

Association Mapping of Agronomic QTLs in U.S. Spring Barley Breeding Germplasm

Duke Pauli, Gary J. Muehlbauer, Kevin P. Smith, Blake Cooper, David Hole, Don E. Obert, Steven E. Ullrich, and Thomas K. Blake*

Abstract

The use of genome-wide association studies (GWAS) to detect quantitative trait loci (QTL) controlling complex traits has become a popular approach for studying key traits in crop plants. The goal of this study was to identify the genomic regions of barley (*Hordeum vulgare* L.) that impact five agronomic and one quality trait in U.S. elite barley breeding lines, as well as to identify markers tightly linked with these loci for further use in barley improvement. Advanced recombinant inbred lines submitted to the U.S. Barley Coordinated Agricultural Project (CAP) were genotyped using a platform of 3072 single nucleotide polymorphism (SNP) markers from the barley oligonucleotide pool assays (BOPAs) 1 and 2. In each of 4 yr, approximately 770 lines were evaluated in a replicated, randomized complete block design under both irrigated and dryland conditions. This gave an overall population size of >3000 lines, which we analyzed in a hierarchical fashion, including analyzing the lines in aggregate using a mixed model to account for population structure and relatedness among the lines. We identified 41 significant marker–trait associations, of which 31 had been previously reported as QTL using biparental mapping techniques; 10 novel marker–trait associations were identified. The results of this work show that genes with major effects are still segregating in U.S. barley germplasm and demonstrate the utility of GWAS in barley breeding populations.

THE AIM OF QTL mapping in agronomically important crops is to find associations between genomic regions and traits that are of importance to the value of that crop. Once a significant association is found, the markers delimiting what genomic region can be used for selection in a breeding program, with the aim of increasing gain per unit time (Lande and Thompson, 1990). By identifying regions of the genome that impact traits of interest, we also gain insight into the overall genetic architecture of traits, which may subsequently influence breeding methods used to improve the trait. The traditional approach to identify genomic regions of interest has been biparental mapping. In biparental mapping, a population segregating for the trait of interest is created by intermating two parents that differ for the respective trait. This technique has proven extremely useful, as indicated by the plethora of published studies using the approach (Bernardo, 2008). However, biparental mapping is not without its limitations, including: only a small fraction of species-wide allelic diversity is sampled by using two parents; the estimation of the allelic effects are

D. Pauli and T.K. Blake, Plant Sciences and Plant Pathology Dep., Montana State Univ., 119 Plant Bioscience Building, Bozeman 59717; G.J. Muehlbauer and K.P. Smith, Dep. of Agronomy and Plant Genetics, Univ. of Minnesota, 411 Borlaug Hall, 1991 Upper Buford Cir., St. Paul 55108; B. Cooper, Limagrain Cereal Seeds, 2040 SE Frontage Rd., Fort Collins, CO 80525; D. Hole, Plants, Soils, and Climate Dep., Utah State Univ., 2325 Old Main Hill RGH 112A, Logan, UT 84322; D.E. Obert, Limagrain Cereal Seeds, 4846 E. 450 N., Lafayette, IN 47905; S.E. Ullrich, Dep. of Crop and Soil Science, Washington State Univ., Pullman 99164. Received 15 Nov. 2013. *Corresponding author (blake@montana.edu).

Abbreviations: BOPA, barley oligonucleotide pool assay; CAP, U.S. Barley Coordinated Agricultural Project; FDR, false discovery rate; GWAS, genome-wide association studies; LD, linkage disequilibrium; MAF, minor allele frequency; P3D, population parameters previously determined; PC, principal component; PCA, principal component analysis; QTL, quantitative trait loci; SNP, single nucleotide polymorphism.

Published in The Plant Genome 7
doi: 10.3835/plantgenome2013.11.0037
© Crop Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

restricted to the genetic backgrounds used; the limited number of recombination events within the population leads to poor localization of the QTL; and the need to construct a population segregating for the trait of interest (Jannink et al., 2001). Although biparental mapping has proved valuable to crop improvement, advances in genomic technologies and statistical methods have led to the implementation of GWAS.

Genome-wide association studies take advantage of historical recombination that has occurred within a germplasm collection by sampling numerous lines from a broader population not constrained by specific crossing. The associations between genotype and phenotype are dependent on linkage disequilibrium (LD), the nonrandom association of alleles, being broken down by many generations of recombination. In contrast to biparental mapping, the number of markers implemented in GWAS must be substantially larger so that LD is ensured between causative alleles and markers (Newell et al., 2012). High marker density has become possible by the rapidly declining cost of marker development and assays. Genome-wide association studies can realize higher mapping resolution of QTL due to the decreased size of linkage blocks. This has important implications in applied breeding, since the significant marker will be more tightly linked to the QTL of interest, and therefore the likelihood of recombination between the marker and QTL will be reduced. Genome-wide association studies are not without their own complications—most notably, the potential for the confounding effect of population structure which can lead to spurious associations (Marchini et al., 2004). Several statistical methods have been used to account for population structure, including principal component analysis (PCA; Price et al., 2006) as well as fitting mixed linear models to account for population structure, marker, and polygenic effects, termed $Q + K$ model (Yu et al., 2006).

Barley represents one of the earliest domesticated crops, and today is the fourth largest cereal crop in terms of area and biomass harvested (<http://faostat.fao.org>, verified 8 Apr. 2014). Barley has also had the advantage of being the focus of a large international collaborative effort to develop new genomic technologies to aid in the genetic understanding of the crop, as well as to assist in breeding efforts. The Barley CAP was designed to conduct association mapping within breeding program materials provided by public and private U.S. breeding programs.

The overall structure of the Barley CAP was for each of 10 breeding programs to submit 96 elite lines each for 4 yr for evaluation. The submitted germplasm contained a wide array of genetic diversity, as lines were two and six-rowed, winter and spring growth habit, and of malt, feed, and food end use (Hamblin et al., 2010). Subsets of the submitted entries were evaluated in numerous trials, spanning the entire region of barley production in the United States so that phenotype data could be collected and cataloged for further study. Since the germplasm of the Barley CAP consisted of elite breeding material,

identified beneficial alleles could be incorporated into the creation of new cultivars with limited deleterious effects (i.e., linkage drag). By identifying novel, useful alleles and associated diagnostic markers, barley breeders will be better able to develop cultivars at a more rapid pace to meet environmental changes as well as production and market demands.

Key agronomic traits that are routinely evaluated in barley yield trials include grain yield, plant height, heading date, grain test weight, and kernel plumpness. Grain protein content is also commonly measured due to its large impact on malting barley acceptance, since protein levels that are too high lead to grain being rejected for malt use (Schwarz and Li, 2011; See et al., 2002). Improvement of these key traits, especially grain yield, kernel plumpness, and grain protein content, is critical in the development of new barley varieties (Schwarz and Li, 2011).

A large number of QTL studies have been conducted on agronomic traits in barley using traditional biparental mapping populations (Hayes et al., 2003), but the employment of these QTLs in applied breeding has been limited due to the shortcomings listed above. Conducted in barley, GWAS have already proven useful in studying multiple traits (Cockram et al., 2010; Comadran et al., 2008, 2011; Massman et al., 2010; Ramsay et al., 2011; Roy et al., 2010; Wang et al., 2011, 2012; Zhou and Steffenson, 2013a, 2013b) so its application to agronomic traits should prove to be beneficial. Because agronomic traits are evaluated and used in selection on a continual basis, efficient marker-assisted selection for these traits could accelerate the development of new barley lines.

In the present study, we sought to identify marker-trait associations for yield, heading date, plant height, test weight, kernel plumpness, and protein content. To achieve this, we evaluated all of the spring barley entries in the Barley CAP over the course of 4 yr in one location under dryland and irrigated conditions. Using these data, in conjunction with the genotypic data, we conducted GWAS to identify the regions of the barley genome that impact these traits in our given target environment. By using such a large and diverse population for study, we were able to identify several major QTLs impacting these traits, supporting much of the previous work conducted on agronomic trait analysis. We were also able to identify 10 novel marker-trait associations that impact plant height (three associations), kernel plumpness (four), test weight (two), and yield (one). Some of these associations had a relatively large impact on the respective trait and should be quite useful. Overall, the marker-trait associations identified in this work will be valuable tools which can be directly utilized for the improvement of barley cultivars.

Material and Methods

Germplasm and Experimental Design

The entire spring Barley CAP germplasm panel was used for these analyses. This panel consists of lines submitted by eight spring barley breeding programs which include the following: Montana State University; USDA-ARS; Aberdeen, ID; Busch Agricultural Resources Inc.; University of Minnesota; North Dakota State 2-rowed; North Dakota State 6-rowed; Utah State University; and Washington State University. Each year, participating breeding programs submitted 96 lines, each of advanced material that included varieties and advanced lines that were inbred to at least the F_4 generation, which gave a total annual population size of 768 individuals. The evaluation of the Barley CAP spring lines was conducted over a period of 4 yr, giving a grand total of 3070 lines (Aberdeen omitted 2 lines in 2006). An irrigated and dryland field trial was conducted in the years 2006, 2007, and 2009; 2008 was lost due to hail. In 2010, only the dryland trial was conducted. The germplasm used in this panel represented the current genetic diversity of U.S. spring barley breeding efforts.

The lines that were submitted by each program in the respective year were evaluated in Bozeman, MT, at the Arthur Post Research Farm (45°40' N, 111°09' W). The representative panel of each year was planted in a randomized complete block design with two replications in two environments: dryland and irrigated. Plots were 1.22 m in length by four rows wide, seeded at a rate of 20 g per plot and managed with standard practices. Feekes scale was used to assess the developmental stages of the plants (Large, 1954). The 10 common check cultivars used throughout the experiments were: AC Metcalfe, Baronesse, Craft, Eslick, Geraldine, Harrington, Haxby, Hockett, Robust, and Tradition. Checks were included in subsequent association analyses.

Data from an onsite weather station was also collected and analyzed to assess yearly growing conditions using PROC GLM in SAS v. 9 (SAS Institute, 2012).

Genotype and Phenotype Data

All submitted lines in the Barley CAP were genotyped at the USDA-ARS Biosciences Research Lab located in Fargo, ND. Lines were genotyped using two Barley Oligo Pool Assays (BOPA1 and BOPA2), composed of 1536 SNPs on the Illumina GoldenGate platform (Illumina Inc., San Diego, CA). The design and development of the two BOPAs is described in detail in Close et al. (2009). For placement of markers, the genetic map generated by Close et al. (2009) was used. It is available for downloading on barley HarvEST database (Wanamaker and Close, 2012).

