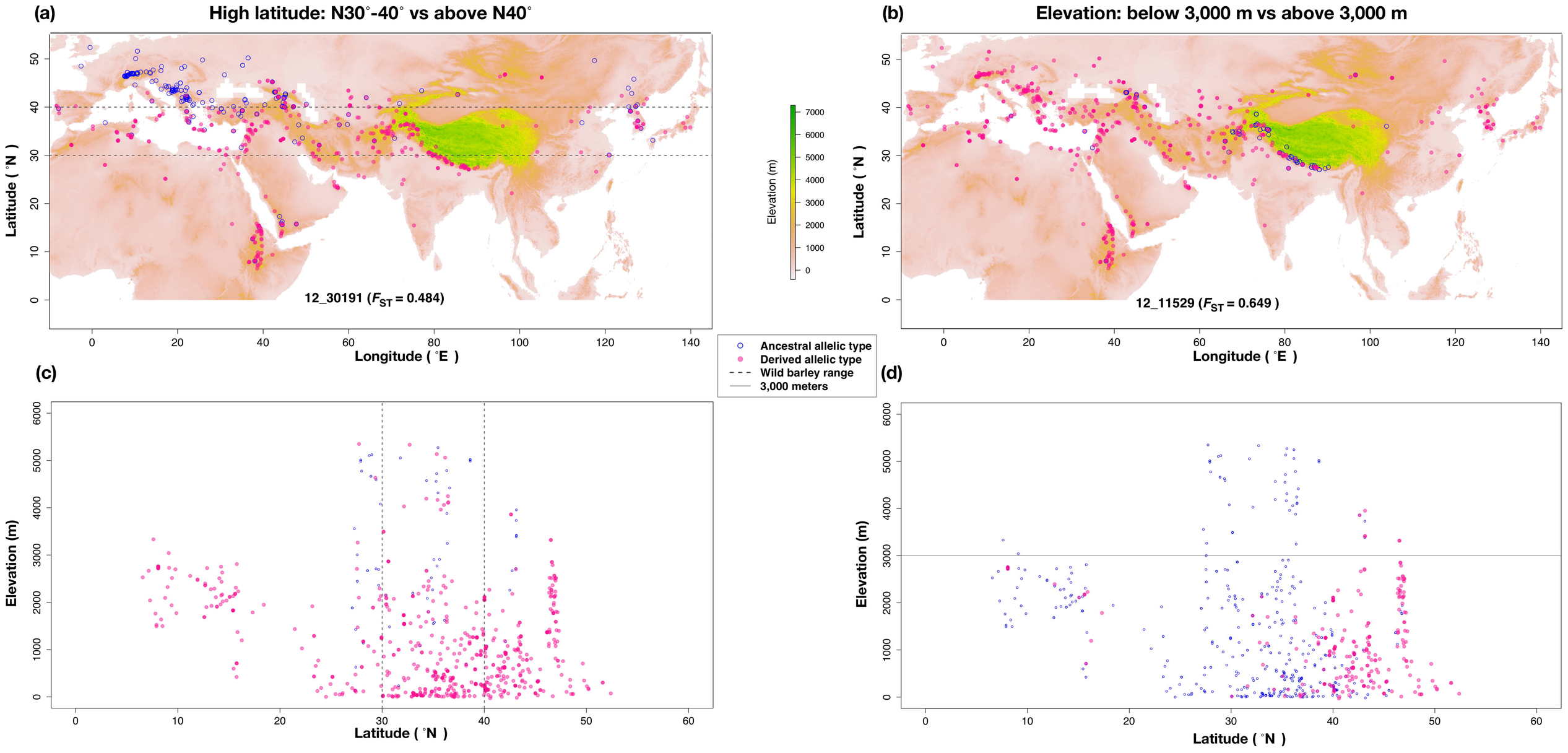

(a)
High latitude: N30˚-40˚ vs above N40˚
12_30191 (F_{ST} = 0.484)
Latitude (˚N)
Longitude (˚E)
Elevation (m)
(b)
Elevation: below 3,000 m vs above 3,000 m
12_11529 (F_{ST} = 0.649)
Ancestral allelic type
Derived allelic type
Wild barley range
3,000 meters
(c)
Elevation (m)
Latitude (˚N)
(d)
Elevation (m)
Latitude (˚N)

