Association Mapping of Malting Quality Quantitative Trait Loci in Winter Barley: Positive Signals from Small Germplasm Arrays


Abstract

Malting quality comprises one of the most economically relevant set of traits in barley (Hordeum vulgare L.). It is a complex phenotype, expensive and difficult to measure, that would benefit from a marker-assisted selection strategy. Malting quality is a target of the U.S. Barley Coordinated Agricultural Project (CAP) and development of winter habit malting barley varieties is a key objective of the U.S. barley research community. The objective of this work was to detect quantitative trait loci (QTL) for malting quality traits in a winter breeding program that is a component of the U.S. Barley CAP. We studied the association between five malting quality traits and 3072 single nucleotide polymorphisms (SNPs) from the barley oligonucleotide pool assay (BOPA) 1 and 2, assayed in advanced inbred lines from the Oregon State University (OSU) breeding program from three germplasm arrays (CAP I, CAP II, and CAP III). After comparing 16 models we selected a structured association model with posterior probabilities inferred from software STRUCTURE (QK) approach to use on all germplasm arrays. Most of the marker-trait associations are germplasm- and environment-specific and close to previously mapped genes and QTL relevant for malt and beer quality. We found alleles fixed by random genetic drift, novel unmasked alleles, and genetic-background interaction. In a relatively small population size study we provide strong evidence for detecting true QTL.

Barley (Hordeum vulgare L.) is the fourth most important cereal crop in terms of total world production (FAO-STAT, 2008), and malting quality is one of its most economically relevant traits. The malt quality phenotype is difficult and expensive to measure because there are many interrelated component traits that lead to malt of high quality (Hayes and Jones, 2000). The component traits, in turn, are encoded by many genes, each of which is subject to interaction with other genes and regulation by environmental factors such as temperature, moisture, and nutrition (Sparrow, 1971). Breeding for superior malting quality is challenging because many of the component traits are not normally measured until late in the breeding cycle when sufficiently large and uniform grain

Published in The Plant Genome 4:256–272. Published 5 Dec. 2011. doi: 10.3835/plantgenome2011.07.0020 © 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.
samples can be obtained for traditional micromalting and malting quality analyses (Mather et al., 1997). Assessment of these samples is further hampered by the time required for micromalting and malting quality analysis (Han et al., 1997a). Therefore, most breeding programs resort to making relatively few “good” quality crosses and ultimately testing a limited number of advanced lines for the target trait. As a consequence, there have been modest genetic gains for malting quality (Han et al., 1997b; Muñoz-Amatriain et al., 2010b). If the process of developing superior malting barley varieties is to be accelerated, other breeding methods are needed.

One approach is to discover the genes determining the components of malting quality and to use this information for marker-assisted selection (MAS) or genomic selection (GS). To date, biparental quantitative trait locus (QTL) mapping studies have been conducted to identify chromosome regions responsible for superior malting quality (see Szűcs et al. [2009] for a review). The results are often difficult to directly apply to cultivar development since QTL identified are frequently background specific and the parents of mapping populations are often exotic or one of the parents lack malting quality. Furthermore, favorable alleles at major QTL impacting target traits may already be fixed in relevant germplasm. Finally, some biparental QTL mapping studies may have low accuracy due to small population size (Dekkers and Hospital, 2002; Yu and Buckler, 2006).

Genome-wide association (GWA) mapping may overcome many of the limitations of the biparental approach (Jannink et al., 2001) and provide targets for MAS in relevant germplasm (Mather et al., 1997). The advantages of GWA include simultaneous assessment of broad diversity, higher resolution for fine mapping, effective use of historical data, and immediate applicability to cultivar development because the genetic background in which QTL are estimated is directly relevant for plant breeding (Kraakman et al., 2004). Additionally, GWA has been proposed as a promising tool for self-pollinating crop improvement and specifically barley breeding (Kraakman et al., 2004, 2006; Hayes and Szűcs, 2006; Stracke et al., 2009; Waugh et al., 2009; Roy et al., 2010; Bradbury et al., 2011; von Zitzewitz et al., 2011).

Many different models have been proposed for studying the association between molecular markers and phenotypic traits. The first association mapping analyses conducted on plants recommended correcting for spurious associations caused by population structure using STRUCTURE (Pritchard et al., 2000). This was the beginning of the posterior probabilities matrix inferred from software STRUCTURE (Q) model. Later, Parisseaux and Bernardo (2004) integrated the relationship matrix as random effects into a mixed model. Yu et al. (2006) proposed the use of a unified mixed models approach to correct for spurious associations caused not only by population structure but also by individual relatedness (and identical-by-descent alleles) using the mixed model including kinship matrix (K). They proposed the use of the structured association model with posterior probabilities inferred from software STRUCTURE (QK) that would account for the fixed effects of population structure (Q) and the random effects of individual relatedness (K).