All data used in this study are available from The Triticeae Toolbox, <http://triticeaetoolbox.org> (verified 8 Apr. 2014), which now houses The Hordeum Toolbox (Blake et al., 2012). The genotype data were downloaded from The Hordeum Toolbox, setting the minor allele frequency (MAF) to 0.01, and the maximum missing data to 20%. The associated phenotype data was downloaded as well.

Five agronomic traits: plant height, heading date, percentage kernel plump, grain test weight, and yield; and one quality trait: grain protein content, were analyzed in this study. Plant height was measured as the average height of an arbitrary sample of plants gathered from the middle of the plot. Height is reported as distance, in centimeters, from the ground to the top of the inflorescence, excluding awns. Heading date was assigned when 50% of the spikes had emerged from the boot, reported as day of the year. Protein and kernel plumpness were both reported as percentages. Kernel plumpness is the percentage of sample remaining on top of a 0.24 by 1.9 cm slotted screen after 30 s on a plump shaker. Grain protein content was measured using near infrared spectroscopy with an Infratec Grain Analyzer (Foss, Hillerod, Denmark). Test weight was reported as grams/liter, and yield as kilograms/hectare. Data for protein and kernel plumpness were not transformed, so that results and estimated allele effects could be interpreted in a meaningful manner. Descriptive statistics were calculated using PROC MEAN and trait means were analyzed using PROC GLM in SAS v. 9. To assess the homogeneity of variances of the six different traits among respective data sets (e.g., 2006 dryland, 2006 irrigated, 2007 dryland), Levene's test was conducted in SAS v. 9.3 using PROC GLM.

Population Structure

Given that all lines from the eight barley breeding programs were analyzed in aggregate for each year as well as all years, strong population structure was expected. To assess this structure, PCA was conducted using the SNP marker data for all lines in a given year using the program TASSEL v. 3.0 (Bradbury et al., 2007). Single nucleotide polymorphism data was imputed using a three-nearest neighbors algorithm, with distances measured in Manhattan distance with an unweighted average. The covariance and number of components options were selected in the PCA analysis window, and the resulting principal component (PC) loadings for each line were exported for the creation of graphs in the statistical program R (R Development Core Team, 2012).

Population structure was also investigated using the program STRUCTURE (v. 2.3.4) (Falush et al., 2003; Pritchard et al., 2000). A reduced set of 136 markers that were spaced approximately 9 cM apart were used to estimate population structure and to calculate a subpopulation membership matrix (**Q**). Two to 15 hypothetical subpopulations were modeled using a burn-in of 10,000 cycles with 50,000 iterations, with 20 independent runs at each subpopulation level. The entire Barley CAP spring panel was used for assessing population structure, and the values from the **Q** matrix for each of the lines were used in subsequent analyses. The optimal subpopulation level was chosen using the method of Evanno et al. (2005). Once the optimal subpopulation number was found, it was used in the final analysis conducted by using a burn-in of 50,000 and 200,000 iterations to calculate the final **Q** matrix.

Association Analyses

Since the goal of the study was to identify marker-trait associations in the entire U.S. spring Barley CAP population, GWAS were conducted in three steps, based on how the phenotype data was utilized. Analysis 1 was conducted using the least square means for each environment and year. PROC GLM in SAS v. 9.3 was used to calculate the line least square means in each environment for each year (i.e., 2006 dryland, 2006 irrigated, 2007 dryland, 2007 irrigated). Genome-wide association studies were then performed on each set of phenotype data, and the results were compared across environments and years. Marker-trait associations detected in multiple years and/or across multiple environments were deemed most significant. Analysis 2 was based on using yearly averages. Again, the PROC GLM statement was used to calculate line least square means across each replication in both environments, so that there was one phenotype data set for years 2006, 2007, and 2009. In 2010, only the dryland trial data was available; therefore, the yearly averages were not used in this analysis. Analysis 3 was based on using standardized data sets. The PROC STANDARD statement was used to standardize the adjusted means data from each year in both the dryland and irrigated conditions using their respective means and standard deviations giving distributions of $N(0,1)$. Since each distribution was similar, they were then combined across years for each condition giving two overall data sets; the dryland standardized set comprised of 3057 lines and the irrigated standardized set comprised of 2293 lines (again, no 2010 irrigated trial).

Association analyses were conducted using the R package GAPIT (Lipka et al., 2012). The **Q + K** mixed-model approach was used, setting the number of PCs to 3 for all analyses conducted, given the population structure present in the germplasm. The options of compression and population parameters previously determined, P3D, were selected for all analyses as well (see Lipka et al., 2012, for further explanation). Marker-trait associations that had a false discovery rate (FDR) adjusted p -value < 0.05 were retained for further analyses. Markers with a MAF $< 5\%$ were not retained for analysis. The results of the analyses were checked using the program TASSEL v. 3.0 (Bradbury et al., 2007) to verify that the obtained results were consistent. This was done since both programs incorporate the same features such as P3D and compression.

The association analyses were also performed in an identical manner using the **Q** matrix generated from STRUCTURE.

Combined Results

Since the same data were being used in all three analyses, the results were compared across the three analyses. Marker-trait associations were first grouped together based on mapping to a common 2 to 3 cM region of the genome. These marker-trait association groups were further ranked by assigning a simple rating of one to three to indicate whether they were found in one, two, or all three analyses.

Those with a rating of two or three were retained for further analysis. The FDR adjusted p -value was lowered to 0.01, and marker-trait association groups that did not have at least one marker meeting this criterion were removed. Marker-trait association groups were also assessed on whether the same marker was repeatedly found significant, as well as multiple markers that had the same map position. These marker-trait associations were further cross-referenced against all reported QTLs in the GrainGenes 2.0 database (USDA-ARS, 2012), as well as through reports in the literature. Although the QTLs reported in GrainGenes are based on various biparental populations with varying map lengths, comparing significant marker-trait associations against this database provided a qualitative way to compare our results with prior studies.

Results

Phenotypic and Genotypic Data

Summary information for phenotypic data for all 4 yr, as well as the two environments (dryland and irrigated), are presented in Fig. 1 as well as Table 1. The results of an ANOVA to assess the means of the six traits with respect to genotype, environment, and year found that there were highly significant differences ($p < 10 \times 10^{-5}$) for all of the comparisons (results not shown). The summary information as well as the ANOVA results readily indicates that there was a large amount of phenotypic variation present in the Barley CAP spring barley germplasm. The standard deviations for each trait across years and environments appear to be stable, indicating that the amount of phenotypic variation in a trait was similar across years. However, the results of Levene's test for the homogeneity of variance did indicate that there were significant differences ($p < 0.001$) in the variances for all six traits (results not shown). The four growing seasons were also not dramatically different (Supplemental Tables S1 and S2, Supplemental Fig. S1).

The summary of genotypic data is presented in Table 2. An average of 775 genotypes were analyzed each year (including the 10 common checks included in each year) with the average number of markers meeting the requirements of MAF > 0.01 and maximum missing data $< 20\%$ being 2603. The trait heritabilities for each year and environment did fluctuate as expected, but overall were fairly consistent. The following are the mean heritabilities; yield, 0.50; plant height, 0.56; heading date, 0.55; kernel plumpness, 0.68; test weight, 0.64; and protein content, 0.60. These estimates are all relatively high which is beneficial for conducting GWAS.

Population Structure

The population structure of the U.S. Barley CAP spring barley lines was investigated so population structure could be accounted for in the association analyses. The results of the PCA revealed that there was significant population structure, which came as little surprise given that breeding lines comprised the population. On average, PC 1

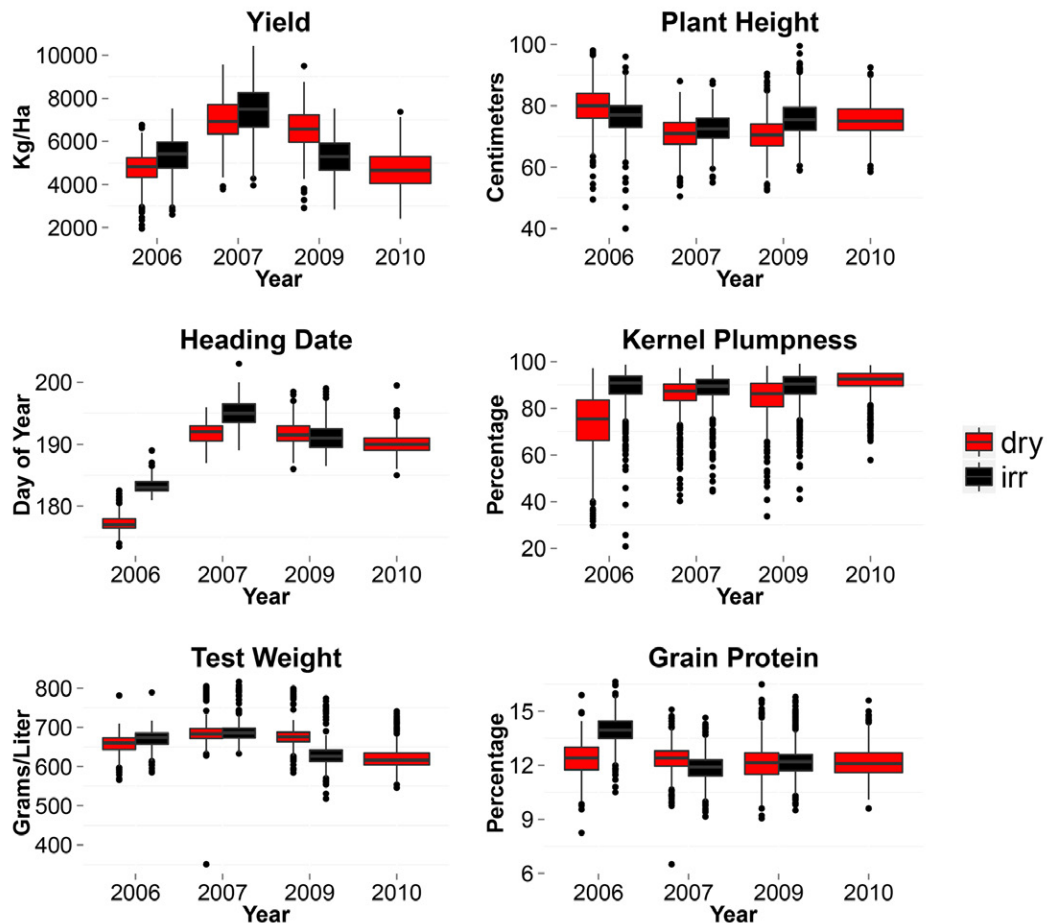


Figure 1. Boxplots of the phenotypic data showing distributions for the six traits investigated based on year for both dryland (dry) and irrigated (irr) environments. Note width of box for 2010, indicating that only the dryland data was available for that year.

accounted for 30.98% of the variation in the germplasm (Table 2) and served to separate the genotypes into two subpopulations of two and six-rowed varieties (Fig. 2). This division also clearly demarcates the eight breeding programs and the amount of genetic diversity contained therein. The second PC accounted for 6.01% of the variation in the germplasm and further separated the two-rowed programs, but had limited impact on the six-rowed breeding programs. Finally, the third PC accounted for 4.33% of the variation in the CAP germplasm. The structure present in the Barley CAP germplasm was consistent across years; the PC values were relatively similar for each year, with little variation among them.

