File S1

**Physical Mapping of 9K SNPs**

A number of genetic mapping studies have reported genetic map positions for barley SNPs (Mayer *et al.*, 2011; Comadran *et al.*, 2012; Muñoz-Amatriaín *et al.*, 2014) , and the physical locations for a portion of iSelect SNPs relative to the barley reference genome (Mascher *et al.*, 2017) have been reported previously (Mayer *et al.*, 2011; Comadran *et al.*, 2012; Cantalapiedra *et al.*, 2015; Colmsee *et al.*, 2015; Bayer *et al.*, 2017). We used the contextual sequences of 7,864 SNPs from the 9K Illumina Infinium iSelect Custom Genotyping BeadChip (typically either 121 or 241 bp long) (Comadran *et al.*, 2012) to perform BLASTn (Altschul *et al.*, 1990)searches against the masked reference genome (Mascher *et al.*, 2017). We configured BLASTn to reject hits with an expect value greater than 0.000001 and identity less than 90%. For sequences where BLAST did not identify a unique position, we used previously reported genetic map positions (Muñoz-Amatriaín *et al.*, 2014) to infer the most likely chromosome of origin. If two BLAST hits were inferred to be within 100 kb from each other, we systematically chose the one with smaller value along the pseudomolecule as the physical position. Otherwise, we identified the physical position with the hit closest to the genetic position. The mapping of 9K SNPs was performed using the Python program SNP\_Utils (https://github.com/mojaveazure/SNP\_Utils). We identified 425 SNPs that were not aligned due to either no hits above the e-value threshold ≥0.000001 or identity ≤ 90% or multiple hits within ≤ 100 Kb for SNPs with no estimated genetic positions. For the 425 SNPs, we used a manual BLAST search of contextual sequence using the IPK web server (http://webblast.ipk-gatersleben.de/barley\_ibsc/viroblast.php) with default parameters. The BLAST search used no threshold and involved selecting the hits with the highest combined rank of identity and score. If the contextual sequence did not have a unique BLAST hit in the genome, we used SNPMeta (Kono *et al.*, 2014) to identify the potential gene where the SNP resides (see [link to Barley\_SNP\_Annotations](http://conservancy.umn.edu/bitstream/handle/11299/181367/Barley_SNP_Annotations) in supplemental text for SNPmeta results), then a BLAST search of the gene against the masked reference genome to identify the physical location of the best hit.

**Exome capture sequence data handling and SNP calling**

Sequence quality assessment used FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) and seqqs (https://github.com/vsbuffalo/seqqs). Reads were counted using bioawk scripts (https://github.com/lh3/bioawk). Adapter trimming used Scythe (https://github.com/vsbuffalo/scythe). Reference-based read mapping against the draft barley reference genome (Mascher *et al.*, 2017) was conducted with the Burrows-Wheeler Aligner (BWA-MEM) (Li and Durbin, 2009). Read mapping used default parameters for BWA-MEM except for the following: 8 threads, minimum seed length of 10, re-seed value of 1.0, a gap penalty of 8, a mismatch penalty of 4, and a minimum reporting threshold of 85. Read mapping parameters were chosen to permit a ~2% mismatch between reads and the reference sequence, the highest estimated nucleotide diversity reported based on Sanger resequencing in wild and cultivated barley(Morrell *et al.*, 2003; Morrell *et al.*, 2006; Morrell *et al.*, 2014) Samtools (Li *et al.*, 2009) and Picard (<http://broadinstitute.github.io/picard/>), were used for alignment sorting, de-duplicating, and adding read groups to the resulting BAM files. Estimates of read depth and coverage made use of ‘bedtools genomecov’ (Quinlan and Hall, 2010) relative to an empirical estimate of exome coverage (Kono *et al.*, 2018). Briefly, this estimate was based on mapping of roughly 241X exome capture reads from Morex, reported by (Mascher *et al.*, 2013), back to the Morex draft genome (Mascher *et al.*, 2017). Regions with > 50X were considered covered by exome capture (https://github.com/MorrellLAB/captured\_50x\_BED). This results in ~80 Mb of exome coverage (Kono *et al.*, 2018) relative to the 60 Mb capture design by (Mascher *et al.*, 2013).