The use of the Q fixed-effect matrix in this model was challenged by Price et al. (2006), who proposed the use of the fixed-effects matrix from principal component analysis (P) to account for population structure instead of the Q matrix. Empirical data and simulations showed that PK models outperformed QK in producing results with smaller type I error, larger power to detect true associations, and shorter computation time to obtain the fixed-effects matrix (Zhao et al., 2007). More recently, Zhu and Yu (2009) proposed the use of the fixed-effects matrix from nonmetric multidimensional scaling (M) to account for population structure instead of the Q or P matrix. They showed how this method outperforms previously proposed approaches using both simulated and empirical data sets. The choice of the random-effect matrix in the mixed model approach was also challenged (Zhao et al., 2007; Zhu and Yu, 2009). The most commonly used kinship matrices are K obtained from TASSEL (Bradbury et al., 2007), K obtained by the efficient mixed model association (EMMA) implementation (Kang et al., 2008), and K obtained by SPAGEDi (Hardy and Vekemans, 2002) using Ritland’s (1996) and Loiselle’s (Loiselle et al., 1995) kinship coefficients. It is, therefore, not completely clear which fixed-effect (population structure) and which random-effect (kinship) matrix should be used for association mapping of economically important traits, such as malting quality in barley.

Malting quality is one of the principal traits targeted for analysis and improvement by the U.S. Barley Coordinated Agricultural Project (Barley CAP, http://www.barleycap.org/, verified 29 Oct. 2011), and the development of winter habit malting barley varieties is a key objective of U.S. public and private sector barley breeding programs. Therefore, under the auspices of the Barley CAP, the Oregon State University (Corvallis, OR) barley breeding program has focused on malting quality in winter germplasm. Here we report results from association mapping using 3072 single nucleotide polymorphism (SNP) markers and five malting quality parameters: grain protein content (GPC), malt extract (ME), diastatic power (DP), α-amylase (AA) activity, and wort β-glucan (BG) content phenotyped in three germplasm arrays (CAP I, CAP II, and CAP III) from the Oregon State University (OSU) breeding program.

**MATERIALS AND METHODS**

**Germplasm**

We used the Oregon Barley CAP I, CAP II, and CAP III germplasm sets for this research. Full descriptors of this germplasm are available at The Hordeum Toolbox through the Triticea-CAP site (Triticea Coordinated Agricultural Project, 2010). Briefly, these germplasm sets consist of varieties, advanced lines, and genetic stocks of immediate use and relevance for breeding winter malting barley (see Fig. 1 for a description). Coordinated agricultural project I consists primarily of winter and facultative six-rowed lines from the OSU program. Coordinated agricultural
Figure 1. Population structure and relationship among lines in three association mapping populations (CAP I, CAP II, and CAP III). S: posterior probability of group belonging from STRUCTURE software. C1: light-color (OIL), Oregon intermated lines, lines derived by several cycles of inter-mating progeny derived from crosses among relatively few parents (Kold, Strider, 8BAb536, Excel, Legacy, and Orca); dark-color (NL), non-Oregon intermated lines, lines that are genetically distant to Oregon lines: Hundred, Eight-Twelve, UTWB940119, UTWB94061, UTWB971412, Dicktoo, and Luca. C2: dark-colors (dark-blue and blue) are Oregon CAP I related lines, sister lines with a sib in CAP I; blue are sister lines with a sib in CAP I (OIL); dark-blue are Oregon lines derived from crosses with a non-CAP I donor (Bu27); light-colors (white and light-blue) are unrelated germplasm; light-blue are Idaho lines not used as parents of CAP I, Nebraska lines not used as parents of CAP I, and Oregon hooded lines that share one parent (Kold) with CAP I while the other parent (Hoody) is not a parent of CAP I lines (SDNE); white are other unrelated lines. C3: light-color (Doyce), lines derived from crosses with Doyce; dark-color (OC3), Oregon intermated lines.
implemented in SPAGeDi (Hardy and Vekemans, 2002). Proposed by Bradbury et al., 2007 (TASSEL), Kang et al., 2008 [efficient mixed nonmetric multidimensional scaling (M)] or using one of the three kinship matrices described. K.T, kinship matrix estimated using TASSEL (Bradbury et al., 2007); K.E, kinship matrix estimated using efficient mixed models approach (Kang et al., 2008); K.S, kinship matrix estimated using SPAGeDi (Hardy and Vekemans, 2002).

Table 1. The 16 models compared in the Barley CAP I germplasm array.