The results from the STRUCTURE analysis found that the optimal number of subpopulations, K , was equal to six (results not shown), implying significant population structure. These results are consistent with those reported by Wang et al. (2012), who also found that $K = 6$ in the Barley CAP spring germplasm. The six subpopulations were divided into three, two-rowed populations and three, six-rowed populations, which approximately correspond to the breeding programs. The Utah germplasm was unique in that the subpopulation was comprised almost entirely of Utah lines. Compared with the PCA, the groupings were similar.

Association Analyses of U.S. Barley CAP Spring Germplasm

Association mapping was conducted for one quality and five agronomic traits to identify marker-trait associations in the U.S. Barley CAP spring germplasm. Table 3 summarizes the results found in each analysis and presents summary information of overall results. Our approach to these hierarchical analyses was to start with the single year, single environment data sets to identify environment specific or moderate effect marker-trait associations, looking for associations that were detected across years and environments. We then proceeded to the yearly adjusted means and standardized data sets so that we could assess the consistency of these marker-trait associations by comparing their results in all three analyses, focusing on the associations that were detected in multiple analyses. By using this hierarchical approach, we were better able to evaluate the validity of the marker-trait associations. The number of significant marker-trait associations that identified unique genomic regions (accounting for overlap in results among the three analyses) is as follows: heading date, 5; plant height, 9; kernel plumpness, 11; protein content, 9; test weight, 5; and yield, 2. The number of associations with supporting information in GrainGenes, as well as the published

Table 1. Summary information for phenotypic data of the six traits investigated. Data is from all 4 yr across both dryland (Dry) and irrigated (Irr) environments. Trait heritabilities estimated by the R package GAPIT.

Trait	Year and environment		Mean	SD	Range	h^2
Yield, kg/ha	2006	Dry	4761.7	727.4	4812.0	0.45
		Irr	5323.2	898.9	4939.0	0.56
	2007	Dry	6999.7	966.9	6103.0	0.43
		Irr	7483.9	1059.2	6479.0	0.46
	2009	Dry	6659.2	893.6	6588.0	0.65
		Irr	5377.6	878.6	4689.0	0.41
	2010	Dry	4744.9	906.4	5552.0	0.57
	Plant height, cm	2006	Dry	79.8	6.6	48.5
Irr			76.7	5.8	56.0	0.50
2007		Dry	70.7	5.2	37.5	0.49
		Irr	72.4	5.3	33.0	0.50
2009		Dry	70.7	5.6	38.0	0.66
		Irr	75.9	6.1	40.5	0.58
2010		Dry	75.2	5.2	34.5	0.66
Heading date, DOY [†]		2006	Dry	177.4	1.4	9.0
	Irr		183.2	1.1	8.0	0.57
	2007	Dry	191.8	1.8	9.0	0.32
		Irr	194.9	2.1	14.0	0.42
	2009	Dry	192.0	2.0	13.0	0.59
		Irr	191.4	2.6	14.5	0.68
	2010	Dry	190.2	1.7	15.5	0.67
	Kernel plumpness, %	2006	Dry	73.8	12.5	67.5
Irr			88.5	8.4	77.9	0.64
2007		Dry	84.4	8.6	57.1	0.64
		Irr	86.4	8.9	54.2	0.67
2009		Dry	84.9	8.6	64.5	0.82
		Irr	88.7	7.6	58.0	0.81
2010		Dry	90.7	8.5	58.1	0.89
Test weight, g/L		2006	Dry	657.0	23.5	215.0
	Irr		671.0	22.2	204.2	0.38
	2007	Dry	681.0	31.6	480.2	0.60
		Irr	684.0	25.9	229.8	0.62
	2009	Dry	680.9	31.4	228.9	0.84
		Irr	633.1	32.4	256.1	0.81
	2010	Dry	622.6	28.9	217.1	0.85
	Grain protein content, %	2006	Dry	12.4	0.9	7.7
Irr			13.9	0.8	6.2	0.38
2007		Dry	12.3	0.9	8.7	0.54
		Irr	11.8	0.8	5.6	0.56
2009		Dry	12.1	1.0	7.5	0.66
		Irr	12.2	0.9	6.3	0.69
2010		Dry	12.0	1.0	7.1	0.82

[†] DOY, day of year.

literature, is as follows: heading date, 5; plant height, 6; kernel plumpness, 7; protein content, 9; test weight, 3; and yield, 1. For the sake of brevity, refer to Table 4 for complete information on the identified marker-trait associations, as well as corresponding references in which QTLs were previously reported. A brief summary of the six traits is presented below, and the results from the STRUCTURE based analysis are reported in

Table 2. Summary information for the data sets as well as values for the population structure parameters, as estimated with principal component (PC) analysis, used in the association analyses.

Parameter	2006	2007	2009	2010	4 yr mean
No. of lines [†]	776	771	778	777	775.5
No. of markers [‡]	2664	2599	2592	2557	2603
PC 1, % [§]	31.44	30.53	29.93	32.01	30.98
PC 2, %	6.01	6.05	5.92	6.04	6.01
PC 3, %	4.91	4.68	4.22	3.49	4.33

[†] Number of lines evaluated in each year.

[‡] Number of markers downloaded meeting the requirements of MAF > 0.01 and maximum missing data < 20%.

[§] The principal components for each year and the percentage variation that it explains.

Supplemental Table S3. The Manhattan plots from the analyses using PCA are shown in Supplemental Fig. S2. The results from the association analyses using the Q matrix generated by STRUCTURE were not substantially different from those generated from using PCA; however, there were some minor differences. For simplicity, the results from the association analyses using PCA are described here, with the differences highlighted.

Heading Date

Our analyses for heading date found five marker-trait associations located on chromosomes 1H, 2H, 3H, 4H, and 7H that were consistently identified. Cross-referencing these significant associations with previous work of others, all five of these marker-trait associations have been previously identified in biparental mapping studies. We identified QTLs in the region near the two photoperiod response genes, *Ppd-H1* and *Ppd-H2*, on chromosomes 2H and 1H, respectively, that collectively explained >2% of the trait variation (Laurie et al., 1995). The other major QTL that we identified was on chromosome 3H at 126.27 cM, which is the genomic region where the *denso* gene has been mapped. The *denso* gene has been shown to impact both flowering time and plant height (Barua et al., 1993; Bezant et al., 1996; Laurie et al., 1995; Pan et al., 1994). Two other marker-trait associations consistently detected were on chromosome 7H at 37.55 cM and 4H at 96.59 cM, both of which have been previously identified (Backes et al., 1995; Laurie et al., 1995; Marquez-Cedillo et al., 2001).

The results of our analysis for heading date are similar to those of Wang et al. (2012), who also investigated this trait using the Barley CAP germplasm and associated genotyping platform. They identified the marker-trait associations on chromosomes 2H at 63.53 cM, 3H at 126.27 cM, and 7H at 37.55 cM; two of their most significant markers were the same as those identified in the present study (markers 12_30265 at 2H, 63.53 cM and 12_30893 at 7H, 37.55 cM). Wang et al. (2012) highlight that the marker 12_30893 is located within the *VRN-H3* gene, which controls vernalization and is a homolog to *Flowering Locus T* in *Arabidopsis* (Yan et al., 2006).

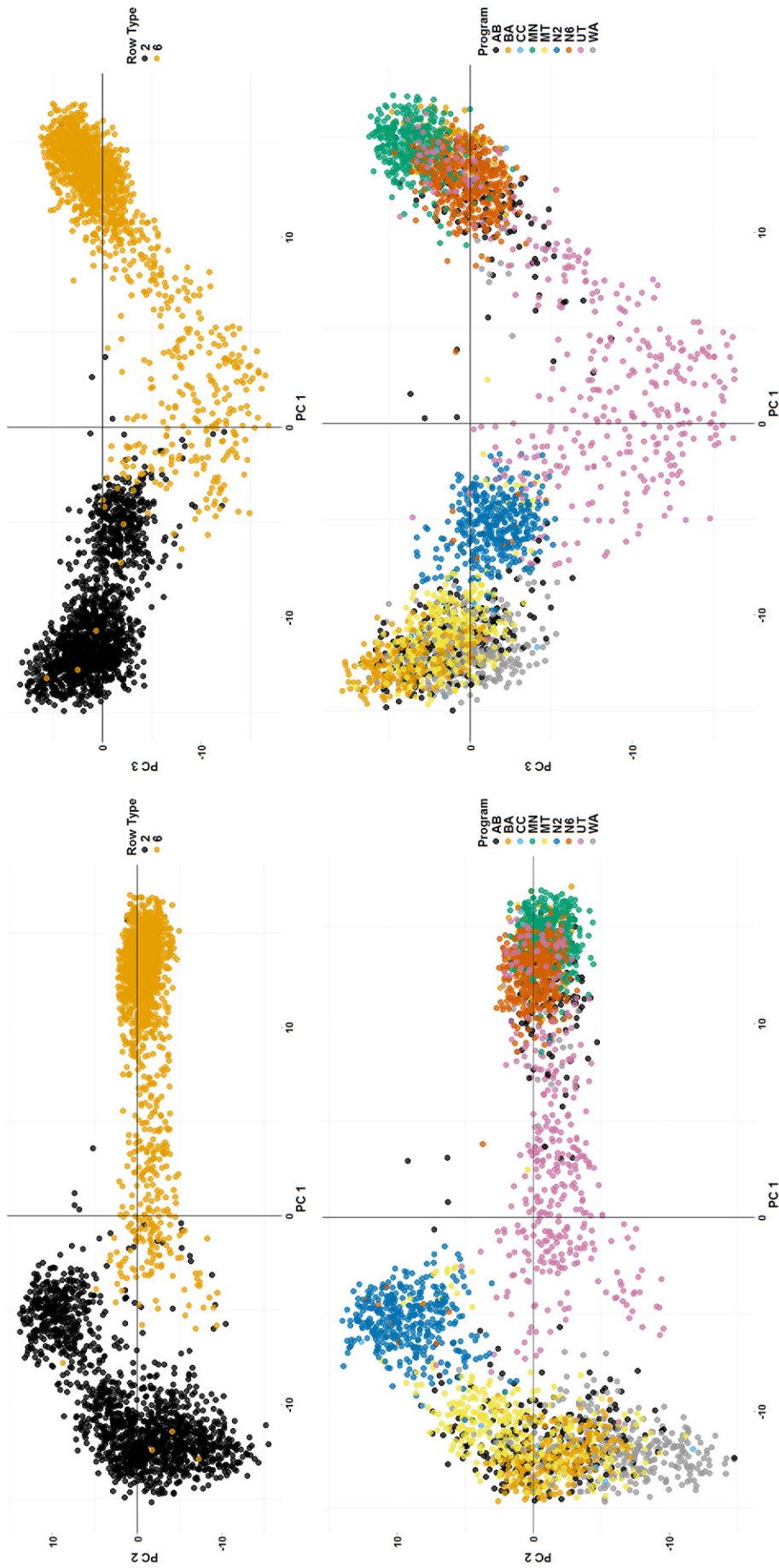


Figure 2. Principal component (PC) analysis of U.S. Barley Coordinated Agricultural Project (CAP) spring germplasm. PC 1 accounts for 30.98% of the variation in the CAP germplasm and separates the lines into two and six-rowed varieties. PC 2 accounts for 6.01% of the variation and further subdivides the two-rowed populations; note clustering of six-rowed breeding programs. The eight breeding programs are University of Idaho (AB), Busch Agricultural Resources Inc. (BA), University of Minnesota (MN), Montana State University (MT), North Dakota State 2-rowed (N2), North Dakota State 6-rowed (N6), Utah State University (UT), and Washington State University (WA). CC stands for Barley CAP Core panel, which are standard barley cultivars.