Alignment processing followed the Genome Analysis Toolkit (GATK) best practices workflow (McKenna *et al.*, 2010; DePristo *et al.*, 2011).Cleaned BAM alignments were realigned around putative insertion/deletion (indel) sites. Individual sample genotype likelihoods were then calculated with the HaplotypeCaller, with a haploid model and a “heterozygosity” (pairwise diversity) value of 0.008 per base pair. This value is the mean estimate of coding nucleotide sequence diversity, based on previous Sanger resequencing experiments (Caldwell *et al.*, 2006; Morrell *et al.*, 2014). SNP calls were made from the genotype likelihoods with the GATK tool GenotypeGVCFs (McKenna *et al.*, 2010). A high-confidence subset of these SNP calls was created by both site-based filtering (position, number of alleles, QUAL, missingness, heterozygosity) and then genotype-based filtering (genotype quality, depth). The site-based filtering was to exclude indels, sites with more than two alleles, sites with a QUAL score < 40 or missing, sites outside the exome capture regions, sites with > 90% heterozygous calls, and sites with ≥ 20% of the calls either missing. The genotype-based filtering was to filter out SNPs with a genotype quality < 9, or with a depth < 5. We used 7,864 9K SNPs with physical positions, 708,620 SNPs from (Kono *et al.*, 2016), and 2,124,487 SNPs from wild barley as prior variants and ran GATK `Variant Recalibrator` and Apply `Recalibration` to recalibrate the variants called by GATK to produce the raw VCF file.

Sites outside the empirically inferred exome capture regions, sites with more than two alleles, and indels were filtered from the raw VCF file. Genotype calls were considered missing if the individual read depths were <5 and >109 (the 95th percentile of coverage), or genotype quality was < 9, or the ratio of reads supported for reference to alternative alleles showed a > 10% deviation from 50:50. SNP positions with > 90% heterozygous calls, a QUAL score < 40, or > 20% missing genotype calls were discarded.

Scripts to perform adapter contamination removal, read mapping, alignment cleaning, implementing the GATK best practices workflow, GATK VariantRecalibrator, ApplyRecalibration, and filtering for VCF files are part of the `sequence\_handling` workflow developed by Hoffman *et al.* (2018).

**Environmental association mapping**

The latitude, longitude, elevation, and BIO1 to BIO19 values of the collection locations for 784 landraces are given in the phenotype data file (Supplemental data 5). To determine if consolidation of components identifies novel variants, the top three Independent Components (IC) were calculated from BIO1 to BIO19 values after standardization of each BIO variable using the icaimax ‘ica’ function from the ica-package in R (Bell and Sejnowski, 1995). We summarized BIO variables with ICs rather than principal component analysis (PCA) because ICs infer the orthogonal coordinate system with a given number of dimensions that capture the data best (maximize the non-Gaussian-ness). PCA is designed to rotate the original orthogonal coordinate system around the origin to capture the highest variance of the data in the order of principal components, an approach that is highly susceptible to redundancy in the data. To compare the summaries of BIO variables, we tested for correlation between the ICs/PCs and individual Bioclim variables. We found that using the top three ICs appears to capture the cold temperature trend better than using top three PCs (Table S3). However, the ICs constitute a somewhat extreme summary of the bioclimatic variables, as the first three ICs included only eight bioclimatic variables (Table S3). Those eight variables are not closely related to the remainder of the BIO variables (Figure S6) and potentially result in a loss of information with regard to associations with remaining variables. Therefore, we report association analysis to both ICs and individual Bioclim variables (Supplemental data 5).

The GAPIT and Efficient Mixed Model Association (EMMA) packages (Zhang *et al.*, 2010; Lipka *et al.*, 2012)were used together with the R packages, MASS (Venables and Ripley, 2002), multtest (Pollard *et al.*, 2005), gplots (Warnes *et al.*, 2016), compiler (Tierney, 2019), and scatterplot3d (Ligges and Maechler, 2003), according to the GAPIT demo script (<http://www.zzlab.net/GAPIT/gapit_tutorial_script.txt>). For GAPIT, default parameter values were used except that the PCA total argument was set to three. We exclude SNPs with minor allele frequency (MAF) > 0.01 from association analysis. To correct for multiple testing, we applied the Benjamini-Hochberg false discovery rate (FDR) correction. We report both adjusted *p*-values and FDR with an FDR threshold ≤ 0.25.

**Permutation and Empirical *P*-value for *F*ST**

The null distribution of *F*ST values was estimated from 1,000 permutations of genotype data for each comparison. For example, 80 randomly selected individuals were assigned to either the spring or winter partition for growth habit. Because there are only 54 samples from the high elevation > 3000 m partition, a random sample of 25 individuals was assigned to each partition for the calculation of the null *F*ST distribution. The R package Hierfstat (Goudet, 2005) was used to calculate *F*ST at each SNP. The *p*-value for each SNP was calculated as the percentage of null *F*ST values (out of 1,000) that exceeded the observed empirical *F*ST value.

**Reference:**

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 1990 Basic local alignment search tool. Journal of Molecular Biology 215: 403-410.