Figure 2

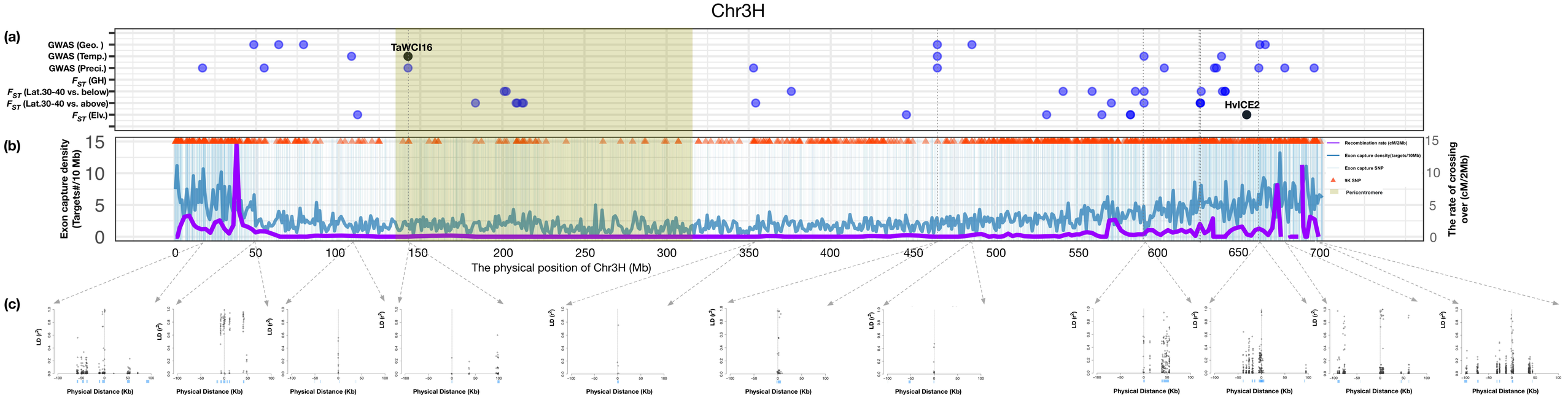
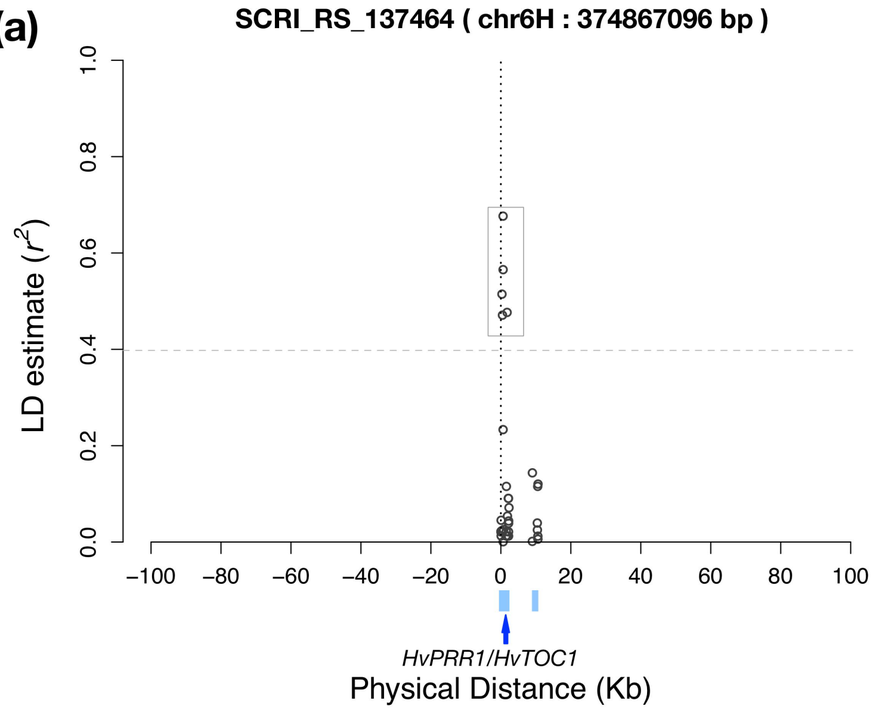
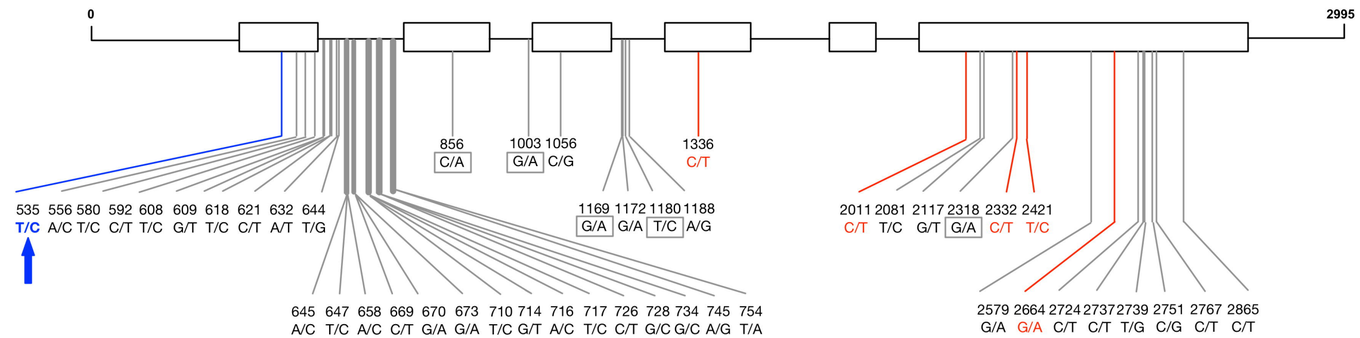