Table 3. Summary of marker-trait associations found in the U.S. Barley Coordinated Agricultural Project using principal component analysis (PCA). The table summarizes the number of significant marker-trait associations found in PCA 1, 2, and 3; numbers include markers mapping to the same position. The table also summarizes how many of those unique genomic regions were identified in 1, 2, or all 3 analyses. The number of putative QTLs as well as QTLs supported in GrainGenes 2.0 and published literature are reported. Average false discovery rate (FDR) adjusted *p*-values and average minor allele frequencies are also reported for each separate analysis. When FDR < 0.05, markers were below the 0.05 threshold but above the 0.01 significance level.

Trait	No. of significant associations						No. of significant associations in combined results				
	PCA 1 [†]		PCA 2		PCA 3		1 PCA	2 PCA	3 PCA	Putative QTLs	GrainGenes and literature
	FDR < 0.01	FDR < 0.05	FDR < 0.01	FDR < 0.05	FDR < 0.01	FDR < 0.05					
Heading date	3	13	7	1	10	11	6	2	3	5	5
Plant height	17	8	8	13	5	15	13	4	5	9	6
Kernel plumpness	22	28	33	51	123	8	6	9	10	11	7
Protein content	20	12	27	17	24	32	7	8	7	9	9
Test weight	19	26	26	28	29	51	5	12	2	5	3
Yield	1	7	1	4	11	30	3	1	1	2	1
Total	176		216		349		40	36	28	41	31
Average FDR <i>p</i> -value	0.017		0.016		0.014						
Average MAF	0.28		0.3		0.27						

[†] Principal Component Analysis 1 used the single-year and single-environment data sets for association analyses. Analysis 2 used the yearly adjusted means data set, averaged over dryland and irrigated environments. Analysis 3 used standardized data. The single-year and single-environment data sets were standardized using their respective means and standard deviations. These were then combined across environments to generate two data sets, standardized dryland and standardized irrigated.

Although none of the significant marker-trait associations that we are reporting are novel, the consistency of our results, compared with those of previous studies, lends credence to the basic genetic architecture of heading date in barley, namely that it is a trait controlled by a few, large-effect QTLs, with other minor effect QTLs most likely involved.

Plant Height

We identified nine significant marker-trait associations for plant height, located on all of the chromosomes, with the exception of chromosome 2H; three of these associations were novel. Plant height and heading date are traits that are typically correlated in barley, which is indicative of pleiotropic gene action (Bezant et al., 1996). This was borne out by our results, in which two of the significant marker-trait associations for plant height were found in the same genomic locations as those for heading date, but not in identical positions. These marker-trait associations were on chromosomes 3H at position 127.1 cM, and 7H at position 39.04 cM, and have been previously reported (Barua et al., 1993; Bezant et al., 1996; Laurie et al., 1995). Both of these associations had relatively high r^2 values in the context of our results, 1.71 and 3.31%, respectively. The other significant associations for plant height were found on chromosomes 1H, 3H, 4H, and 5H, and collectively accounted for 6.87% of plant height variation. All have been previously reported (see Table 4 for references). The novel marker-trait associations that we identified in this work are on chromosomes 1H, 5H, and 6H, which together explained an additional 6.14% of plant height variation. The associations on 5H at position 113.83 cM and 6H at position 0 cM did have low minor allele frequencies (0.07), which is at the limit of detection. However, their FDR *p*-values and multiple

detections in our analyses provides strong evidence that these are not false positive associations but are true significant regions of the barley genome that impact plant height. The novel marker-trait association on chromosome 1H at 47.47 cM was the only association in this study that was not also found in the analyses using the STRUCTURE generated Q matrix. The FDR *p*-values from the two analyses in which it was identified were 0.067 and 0.071. Although the FDR *p*-values are above the 0.05 threshold, they are only marginally so, and given the results from the analyses using PCA, it seems reasonable this could be a true association.

Kernel Plumpness

We detected eleven significant marker-trait associations for kernel plumpness on chromosomes 1H, 2H, 4H, and 7H, of which four were novel. Together, these QTLs explained 8.86% of the trait variation (see Table 4 for references). These associations all had very good FDR *p*-values ($p < 10 \times 10^{-6}$), with corresponding MAF values between 0.1 and 0.2. The combination of the low FDR *p*-values and minor allele frequencies providing adequate sample sizes lends strength that these associations are not false positives, but represent genomic regions affecting kernel plumpness. A noticeable feature of the kernel plumpness trait was the high number of significant markers that were below an FDR *p*-value of 0.01. This was due to areas of the genome that were identified as significant which had multiple markers map to the same position (e.g., eight markers all mapped to chromosome 1H at 55.49 cM, a region found to be significant). The results from the analyses using the Q matrix from STRUCTURE identified all the same marker-trait associations that were found using PCA. The difference in results were that marker-trait associations were not found in as many of

Table 4. Significant marker-trait associations identified in the U.S. Barley Coordinated Agricultural Project spring germplasm based on analyses using principal component analysis to control for population structure.

Trait	Marker	Chr	cM	MAF	FDR	r ²	QTL	Effect of minor allele	Rating	Type	Map	References
Hddt	11_20772	1H	139.79	0.16	5.98E-05	0.73	QHd1H.140	Pos	2	QHd.HaMo-1H.2 (145.72)	Hordeum-QTLConsensus-Agronomic-1	Laurie et al. (1995); Marquez-Cedillo et al. (2001)
Hddt	12_30265	2H	63.53	0.39	0.0014	1.38	QHd2H.64	Pos	3	QHd.lgDo-2H (61.82) QHd.HaMo-2H (67.26) QHd.umr-2H.1	Hordeum-QTLConsensus-Agronomic-2 Hordeum-QTLConsensus-Agronomic-2	Backes et al. (1995); Laurie et al. (1995); Marquez-Cedillo et al. (2001); Mestlin et al. (2003)
Hddt	12_31525	3H	126.27	0.09	0.0014	1.29	QHd3H.126	Pos	3	QHd.DiMo-3H (118.86–138.62)	Hordeum-QTLConsensus-Agronomic-3	Pan et al. (1994)
Hddt	12_30554	4H	96.59	0.32	0.0069	0.42	QHd4H.97	Pos	2	QHd.lgTr-4H (90.93)	Hordeum-QTLConsensus-Agronomic-4	Laurie et al. (1995)
Hddt	12_30893	7H	37.55	0.32	0.0017	1.23	QHd7H.36–38	Pos	3	QHd.lgDo-7H.1 (41.42) QHd.lgTr-7H.1 (41.42) QHd.HaMo-7H (40.15–48.60)	Hordeum-QTLConsensus-Agronomic-7 Hordeum-QTLConsensus-Agronomic-7 Hordeum-QTLConsensus-Agronomic-7	Backes et al. (1995); Laurie et al. (1995); Marquez-Cedillo et al. (2001)
Plht	12_31276	1H	27.35	0.23	0.0069	1.50	QH1H.27	Neg	2	QH.SiMo-1H (14.76–26.80) QPh-1H.30 (30.15)	Hordeum-QTLConsensus-Agronomic-1 Barley, OPA 2011, Consensus	Hayes et al. (1999); Berger et al. (2012)
Plht	12_31467	1H	47.47	.50	0.0069	1.40	QH1H.47	Pos	2	novel	Hordeum-QTLConsensus-Agronomic-3	Marquez-Cedillo et al. (2001); Rode et al. (2012)
Plht	11_10224	3H	56.4	0.41	0.00076	2.34	QH3H.56	Pos	2	QH.HaMo-3H.3 (49.06–62.23) QH-3H.56	Hordeum-QTLConsensus-Agronomic-3	Marquez-Cedillo et al. (2001); Rode et al. (2012)
Plht	12_30096	3H	127.1	0.39	0.0054	1.71	QH3H.127	Neg	3	QH3HL	Hordeum-QTLConsensus-Agronomic-3	Barua et al. (1993); Bezaat et al. (1996); Laurie et al. (1993)
Plht	11_21070	4H	26.19	0.4	8.34E-05	1.03	QH4H.26	Neg	3	QH.ei.pil-4H.1 (24)	Hordeum-Pillen-4H	Pillen et al. (2003)
Plht	11_10870	5H	0	0.06	0.0025	2.00	QH5H.0	Pos	3	QH.HaTr-5H.1 (9.05)	Hordeum-QTLConsensus-Agronomic-5	Spanner et al. (1999); Tinkler et al. (1995)
Plht	12_30619	5H	113.83	0.07	4.41E-05	2.63	QH5H.114	Pos	2	novel	Hordeum-QTLConsensus-Agronomic-5	
Plht	12_30319	6H	0	0.07	0.0024	2.11	QH6H.0	Neg	3	novel	Hordeum-QTLConsensus-Agronomic-5	Bezaat et al. (1996)
Plht	12_10218	7H	39.04	0.3	5.15E-05	3.31	QH7H.39	Neg	3	QH7HL	Hordeum-QTLConsensus-Agronomic-5	
Plump	12_30348	1H	55.49	0.12	7.35E-05	0.69	QKp1H.55	Pos	3	novel	Hordeum-HxM-base-2H	Marquez-Cedillo et al. (2001)
Plump	11_20340	2H	85.92	0.45	4.61E-12	1.75	QKp2H.86	Neg	3	QKps.nab-2H (89.30–97.1)	Hordeum-HxM-base-2H	Marquez-Cedillo et al. (2001)
Plump	12_31095	2H	116.49	0.22	0.0029	0.42	QKp2H.113–116	Neg	3	QKp.HaMo-2H (113.49)	Hordeum-OWB-OPA2008-2H	Marquez-Cedillo et al. (2001); Szucs et al. (2009)
Plump	11_10065	2H	130.01	0.17	7.62E-06	0.84	QKp2H.130	Neg	3	novel	Hordeum-OWB-OPA2008-2H	
Plump	11_20777	4H	26.66	0.41	4.17E-07	3.08	QKp4H.27	Pos	3	QKp.HaMo-4H (25.94)	Hordeum-OWB-OPA2008-4H	Marquez-Cedillo et al. (2001); Szucs et al. (2009)
Plump	11_20894	5H	2.09	0.19	0.0019	0.45	QKp5H.2	Neg	3	QKp.HaTr-5H (0)	Hordeum-OWB-OPA2008-5H	Mather et al. (1997); Szucs et al. (2009)
Plump	11_20708	5H	51	0.32	0.0029	0.42	QKp5H.51	Pos	2	QKp.nab-5H.1 (42–49.5)	Hordeum-HxM-base-5H	Marquez-Cedillo et al. (2001)
Plump	11_21061	5H	110.26	0.46	1.60E-04	2.31	QKp5H.110	Pos	3	QKp.HaMo-5H.1 (113.77)	Hordeum-OWB-OPA2008-5H	Marquez-Cedillo et al. (2001); Szucs et al. (2009)
Plump	11_20076	7H	0	0.11	6.01E-10	1.43	QKp7H.0	Neg	3	novel	Hordeum-OWB-OPA2008-5H	
Plump	11_10346	7H	61.32	0.16	0.0026	0.43	QKp7H.61	Neg	3	QKp.nab-7H (60.60–76.7)	Hordeum-HxM-base-7H	Marquez-Cedillo et al. (2001)
Plump	11_21201	7H	98.5	0.12	3.18E-08	1.18	QKp7H.98	Neg	3	novel	Hordeum-HxM-base-7H	
Protein	11_11400	2H	53.53	0.06	2.50E-04	0.64	QGr2H.54	Neg	2	QGr.SiMo-2H.1 (55.21)	Hordeum-OWB-OPA2008-2H	Hayes et al. (1993); Szucs et al. (2009)
Protein	11_20340	2H	85.92	0.45	1.64E-05	1.72	QGr2H.86	Neg	2	QGr.nab-2H.1 (89.3–97.1)	Hordeum-HxM-base-2H	Marquez-Cedillo et al. (2001)
Protein	11_21070	4H	26.19	0.4	3.50E-04	0.74	QGr4H.26	Neg	3	QGr.HaMo-4H (25.94)	Hordeum-OWB-OPA2008-4H	Marquez-Cedillo et al. (2001); Szucs et al. (2009)
Protein	12_20770	5H	42.32	0.08	0.0040	1.27	QGr5H.42	Neg	2	QGr.DiMo-5H.1 (44.85)	Hordeum-OWB-OPA2008-5H	Oziel et al. (1996); Szucs et al. (2009)