Bayer, M. M., P. Rapazote-Flores, M. Ganal, P. E. Hedley, M. Macaulay *et al.*, 2017 Development and evaluation of a barley 50k iSelect SNP array. Frontiers in plant science 8: 1792.

Bell, A. J., and T. J. Sejnowski, 1995 An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7: 1129-1159.

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

Cantalapiedra, C. P., R. Boudiar, A. M. Casas, E. Igartua, and B. Contreras-Moreira, 2015 BARLEYMAP: physical and genetic mapping of nucleotide sequences and annotation of surrounding loci in barley. Molecular Breeding 35: 13.

Colmsee, C., S. Beier, A. Himmelbach, T. Schmutzer, N. Stein *et al.*, 2015 BARLEX – the barley draft genome explorer. Molecular Plant 8: 964-966.

Comadran, J., B. Kilian, J. Russell, L. Ramsay, N. Stein *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.

Depristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire *et al.*, 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43: 491-498.

Goudet, J., 2005 hierfstat, a package for R to compute and test hierarchical F-statistics. Molecular Ecology Notes 5: 184-186.

Hoffman, P. J., S. R. Wyant, T. J. Y. Kono, and P. L. Morrell, 2018 MorrellLAB/sequence\_handling: Release v2.0: SNP calling with GATK 3.8.

Kono, T. J. Y., C. Liu, E. E. Vonderharr, D. Koenig, J. C. Fay *et al.*, 2018 The fate of deleterious variants in a barley genomic prediction population. bioRxiv :10.1101/442020 (Preprint posted October, 2018).

Kono, T. J. Y., F. Fu, M. Mohammadi, P. J. Hoffman, C. Liu *et al.*, 2016 The Role of Deleterious Substitutions in Crop Genomes. Molecular Biology and Evolution 33: 2307-2317.

Kono, T. J. Y., K. Seth, J. A. Poland, and P. L. Morrell, 2014 SNPMeta: SNP annotation and SNP metadata collection without a reference genome. Molecular Ecology Resources 14: 419-425.

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

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.

Ligges, U., and M. Maechler, 2003 scatterplot3d - an R package for visualizing multivariate data. Journal of Statistical Software 8: 1-20.

Lipka, A. E., F. Tian, Q. Wang, J. Peiffer, M. Li *et al.*, 2012 GAPIT: genome association and prediction integrated tool. Bioinformatics 28: 2397-2399.

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

Mascher, M., T. A. Richmond, D. J. Gerhardt, A. Himmelbach, L. Clissold *et al.*, 2013 Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond. The Plant Journal 76: 494-505.

Mayer, K. F. X., M. Martis, P. E. Hedley, H. Šimková, H. Liu *et al.*, 2011 Unlocking the barley genome by chromosomal and comparative genomics. The Plant Cell 23: 1249-1263.

Mckenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20: 1297-1303.

Morrell, P. L., A. M. Gonzales, K. K. T. Meyer, and M. T. Clegg, 2014 Resequencing data indicate a modest effect of domestication on diversity in barley: a cultigen with multiple origins. Journal of Heredity 105: 253-264.

Morrell, P. L., K. E. Lundy, and M. T. Clegg, 2003 Distinct geographic patterns of genetic diversity are maintained in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite migration. Proceedings of the National Academy of Sciences of the United States of America 100: 10812-10817.

Morrell, P. L., D. M. Toleno, K. E. Lundy, and M. T. Clegg, 2006 Estimating the contribution of mutation, recombination and gene conversion in the generation of haplotypic diversity. Genetics 173: 1705-1723.

Muñoz-Amatriaín, M., A. Cuesta-Marcos, J. B. Endelman, J. Comadran, J. M. Bonman *et al.*, 2014 The USDA barley core collection: genetic diversity, population dtructure, and potential for genome-wide association studies. PLOS ONE 9: e94688.

Pollard, K. S., S. Dudoit, and M. J. Van Der Laan, 2005 Multiple Testing Procedures: the multtest Package and Applications to Genomics, pp. 249-271 in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, edited by R. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry, and S. Dudoit. Springer New York, New York, NY.

Quinlan, A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-842.

Tierney, L., 2019 A Byte Code Compiler for R. 123.

Venables, W. N., and B. D. Ripley, 2002 *Modern Applied Statistics with S*. Springer-Verlag, New York.

Warnes, G. R., B. Bolker, L. Bonebakker, R. Gentleman, W. H. A. Liaw *et al.*, 2016 gplots: Various R Programming Tools for Plotting Data version 3.0.1.1 from CRAN.

Zhang, Z., E. Ersoz, C.-Q. Lai, R. J. Todhunter, H. K. Tiwari *et al.*, 2010 Mixed linear model approach adapted for genome-wide association studies. Nature Genetics 42: 355-360.