Chr3H
(a)
GWAS (Geo.)
GWAS (Temp.)
GWAS (Preci.)
F_{ST} (GH)
F_{ST} (Lat.30-40 vs. below)
F_{ST} (Lat.30-40 vs. above)
F_{ST} (Elv.)
TaWCI16
HvICE2
(b)
Exon capture density (Targets#/10 Mb)
The rate of crossing over (cM/2Mb)
Recombination rate (cM/2Mb)
Exon capture density(targets/10Mb)
Exon capture SNP
9K SNP
Pericentromere
The physical position of Chr3H (Mb)
(c)
LD (r^2)
Physical Distance (Kb)

Figure 3

(a)



(b) ***HvPRR1/HvTOC1***



Red: Nonsynonymous SNP

Blue: Significant associated SNP: SCRI_RS_137464

▭ : LD with significant associated SNP (r^2>0.4)

(c)

Sites	535	556	580	592	608	609	618	621	632	644	645	647	658	669	670	673	710	714	716	717	726	728	734	745	754	856	1003	1056	1169	1172	1180	1188	1336	2011	2081	2117	2318	2332	2421	2579	2664	2724	2737	2739	2751	2767	2865	General				Subsets			
Physical position	7096	7117	7141	7153	7169	7170	7179	7182	7193	7205	7206	7208	7219	7230	7231	7234	7271	7275	7277	7278	7287	7289	7295	7306	7315	7417	7564	7617	7730	7733	7741	7749	7897	8572	8642	8678	8879	8893	8982	9140	9225	9285	9298	9300	9312	9328	9426	Elevation		Latitude		L_Latitude Elevation		H_Latitude Elevation	
Consensus	T	A	T	C	T	G	T	C	A	T	A	T	A	C	G	G	T	G	A	T	C	G	G	A	T	C	G	C	G	A	T	A	C	C	T	G	G	C	C	G	G	C	C	T	C	C	C								
Ancestral	C	.	C	.	.	T	.	T	-	.	.	.	.	T	-	A	.	.	C	.	T	C	C	.	.	.	.	-	A	G	C	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	-	L	H	L	H	L	H	L	H
Hap1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	5	-	3	2	3	-	2	-
Hap2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	-	1	-	1	-	-	-
Hap3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	15	1	8	8	7	1	8	-
Hap4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	3	-	1	2	1	-	2	-
Hap5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	1	-	-	1	-	-	1	-
Hap6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	3	-	1	2	1	-	2	-
Hap7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	A	.	.	.	.	.	.	.	.	3	-	1	2	-	-	1	-
Hap8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	2	-	1	1	1	-	1	-
Hap9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	G	.	.	.	2	-	1	1	1	-	1	-
Hap10	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	-	1	-	1	-	-	-
Hap11	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	T	.	.	.	.	.	.	.	.	.	.	T	.	2	-	2	-	2	-	-	-
Hap12	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	2	-	2	-	2	-	-	-
Hap13	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	.	1	-	-	1	-	-	1	-
Hap14	C	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	1	-	1	-	1	-	-	-
Hap15	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	-	-	1	-	-	1	-
Hap16	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	.	A	G	C	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	9	2	8	3	6	2	3	-
Hap17	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	.	A	G	C	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	4	-	2	2	2	-	2	-
Hap18	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	.	A	G	C	.	.	.	.	.	A	.	.	.	A	.	.	.	.	.	.	.	4	-	4	-	4	-	-	-
Hap19	C	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	1	-	1	-	1	-	-	-
Hap20	C	C	C	T	C	T	.	T	T	G	C	C	C	T	A	A	C	T	C	C	T	C	C	G	.	.	.	.	A	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	1	-	1	-	1	-	-	-

Figure 4