(cont'd)

Table 4. Continued.

Trait	Marker	Chr	cM	MAF	FDR	r ²	QTL	Effect of minor allele	Rating	Type	Map	References
Protein	11_10095	5H	137.16	0.24	1.50E-04	1.92	<i>Qgp5H.137</i>	Pos	3	QPro.pil-5H.1 (141)	Hordeum-Pillen-5H	Pillen et al. (2003)
Protein	11_10254	5H	179.06	0.45	0.0057	1.28	<i>Qgp5H.177-180</i>	Pos	3	QGpc.DiMo-5H.2 (176.13)	Hordeum-OWB-OPA2008-5H	Oziel et al. (1996); Szucs et al. (2009)
Protein	12_10199	6H	45.44	0.16	9.21E-14	2.12	<i>Qgp6H.45</i>	Neg	3	QGpc.6H-Hvm/74		See et al. (2002)
Protein	12_31199	7H	86.44	0.09	0.0028	1.43	<i>Qgp6H.86</i>	Pos	3	QGpc.HoMo-7H (90.65)	Hordeum-OWB-OPA2008-7H	Marquez-Cedillo et al. (2001); Szucs et al. (2009)
Protein	11_21209	7H	129.91	0.33	0.0074	1.14	<i>Qgp7H.130</i>	Pos	2	QGpc.DiMo-7H (122.25)	Hordeum-OWB-OPA2008-7H	Oziel et al. (1996); Szucs et al. (2009)
Tw	12_11173	1H	101.45	0.09	1.70E-04	1.27	<i>Qtw1H.101</i>	Neg	3	novel		
Tw	11_20340	2H	85.92	0.45	6.36E-06	1.30	<i>Qtw2H.86</i>	Neg	2	QTw.nob-2H (89.3-97.1) Qtw2H.89 (89.68)	Hordeum-HxM-base2H Barley, OPA 2011, Consensus	Marquez-Cedillo et al. (2001); Berger et al. (2012)
Tw	12_30865	4H	33.38	0.43	0.0017	0.46	<i>Qtw4H.33</i>	Pos	2	QTw.nob-4H.2 (25-35.8)	Hordeum-HxM-base-4H	Marquez-Cedillo et al. (2001)
Tw	11_20682	6H	75.21	0.41	0.0064	1.47	<i>Qtw6H.75</i>	Neg	2	QTw.nob-6H.2 (67.6-75)	Hordeum-HxM-base-6H	Marquez-Cedillo et al. (2001)
Tw	12_31441	7H	70.4	0.45	0.00078	0.83	<i>Qtw7H.70</i>	Neg	3	novel		
Yld	12_10579	2H	132.48	0.37	0.00019	1.70	<i>QYld2H.132</i>	Neg	2	QYld.Bkly2H.2 (134.8-146.2)	Hordeum-QTLConsensus-Agronomic-2	Bezant et al. (1997)
Yld	12_31010	3H	55.57	0.08	0.015	1.45	<i>QYld3H.52</i>	Pos	3	novel		

[†]Agronomic traits: Hddt, heading date; Plht, plant height; Plump, kernel plumpness; Protein, grain protein content; Tw, Grain test weight; Yld, grain yield. Marker, most significant marker showing association with agronomic trait based on Barley Oligonucleotide Pool Assay nomenclature. Chr, chromosome on which significant marker-trait association was identified. cM, position on chromosome, in centimorgans, where significant marker-trait association was identified according to Close et al. (2009). MAF, minor allele frequency. FDR, false discovery rate *p*-value, calculated according to Benjamini and Hochberg (1995). r², percentage of trait variation explained by the SNP, taken from most significant association from the three analyses. QTL, tentative qualitative trait loci name for significant marker-trait association; name consists of trait, chromosome, and location. Effect of minor allele indicates whether the minor allele has a positive or negative impact on the respective trait. Rating indicates the rating given to the marker-trait association, see text for explanation. Type indicates whether the marker-trait association has been previously identified; if so, reported QTL name is given. Map gives the name of the supporting map in GrainGenes 2.0 in which previously identified QTL are reported.

the data sets using the Q matrix; for example, an association was only found in Analysis 2, whereas with PCA it may have been found in analyses 1, 2, and 3.

Grain Protein Content

We identified nine marker-trait associations for grain protein content on chromosomes 2H, 4H, 5H, 6H, and 7H (see Table 4 for complete information and references). The marker-trait association detected on chromosome 6H had the lowest FDR *p*-value. This association was located at 45.44 cM and accounted for 2.12% of the trait variation, which is relatively high with respect to our results. The association at 45.44 cM is the marker-trait association that has the largest impact; in other analyses we have conducted it generally accounts for a 1% difference in protein levels between lines (data not shown). This QTL has been previously mapped by See et al. (2002) using a biparental population made by crossing Karl, a low protein cultivar, with ‘Lewis’. Further work conducted on this population by Distelfeld et al. (2008) suggested that the gene responsible for this QTL is the barley NAC transcription factor *HvNAM-1*. This QTL has been named *Gpc-6H*. Given the high FDR *p*-value for the marker 12_10199 ($p < 9.21 \times 10^{-14}$), and its large effect on protein content, this marker appears to be in tight linkage with the *HvNAM-1* gene and serves as an excellent marker for marker-assisted breeding in the reduction of grain protein content.

Test Weight

In our analyses for test weight, we identified five significant marker-trait associations with two of those being novel. Associations were identified on all of the chromosomes except 3H and 5H. The three associations that were not novel have been mapped by Marquez-Cedillo et al. (2001) using the Harrington-Morex double haploid population. The locations of those three associations are as follows: 2H at 85.92 cM, 4H at 33.38 cM, and 6H at 75.21 cM.

One of the novel associations that we detected was located on chromosome 1H at 101.45 cM. The MAF value for this marker was low, 0.09, but the FDR *p*-value of 1.7×10^{-4} indicates that this not a false positive. The marker-trait association was also detected in all three analyses conducted, which also supports it being a significant marker-trait association. The other significant marker-trait association that we identified was located on chromosome 7H at 70.4 cM. The MAF value for this allele was 0.45, with a corresponding FDR *p*-value of 7.8×10^{-4} . This marker-trait association was also found in all three analyses, which strengthens the position that this is a true association that could be utilized for variety improvement.

Yield

We identified fewer marker-trait associations for yield compared with the other traits we investigated. Given the heterogeneous nature of the CAP lines, the multiple growing seasons, and the use of small, 1.22-m plots, it was challenging to get an accurate estimate of the true yield potential of the lines. The estimates of heritability for yield

were typically lower, which would influence the results we obtained (Table 1). In our analyses, we identified two marker-trait associations that were on chromosomes 2H and 3H. The chromosome 2H QTL at 132.48 cM may be the same as that identified by Bezant et al. (1996), using a traditional biparental mapping population. Their identified QTL spanned a region from 134.8 to 146.2 cM. Although our identified association falls just outside of this region, it may be the same QTL. In the analysis using the Q matrix form STRUCTURE, this association was only identified in one of the data sets, whereas in the PCA-based analysis, it was identified in analyses 1 and 3. The other marker-trait association that we identified was located on chromosome 3H at 55.57 cM; this association was novel. The MAF value for this allele was 0.08 and the corresponding FDR p -value was only 0.015, but this association did show up in all three analyses. The r^2 values for the chromosome 2H and 3H QTLs were 1.7 and 1.45%, respectively. With regard to the association on chromosome 3H, we also identified this genomic region in our plant height analyses. The correlation between plant height and yield varies considerably, so it is hard to say if it is pleiotropic gene action or two distinct QTLs.