**Supplemental materials:**

Methods S1

Figure S1: The principal components analysis based on genotyping data from 5,800 SNPs. (A). The proportion of variance explained by each of 20 principal components. (B) The population structure displayed by PC1 and PC2. (C) The population structure displayed by PC1 and PC3.

(D) The population structure displayed by PC2 and PC3.

Figure S2: The distribution of the pairwise genetic distance (Manhattan distance) from 784 barley landraces.

Figure S3: Relationship of barley landrace accessions based on principal components. PC1 and PC2 are depicted relative to a map of the landrace distribution in Africa and Eurasia.

Figure S4: Exome capture target density (dark blue line), crossover rate in cM/Mb (purple line), the genomic distribution of SNPs identified in 62 barley landraces (vertical light blue lines), and 9K iSelect SNPs (red triangle) for seven chromosomes. Crossover rates were calculated using 9K SNPs. SNP genetic positions are based on the genetic map of Muñoz-Amatriaín et al. (2011).

Figure S5: The derived site frequency spectrum (SNPs with the inferred ancestral state) for: (A) 2,806 SNPs from the 9K iSelect genotyped in the 784 landraces and (B) 340,260 SNPs with exome-capture resequencing data in 62 landraces. Ancestral state was inferred based on majority state from *H. murinum spp. glaucum* resequencing mapped to the Morex assembly. For all SNPs: (C) Minor allele frequency, for 5,800 SNPs from the 9K iSelect and (D) for 482,714 SNPs from exome-capture resequencing.

Figure S6: The heat map of the Pearson pairwise correlation coefficient for 22 environmental variables.

Figure S7: The distribution of ranked *F*ST from two- and three-level comparisons of elevation and latitude.

Figure S8: Venn diagram for the *F*SToutliers from the comparisons of elevation, high and low latitude, and growth habit.

Figure S9: The geographic distribution of the SNPs with high *F*ST. (A) The geographic distribution of allelic types of 9K SNP SCRI\_RS\_153793 with highest *F*ST = 0.505. The *F*ST was from the low latitude (LL) comparison. (B) The geographic distribution of allelic types of 9K SCRI\_RS\_134850 with highest *F*ST = 0.390. The *F*ST was from the 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 indicate the ancestral allele.

Figure S10: Venn diagram for the candidate SNPs significantly associated with three categories of environmental variables: precipitation, temperature, and geographic factors.

Figure S11: The difference between the replaced SNPs and queried SNPs not in the exome-capture data. Panel (A) shows the minor allele frequency (MAF) and (B) the physical distance.

Figure S12: 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 dashed lines are candidate SNP locations. The negative signs on the x-axis refer to positions downstream of the candidate SNP.

Figure S13: (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 tolerance-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.

Figure S14: (A) The linkage disequilibrium (LD) analysis of candidate SNP SCRI\_RS\_235243 significantly 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: Details of the exome capture data from 62 landraces.

Table S3: Summary statistics for the top three independent components (IC) and principal components (PCs) calculated from 19 BIOs.

Table S4: The samples size for each partition for FST comparisons.

Table S5: Known flowering time-related genes list.

Table S6: Known cold tolerance-related genes list.

Table S7: Known drought tolerance-related genes list.

Table S8: The annotation of the 155 significant SNPs identified by environmental association.

Table S9: The average and standard deviation of *F*ST calculated by various comparisons.

Table S10: 203 *F*ST outliers from elevation, high and low latitude, and growth habit comparisons.

Table S11: The nine overlapping SNPs identified by *F*ST outliers and association analysis approaches.

Table S12: *F*ST outliers from elevation comparison in a putative inverted region reported by Fang et al. 2014 without culling SNPs in strong LD.

Table S13: Barley flowering time genes with a signal of adaption, the numbers of SNPs genotyped those genes and the SNPs outlier by environmental association and *F*ST outlier SNPs.

**Supplemental datasets:**

Supplemental data 1: VCF file for the 6,152 SNPs without culling SNPs in complete LD.

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

Supplemental data 3: The physical positions of 9K SNPs.

Supplemental data 4: The annotations for SNPs called from exome capture resequencing data from 62 landraces.

Supplemental data 5: Phenotype matrix with 25 geographic and climatic variables for the environmental association.

Supplemental data 6: Inferred ancestral status for each 9K SNP.

Supplemental data 7: Inferred ancestral status for each exome resequencing SNP from 62 landraces.

Supplemental data 8: VCF file for SNPs called from exome-capture resequencing data from 62 landraces

Supplemental data 9: All *p*-values and Benjamini-Hochberg FDR-values from the environmental associations for 25 variables.

Supplemental data 10: All *p*-values and *F*ST from elevation, low and high latitude, longitude, and growth habit.