Discussion

One of the main goals of the Barley CAP was to integrate new genomic technologies into the development of superior cultivars. To achieve this goal, the development and implementation of a common marker based platform was used to identify QTLs that could be used in barley breeding. Over the course of 4 yr, a total of approximately 3000 advanced generation lines were phenotyped in one location in two environments to provide a large data set that could be leveraged for conducting GWAS. By conducting these metaanalyses, we hoped to take full advantage of the Barley CAP germplasm, which represents the current genetic diversity in elite spring barley breeding programs in the United States. In analyzing this data set, we sought to identify the regions of the barley genome that impact five agronomic traits (heading date, plant height, kernel plumpness, grain test weight, and yield) as well as one quality trait (grain protein content). This work demonstrates the utility of using breeding populations for GWAS.

Structure of U.S. Barley Spring Breeding Populations

It has been well demonstrated that population structure can lead to an increased number of false positive associations in GWAS (Atwell et al., 2010; Pritchard et al., 2000). Many crop species exhibit strong population structure due to factors such as geographical origins, reproductive nature, and selective breeding efforts by humans. Although population structure can adversely affect the results of association mapping, studies have been conducted in the Poaceae family, which exhibits highly structured populations. These studies include rice (*Oryza sativa* L., Agrama et al., 2007), wheat (*Triticum aestivum* L., Breseghello and Sorrels, 2006), and barley (Cockram et al., 2010; Massman et al., 2010; Wang et al., 2011). Since the Barley CAP consists of lines from

breeding populations which have been under high selective pressure, correction for population structure was needed. This was accomplished using PCA and a model-based analysis using the program STRUCTURE (Price et al., 2006; Pritchard et al., 2000).

The major population subdivision was detected between two and six-rowed barley types, which has been a common feature of population structure in barley (Brantestam et al., 2007; Hamblin et al., 2010; Hayes and Szucs, 2006). Breeders typically work within these groups, as crosses between the two groups result in progeny that are rarely suitable for cultivar development (Hayes et al., 2003). Because of this, the subdivision between two- and six-rowed barleys also led to the separation of the breeding programs, the two major six-rowed breeding programs are the University of Minnesota and the North Dakota six-rowed program, which are clearly demarcated in Fig. 2. The North Dakota two-rowed program was distinctly different from the other two-rowed programs, as it grouped away from the other two-rowed breeding programs.

The germplasm of the Utah program was the most distinct. It did not group strongly with either the two or six-rowed programs, but instead spread out along the axis of PC 1 and 2. When the third PC was plotted, the Utah germplasm again clustered by itself, indicating its uniqueness among the Barley CAP germplasm. This result was also found in the STRUCTURE analysis, as one subpopulation was composed almost entirely of the Utah germplasm. These results are most likely due to the fact that the Utah breeding program has been selecting for six-rowed barleys for livestock feed and agronomic performance instead of malting. Another contributing factor to Utah's unique genetic variance is the source of the parental germplasm. Most of the Utah breeding germplasm is based on North African material, which is unique among barley breeding programs in the United States. North American breeding programs are typically based on European and Asian germplasm (Horsley and Harvey, 2011).

This division of the Barley CAP germplasm into component subpopulations has been observed before. Hamblin et al. (2010) investigated the population structure of the Barley CAP germplasm, including winter barleys, and found approximately the same groupings of lines, that is, lines fell into respective breeding programs. The Utah population was not considered a separate population due to only 96 lines from the Utah program being included in the analysis. Zhou et al. (2012) also performed an investigation of the entire Barley CAP germplasm using both PCA and the program STRUCTURE. The results from their STRUCTURE analysis found that $K = 9$ was the optimal number of subpopulations, but their analysis included the two additional winter barley breeding programs, which would be expected to be distinctly different from the spring barley germplasm. Their results from using PCA are similar to the groupings that we are reporting.

When the population structure was assessed separately over the 4 yr, the same pattern emerged. The values for the amount of variation explained by each PC

remained consistent which led to similar groupings when graphed (Table 2 and Fig. 2). The consistency of the sub-populations in the Barley CAP from year to year is most likely due to the nature of using breeding populations for a germplasm source, as each program has its own genetic identity that changes slowly.

Association Analyses of Agronomic Traits

Conducting replicated yield trials of all the spring Barley CAP entries at one location afforded us the opportunity to use these data in conducting GWAS in a hierarchical fashion. Our approach was to first conduct the analyses on the single-year, single-environment data sets because this would allow us to detect marker-trait associations that are specific to an environment or perhaps have moderate effects. This also served as a way to identify associations that had large, consistent effects as they would be detected in multiple analyses across years and environments. The second analysis was conducted by averaging across the two environments within a year. By doing this, we hoped to identify those marker-trait associations with moderate to large effects, confirming those found in the single-year, single-environment analyses. Associations that were significant in only one of the seven data sets were more likely viewed as potential false positives.

We next proceeded to the data sets that were standardized using their respective means and standard deviations, as described in the Materials and Methods. Given that the first two analyses were based on data sets in which field variation was controlled (i.e., environment, replication, and block effects) we felt it acceptable to proceed with data standardization. We chose to use data standardization within each year as a way to remove year effects similar to what is practiced in animal breeding by the use of contemporary groups (Bourdon, 2000). This method has also been used in plant breeding (Skovmand et al., 2001). We also chose standardization to deal with the heteroscedasticity of the phenotypic data. Although our approach to these analyses may seem basic, it is this simplicity and reliance on few statistical assumptions that speak to the validity of the results. Given that we were unable to fully account for all the variation due to year effects, but still able to detect strong associations through this excess noise, supports the use of our method. Of the 41 marker-trait associations identified in our study, 26 of those were found in all three analyses. This indicates that the data standardization was able to produce results similar to the other two methods which did not rely on this type of data manipulation.

The use of data standardization allowed us to combine multiple years' worth of data to better localize marker-trait associations and achieve the highest resolution possible in the Barley CAP data. Up to this point, QTL localization has been poor (i.e., QTLs declared in regions spanning 10, even 30 cM). The SNP marker platform developed through the Barley CAP made it feasible to refine these regions to much smaller intervals. By using all of the lines, we were able to take maximum advantage of years of recombination events, leading to higher resolution and tighter

linkage between markers and causative genomic regions. This should be beneficial for the barley breeding community at large, since significant markers should be useful across multiple programs and environments despite being identified in just the Montana environment.

Another impetus for analyzing all lines in aggregate is that it increased the frequency of minor alleles. By increasing the frequencies of minor alleles to more moderate values, specifically those below 0.05 in the single-year, single-environment data sets, the statistical power to detect them increases, reducing the probability of false positives and assisting in identification of some potential moderate effect marker-trait associations (Mackay et al., 2009). This was evident in cases like the association found on chromosome 2H at 53.53 cM for grain protein content which had an overall MAF of 0.06. In the single-year, single-environment data sets, it was below 0.05. The MAF of 0.06 is at the limit of detection, but by combining all the lines we were able to detect it. This proved useful, as this region has been identified as a QTL and has been previously mapped with the Steptoe × Morex mapping population (Hayes et al., 1993).

A final advantage of combining the lines across the different years was the increase in allele replication. A potential drawback to our work is the lack of line replication present in the study. Lander and Botstein (1989) highlight that the power to test for the mean allelic effects between alleles is influenced by replication of lines. However, replication of alleles and population size are large contributing factors to the power of the test for differences between allele means. Knapp and Bridges (1990) investigated the relationship between line replication, allele replication, progeny population size, and QTL parameters. Their results showed that increasing the replication of alleles as well as the population size always led to an increase in power regardless of the number of replications of lines used for the study. Given these findings, the absence of line replication in our work is easily mitigated by the extensive replication of alleles and the vast population size.

Our analyses also incorporated the use of two methods to correct for population structure, the nonparametric approach of PCA, and the model-based approach used in the program STRUCTURE. The differences in results were not substantial, but it does demonstrate the sensitivity of the different uses of population structure control. The use of STRUCTURE does appear to represent a more conservative analysis. This is best exemplified by the results for the kernel plumpness trait. In our results from the PCA, nine of the eleven declared marker-trait associations were found in all three analyses, whereas in the STRUCTURE analysis only two were (both were novel associations). Also, four of those marker-trait associations in the STRUCTURE analysis were only found in one of the three analyses we conducted; three of these four associations were previously identified QTL in biparental mapping studies.

Agronomic Traits

The results we obtained using a genome-wide association approach agree with numerous independent, traditional

biparental mapping population studies conducted for agronomic traits, as can be seen in Table 4. This serves as validation in two ways: our results obtained by using a genome-wide association approach lends support to our method of analysis, and the results also serve to validate those QTLs identified in biparental populations which have their own inherent shortcomings. Although we found only a small number of novel marker-trait associations in this study, we believe that these novel associations are true associations since their effects were large enough to be detected in such a heterogeneous population. Although this large amount of heterogeneity limits our ability to detect marker-trait associations that have smaller effects, it does lead to the identification of those QTLs that have a major impact on traits. Also, detection of QTLs with minor effects is not as important, given that the application of marker-assisted selection for these QTLs would be largely ineffective; the use of genomic selection would be more applicable. By identifying those QTLs with large effects in a population with an assorted genetic background, we can be certain that these QTLs will maintain their effect as they are incorporated into other genetic backgrounds, avoiding the phenomenon of ghost or vanishing QTLs.

The r^2 values that we obtained in our results were relatively low; they ranged from 0.42 to 3.31%. These low values are not surprising, given the large amount of additive genetic variance and heterogeneous composition of the Barley CAP. With a larger amount of additive genetic variance, the trait variation is distributed over more markers than one would observe in a traditional biparental mapping study, as there are more segregating loci. Berger et al. (2012) performed a GWAS on a sample of 329 barley lines from the Virginia Tech winter barley program (also part of the Barley CAP) looking at several traits. The r^2 values that were obtained for the agronomic traits (heading date, plant height, test weight, and yield) ranged from 3 to 10%. Given that those values were obtained within just the Virginia Tech program, the results we obtained seem agreeable since the current study population encompasses eight breeding programs and a much larger sample size. Both of these results indicate that there may be more potential loci that impact these agronomic traits that remain to be found.

The marker-trait associations that we identified in the single-year, single-environment analyses are perhaps the most intriguing. Clearly there are statistical difficulties in validating them due to limited sample size, but they do offer the possibility that they are novel marker-trait associations with minor effects that might have a more prominent role in targeted environments. These marker-trait associations could serve as the genetic variability that could be used to produce optimal cultivars once the beneficial major genes have been fixed in breeding programs.

Our results also show the effects of assortative mating in self-pollinated crops as like phenotypes are more frequently mated with lines of similar phenotype, the basis of advanced cycle breeding (crosses based on good \times good; Falconer and Mackay, 1996). These effects are manifested by marker-trait associations being found

in clusters more often than not throughout the barley genome due to the formation of linkage blocks. These linkage blocks also represent untapped genetic variability that could be utilized through breakage of these blocks through recombination, and lead to new genetic gain in populations with unfavorable linkage (Simmonds, 1979).

The other salient feature of our results is that these major genes are still segregating in U.S. elite barley lines, which creates useful genetic variation for these agronomic traits. Most breeding programs employ advance cycle breeding, which can rapidly degrade genetic variability within a breeding program. With the genomic tools developed by the Barley CAP, breeders should be able to identify lines that are phenotypically similar but vary genetically, based on a kinship analysis using available marker data. This should aid in selection of new parents for line development by maintaining some level of genetic diversity while still intermating lines that are phenotypically similar. An example of this would be first selecting high-yielding lines from programs that are similar with regard to target environment. Next, by looking at a kinship matrix for those selected lines, given by $2f_{ij}V_a$, with f_{ij} representing the coefficient of coancestry and V_a the additive genetic variance, the amount of genetic diversity between those candidate lines could be assessed. Candidate lines that are sufficiently diverse could then be evaluated based on the QTL reported in this study. Candidate lines that carry contrasting favorable alleles at these QTL could then be crossed with the hope of producing transgressive segregates identified with marker-assisted selection. Furthermore, the marker information could be used to monitor the progress of breeding to assess if favorable allele frequencies are moving in the right direction representing genetic gain.

This study addresses one of the goals of the U.S. Barley CAP, which was to identify genetic resources that could be employed to meet the ever-changing demands of production. With this work, we show that there is the potential to identify and incorporate useful genetic variation from adapted lines into the development of new U.S. cultivars with superior performance, and that the resources of the Barley CAP will greatly aid in that endeavor.

Supplemental Information Available

Supplemental information is included with this article, including Supplemental Tables S1, S2, and S3.

Supplemental Figure S1. Average daily temperature at the Arthur Post Research Farm in Bozeman, MT. Dashed line represents the mean temperature for the respective growing season, the blue line represents mean daily temperature, and the red smoothed line is LOESS fitted line. Planting typically occurs approximately at Day 14 to 21 and harvest at approximately Day 110. Data collected at Arthur Post Research Farm.

Supplemental Figure S2. Manhattan plots of the results obtained from all analyses using PCA to control for population structure. Chromosome 10 represents markers that have not been mapped and are not included in the results.

Acknowledgments

This research was supported by USDA-CSREES-NRI Grant No. 2006-55606-16722 and USDA-NIFA Grant No. 2009-85606-05701, "Barley Coordinated Agricultural Project: Leveraging Genomics, Genetics, and Breeding for Gene Discovery and Barley Improvement." Funding was also provided by USDA-NIFA Grant No. 2011-68002-30029, "Triticeae Coordinated Agricultural Project." Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the USDA.

References

- Agrama, H.A., G.C. Eizenga, and W. Yan. 2007. Association mapping of yield and its components in rice cultivars. *Mol. Breed.* 19:341–356. doi:10.1007/s11032-006-9066-6
- Atwell, S., Y.S. Huang, B.J. Vilhjalmsón, G. Willems, M. Horton, Y. Li, D. Meng, A. Platt, A.M. Tarone, T.T. Hu, R. Jiang, N.W. Mulyati, X. Zhang, M.A. Amer, I. Baxter, B. Brachi, J. Chory, C. Dean, M. Debieu, J. de Meaux, J.R. Ecker, N. Faure, J.M. Kniskern, J.D.G. Jones, T. Michael, A. Nemri, F. Roux, D.E. Salt, C. Tang, M. Todesco, M.B. Traw, D. Weigel, P. Marjoram, J.O. Borevitz, J. Bergelson, and M. Nordborg. 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465:627–631. doi:10.1038/nature08800
- Backes, G., A. Graner, B. Foroughi-Wehr, G. Fishbeck, G. Wenzel, and A. Jahoor. 1995. Localization of quantitative trait loci (QTL) for agronomic important characters by the use of a RFLP map in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 90:294–302. doi:10.1007/BF00222217
- Barua, U.M., K.J. Chalmers, W.T.B. Thomas, C.A. Hackett, V. Lea, P. Jack, B.P. Forster, R. Waugh, and W. Powell. 1993. Molecular mapping of genes determining height, time to heading, and growth habit in barley (*Hordeum vulgare*). *Genome* 36:1080–1087. doi:10.1139/g93-143
- Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc., B* 57:289–300.
- Berger, G.L., S. Liu, M.D. Hall, W.S. Brooks, S. Chao, G.J. Muehlbauer, B.-K. Baik, B. Steffenson, and C.A. Griffey. 2012. Marker-trait associations in Virginia Tech winter barley identified using genome-wide mapping. *Theor. Appl. Genet.* 126:693–710. doi:10.1007/s00122-012-2011-7
- Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Sci.* 48:1649–1664. doi:10.2135/cropsci2008.03.0131
- Bezant, J., D. Laurie, N. Pratchett, J. Chojecki, and M. Kearsey. 1996. Marker regression mapping of QTL controlling flowering time and plant height in a spring barley (*Hordeum vulgare* L.) cross. *Heredity* 77:64–73. doi:10.1038/hdy.1996.109
- Bezant, J., D.A. Laurie, N. Pratchett, J. Chojecki, and M. Kearsey. 1997. Mapping QTL controlling yield and yield components in a spring barley (*Hordeum vulgare* L.) cross using marker regression. *Mol. Breed.* 3:29–38. doi:10.1023/A:1009648220852
- Blake, V.C., J.G. Kling, P.M. Hayes, J.-L. Jannink, S.R. Jillella, J. Lee, D.E. Matthews, S. Chao, T.J. Close, G.J. Muehlbauer, K.P. Smith, R.P. Wise, and J.A. Dickerson. 2012. The Hordeum Toolbox: The barley coordinated agricultural project genotype and phenotype resource. *Plant Gen.* 5:81–91. doi:10.3835/plantgenome2012.03.0002
- Bourdon, R.M. 2000. Understanding animal breeding. Prentice Hall, Upper Saddle River, NJ.
- Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635. doi:10.1093/bioinformatics/btm308
- Brantestam, A.K., R. von Bothmer, C. Dayteg, L. Rashaal, S. Tuveesson, and J. Weibull. 2007. Genetic diversity changes and relationships in spring barley (*Hordeum vulgare* L.) germplasm of Nordic and Baltic areas as shown by SSR markers. *Genet. Resour. Crop Evol.* 54:749–758. doi:10.1007/s10722-006-9159-4
- Bresegghello, F., and M.E. Sorrells. 2006. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172:749–758.
- Close, T.J., P.R. Bhat, S. Lonardi, Y. Wu, N. Rostoks, L. Ramsay, A. Druka, N. Stein, J.T. Svensson, S. Wanamaker, S. Bozdog, M.L. Roose, M.J. Moscou, S. Chao, R.K. Varshney, P. Szűcs, K. Sato, P.M. Hayes, D.E. Matthews, A. Kleinjohs, G.J. Muehlbauer, J. DeYoung, D.F. Marshall, K. Madhishetty, R.D. Fenton, P. Condamine, A. Graner, and R. Waugh. 2009. Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics* 10:582. doi:10.1186/1471-2164-10-582
- Cockram, J., J. White, D.L. Zuluaga, D. Smith, J. Comadran, M. Macaulay, Z. Luo, M.J. Kearsey, P. Werner, D. Harrap, C. Tapsell, H. Liu, P.E. Hedley, N. Stein, D. Schulte, B. Steuemagel, D.F. Marshall, W.T.B. Thomas, L. Ramsay, I. Mackay, D.J. Balding, T.A. Consortium, R. Waugh, and D.M. O'Sullivan. 2010. Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. *Proc. Natl. Acad. Sci. USA* 107:21,611–21,616. doi:10.1073/pnas.1010179107
- Comadran, J., J.R. Russell, A. Booth, A. Psarayi, S. Ceccarelli, S. Grando, A.M. Stanca, N. Pecchioni, T. Akar, A. Al-Yassin, A. Benbelkacem, H. Ouabbou, J. Bort, F.A. Van Eeuwijk, W.T.B. Thomas, and I. Romagosa. 2011. Mixed model association scans of multi-environment trial data reveal major loci controlling yield and yield related traits in *Hordeum vulgare* in Mediterranean environments. *Theor. Appl. Genet.* 122:1363–1373. doi:10.1007/s00122-011-1537-4
- Comadran, J., Jr., F.A. Russell, S. van Eeuwijk, S. Ceccarelli, M.B. Grando, A.M. Stanca, N. Pecchioni, A.M. Mastrangelo, T. Akar, A. Al-Yassin, A. Benbelkacem, W. Choumane, H. Ouabbou, R. Dahan, J. Bort, J.L. Araus, A. Psarayi, I. Romagosa, C.A. Hackett, and W.T.B. Thomas. 2008. Mapping adaptation of barley to droughted environments. *Euphytica* 161:35–45. doi:10.1007/s10681-007-9508-1
- Distelfeld, A., A. Korol, J. Dubcovsky, C. Uauy, T. Blake, and T. Fahima. 2008. Colinearity between the barley grain protein content (GPC) QTL on chromosome arm 6HS and the wheat *GPC-B1* region. *Mol. Breed.* 22:25–38. doi:10.1007/s11032-007-9153-3
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x
- Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman, UK.
- Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164:1567–1587.
- Hamblin, M.T., T.J. Close, P.R. Bhat, S. Chao, J.G. Kling, K.J. Abraham, T. Blake, W.S. Brooks, B. Cooper, C.A. Griffey, P.M. Hayes, D.J. Hole, R.D. Horsley, D.E. Obert, K.P. Smith, S.E. Ullrich, G.J. Muehlbauer, and J.-L. Jannink. 2010. Population structure and linkage disequilibrium in U.S. barley germplasm: Implications for association mapping. *Crop Sci.* 50:556–566. doi:10.2135/cropsci2009.04.0198
- Hayes, P., A. Castro, L. Marquez-Cedillo, A. Corey, C. Henson, B.L. Jones, J.G. Kling, D. Mather, I. Matus, C. Rossi, and K. Sato. 2003. Genetic diversity for quantitatively inherited agronomic and malting quality traits. In: R. von Bothmer et al., editors, Diversity in barley (*Hordeum vulgare*). Elsevier, Amsterdam. p. 201–226.
- Hayes, P., B. Liu, S. Knapp, F. Chen, B. Jones, T. Blake, J. Franckowiak, D. Rasmusson, M. Sorrells, and S.E. Ullrich. 1993. Quantitative trait locus effects and environmental interaction in a sample of North American barley germ plasm. *Theor. Appl. Genet.* 87:392–401. doi:10.1007/BF01184929
- Hayes, P., and P. Szucs. 2006. Disequilibrium and association in barley: Thinking outside the glass. *Proc. Natl. Acad. Sci. USA* 103:18,385–18,386. doi:10.1073/pnas.0609405103
- Horsley, R.D., and B.L. Harvey. 2011. Barley breeding history, progress, objectives, and technology: North America. In S.E. Ullrich (ed.) Barley: Production, improvement, and uses. Wiley-Blackwell, Oxford, UK. p. 171–186.
- Jannink, J.-L., M. Bink, and R.C. Jansen. 2001. Using complex plant pedigrees to map valuable genes. *Trends Plant Sci.* 6:337–342. doi:10.1016/S1360-1385(01)02017-9
- Knapp, S.J., and W.C. Bridges. 1990. Using molecular markers to estimate quantitative trait locus parameters: Power and genetic variances for unreplicated and replicated progeny. *Genetics* 126:769–777.
- Lande, R., and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743–756.
- Lander, E.S., and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–199.
- Large, E.C. 1954. Growth stages in cereals illustration of the Feekes scale. *Plant Pathol.* 3:128–129. doi:10.1111/j.1365-3059.1954.tb00716.x

- Laurie, D.A., N. Pratchett, J.H. Bezan, and J.W. Snape. 1995. RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter × spring barley (*Hordeum vulgare* L.) cross. *Genome* 38:575–585. doi:10.1139/g95-074
- Laurie, D.A., N. Pratchett, C. Romero, E. Simpson, and J.W. Snape. 1993. Assignment of the *denso* dwarfing gene to the long arm of chromosome 3(3H) of barley by use of RFLP Markers. *Plant Breed.* 111:198–203. doi:10.1111/j.1439-0523.1993.tb00630.x
- Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.G. Gore, E.S. Buckler, and Z. Zhang. 2012. GAPIT: Genome association and prediction integrated tool. *Bioinformatics* 28:2397–2399. doi:10.1093/bioinformatics/bts444
- Mackay, T.F.C., E.A. Stone, and J.F. Ayroles. 2009. The genetics of quantitative traits: Challenges and prospects. *Nat. Rev. Genet.* 10:565–576. doi:10.1038/nrg2612
- Marchini, J., L.R. Cardon, M.S. Phillips, and P. Donnelly. 2004. The effects of human population structure on large genetic association studies. *Nat. Genet.* 36:512–517. doi:10.1038/ng1337
- Marquez-Cedillo, L.A., P. Hayes, A. Kleinhofs, W.G. Legge, B.G. Rossnagel, K. Sato, S.E. Ullrich, and D.M. Wesenberg. 2001. QTL analysis of agronomic traits in barley based on the double haploid progeny of two elite North American varieties representing different germplasm groups. *Theor. Appl. Genet.* 103:625–637. doi:10.1007/PL00002919
- Massman, J., B. Cooper, R.D. Horsley, S. Neate, R. Dill-Macky, S. Chao, Y. Dong, P. Schwarz, G.J. Muehlbauer, and K.P. Smith. 2010. Genome-wide association mapping of Fusarium head blight resistance in contemporary barley breeding germplasm. *Mol. Breed.* 27:439–454. doi:10.1007/s11032-010-9442-0
- Mather, D.E., N.A. Tinker, D.E. LaBerge, M. Edney, B.L. Jones, B.G. Rossnagel, W. Legge, K.G. Briggs, R.B. Irvine, D.E. Falk, and K.J. Kasha. 1997. Regions of the genome that affect grain and malt quality in a North American two-row barley cross. *Crop Sci.* 37:544–554. doi:10.2135/cropsci1997.0011183X003700020039x
- Mesfin, A., K.P. Smith, R. Dill-Macky, C.K. Evans, R. Waugh, C.D. Gustus, and G.J. Muehlbauer. 2003. Quantitative trait loci for Fusarium head blight resistance in barley detected in a two-rowed by six-rowed population. *Crop Sci.* 43:307–318. doi:10.2135/cropsci2003.0307
- Newell, M.A., F.G. Asoro, M.P. Scott, P.J. White, W.D. Beavis, and J.-L. Jannink. 2012. Genome-wide association study for oat (*Avena sativa* L.) beta-glucan concentration using germplasm of worldwide origin. *Theor. Appl. Genet.* 125:1687–1696. doi:10.1007/s00122-012-1945-0
- Oziel, A., P.M. Hayes, F.Q. Chen, and B. Jones. 1996. Application of quantitative trait locus mapping to the development of winter-habit malting barley. *Plant Breed.* 115:43–51. doi:10.1111/j.1439-0523.1996.tb00869.x
- Pan, A., P.M. Hayes, F. Chen, T.H.H. Chen, T. Blake, S. Wright, I. Karsai, and Z. Bedo. 1994. Genetic analysis of the components of winterhardiness in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 89:900–910.
- Pillen, K., A. Zacharias, and J. Leon. 2003. Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 107(2):340–352. doi:10.1007/s00122-003-1253-9
- Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich. 2006. Principal component analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38:904–909. doi:10.1038/ng1847
- Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ramsay, L., J. Comadran, A. Druka, D.F. Marshall, W.T.B. Thomas, M. Macaulay, K. Mackenzie, C. Simpson, J. Fuller, N. Bonar, P.M. Hayes, U. Lundqvist, J.D. Franckowiak, T.J. Close, G.J. Muehlbauer, and R. Waugh. 2011. INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene TEOSINTE BRANCHED 1. *Nat. Genet.* 43:169–172. doi:10.1038/ng.745
- Rode, J., J. Ahlemeyer, W. Friedt, and F. Ordon. 2012. Identification of marker-trait associations in the German winter barley breeding gene pool (*Hordeum vulgare* L.). *Mol. Breed.* 30:831–843. doi:10.1007/s11032-011-9667-6
- Roy, J.K., K.P. Smith, G.J. Muehlbauer, S. Chao, T.J. Close, and B. Steffenson. 2010. Association mapping of spot blotch resistance in wild barley. *Mol. Breed.* 26:243–256. doi:10.1007/s11032-010-9402-8
- SAS Institute. 2012. SAS/STAT guide for personal computer. v. 9.3. SAS Inst., Cary, NC.
- Schwarz, P., and Y. Li. 2011. Malting and brewing uses of barley. In S.E. Ullrich, editor, *Barley: Production, improvement, and uses*. Wiley-Blackwell. p. 505.
- See, D., V. Kanazin, K. Kephart, and T.K. Blake. 2002. Mapping genes controlling variation in barley grain protein concentration. *Crop Sci.* 42:680–685. doi:10.2135/cropsci2002.0680
- Simmonds, N.W. 1979. Principles of crop improvement. Longman, New York.
- Skovmand, B., M.P. Reynolds, and I.H. Delacy. 2001. Searching genetic resources for physiological traits with potential for increasing yield. In: M.P. Reynolds et al., editors, *Application of physiology in wheat breeding*. CIMMYT, Mexico. p. 17–28.
- Spaner, D., B.G. Rossnagel, W.G. Legge, G.J. Scoles, P.E. Eckstein, G.A. Penner, N.A. Tinker, K.G. Briggs, D.E. Falk, J.C. Afele, P.M. Hayes, and D. Mather. 1999. Verification of a quantitative trait locus affecting agronomic traits in two-row barley. *Crop Sci.* 39:248–252. doi:10.2135/cropsci1999.0011183X003900010038x
- Szucs, P., V.C. Blake, P.R. Bhat, S. Chao, T.J. Close, A. Cuesta-Marcos, G.J. Muehlbauer, L. Ramsay, R. Waugh, and P.M. Hayes. 2009. An integrated resource for barley linkage map and malting quality QTL Alignment. *Plant Gen.* 2:134–140. doi:10.3835/plantgenome2008.01.0005
- Tinker, N.A., D.E. Mather, T.K. Blake, K.G. Briggs, T.M. Choo, L. Dahleen, S.M. Dofing, D.E. Falk, T. Ferguson, J.D. Franckowiak, R. Graf, P.M. Hayes, D. Hoffman, R.B. Irvine, A. Kleinhofs, W. Legge, B.G. Rossnagel, M.A. Saghai Maroof, G.J. Scoles, L.P. Shugar, B. Steffenson, S. Ullrich, and Kasha. 1995. Regions of the genome that affect agronomic performance in two-row barley. *Crop Sci.* 36:1053–1062. doi:10.2135/cropsci1996.0011183X003600040040x
- USDA-ARS. 2012. GrainGenes 2.0. <http://wheat.pw.usda.gov/GG2/index.shtml> (verified 8 Apr. 2014).
- Wanamaker, S., and T.J. Close. 2012. HarvEST: Barley v. 1.83. Available at <http://harvest.ucr.edu/> (verified 8 Apr. 2014). Univ. of California, Riverside.
- Wang, M., N. Jiang, T. Jia, L. Leach, J. Cockram, R. Waugh, L. Ramsay, B. Thomas, and Z. Luo. 2011. Genome wide association mapping of agronomic and morphologic traits in highly structured populations of barley cultivars. *Theor. Appl. Genet.* 124:233–246. doi:10.1007/s00122-011-1697-2
- Wang, H., K.P. Smith, E. Combs, T. Blake, R.D. Horsley, and G.J. Muehlbauer. 2012. Effect of population size and unbalanced data sets on QTL detection using genome-wide association mapping barley breeding germplasm. *Theor. Appl. Genet.* 124:111–124. doi:10.1007/s00122-011-1691-8
- Yan, L., D. Fu, C. Li, A. Blechl, G. Tranquilli, M. Bonafede, A. Sanchez, M. Valarik, S. Yasuda, and J. Dubcovsky. 2006. The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. *Proc. Natl. Acad. Sci. USA* 103:19,581–19,586. doi:10.1073/pnas.0607142103
- Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38:203–208. doi:10.1038/ng1702
- Zhou, H., G.J. Muehlbauer, and B. Steffenson. 2012. Population structure and linkage disequilibrium in elite barley breeding germplasm from the United States. *J. Zhejiang Univ.* 13:438–451. doi:10.1631/jzus.B1200003
- Zhou, H., and B.J. Steffenson. 2013a. Association mapping of septoria speckled leaf blotch resistance in U.S. barley breeding germplasm. *Phytopathology* 103:300–609.
- Zhou, H., and B.J. Steffenson. 2013b. Genome-wide association mapping reveals genetic architecture of durable spot blotch resistance in US barley breeding germplasm. In: G. Zhang et al., editors, *Advance in barley sciences*. Proc. 11th Int. Barley Genetics Symp. Springer and Zhejiang Univ. Press, Dordrecht, the Netherlands. p. 257–267.