<table>
<thead>
<tr>
<th>Kinship matrix(^1)</th>
<th>Structure covariate</th>
</tr>
</thead>
<tbody>
<tr>
<td>None(^2)</td>
<td>Q-STRUCTURE(^3)</td>
</tr>
<tr>
<td>TASSSEL</td>
<td>K.T</td>
</tr>
<tr>
<td>EMMA</td>
<td>K.E</td>
</tr>
<tr>
<td>SPAGeDi</td>
<td>K.S</td>
</tr>
</tbody>
</table>

\(^1\)Models with either no-kinship matrix or kinship matrices as the ones proposed by Bradbury et al., 2007 (TASSEL), Kang et al., 2008 (efficient mixed model association (EMMA)), or the one using Pitman’s (1996) coefficients implemented in SPAGeDi (Hardy and Vekemans, 2002).

\(^2\)No structure covariates used: naive, a simple test of association (Kruskal–Wallis) with no correction for population structure, or a mixed model without interred population structure as cofactor and using one of the three kinship matrices described. K.T, kinship matrix estimated using TASSEL (Bradbury et al., 2007); K.E, kinship matrix estimated using efficient mixed models approach (Kang et al., 2008); K.S, kinship matrix estimated using SPAGeDi (Hardy and Vekemans, 2002).

\(^3\)Q-STRUCTURE, models including population structure estimated from STRUCTURE (Pritchard et al., 2000) either in a fixed-effects model (posterior probabilities matrix inferred from software STRUCTURE [Q], without kinship matrix) or in a mixed model following Yu et al. (2008) with one of the three matrices described.

\(^4\)PCA, principal component analysis. Models including population structure estimated from a principal component analysis implemented in R (R Development Core Team, 2005) following Price et al. (2006) without a kinship matrix (fixed-effects matrix from principal component analysis [P]) or using one of the three kinship matrices described.

\(^5\)nMDS, nonmetric multidimensional scaling. Models including population structure estimated from a nonmetric multidimensional scaling analysis implemented in R (R Development Core Team, 2005) following Zhu and Yu (2009) without a kinship matrix (fixed-effects matrix from nonmetric multidimensional scaling [M]) or using one of the three kinship matrices described.

Project II consists primarily of Oregon winter and facultative six-rowed lines and winter six-rowed and two-rowed lines from the USDA-Agricultural Research Service (ARS) barley breeding program located at Aberdeen, ID, and a set of near isogenic lines for vernalization genes developed by Okayama University (Okayama, Japan). Coordinated agricultural project III consists primarily of winter and facultative six-rowed lines derived from intermating of germplasm represented in CAP I. Although each of the three germplasm sets consists of 96 genotypes, we used reduced sets from CAP I and CAP II. Due to the presence of 17 identical accessions in CAP I, 16 were removed and the final set used for analysis is n = 79. Removal of the near-isogenic lines and spring habit parental lines from CAP II left a final set of 71 accessions. All 96 accessions from CAP III were used.

Genotyping

All accessions were genotyped for 3072 SNPs using two Illumina GoldenGate (Illumina Inc., San Diego, CA) oligonucleotide pool assays developed for the barley CAP. The development and application of the barley oligonucleotide pool assays (BOPA1 and BOPA2) is described in detail in Close et al. (2009) and Szűcs et al. (2009). In this research, we used the consensus SNP map developed by Close et al. (2009), available by downloading the 1.77 version of the barley HarvEST database (Wenamaker and Close, 2011).

Phenotyping

The CAP germplasm sets were grown at two locations in Oregon, Corvallis and Pendleton, over a 3-yr period. Each set was grown in a different year. Because winter barley is planted in the fall of 1 yr and harvested in the summer of the next year, the seasons sampled were as follows: CAP I: 2005/2006 for some traits and 2006/2007 for others; CAP II: 2006/2007; and CAP III: 2008/2009. The CAP trials were grown in single replicate augmented designs following standard yield trial procedures (e.g., plots sizes, fertility, and weed management). Malting quality traits were measured at the USDA-ARS Cereal Crops Research Unit located in Madison, WI, following American Society of Brewing Chemists (ASBC, 1992) protocols (see Budde et al., 2008, for details of specific micromalting and malt quality methodologies).

Statistical Analyses

We compared 16 different models for association mapping in the CAP I population to identify the most appropriate model for all three populations (Table 1). The selected model, a QK mixed model structured association proposed by Yu et al. (2006) and modified by Kang et al. (2008), was used:

\[ Y = X\beta + Qv + Zu + e \]

where Y is the phenotypic vector, X is the molecular marker matrix, \( \beta \) is the unknown vector of allele effects to be estimated, Q is the posterior probabilities matrix of belonging to each population obtained from STRUCTURE (Pritchard et al., 2000), v is the vector of population effects (parameters), Z is a matrix that relates each measurement to the individual from which it was obtained (an identity matrix in our case), u is the vector of random background polygenic effects, and e is the residual error.

The population structure matrix (Q matrix) was obtained from STRUCTURE (Pritchard et al., 2000) with the linkage model (Falush et al., 2003) using 1083 SNPs for CAP I, 1339 SNPs for CAP II, and 1037 SNPs for CAP III selected after removing markers with minor allele frequencies, more than 10% missing data, and low marker quality (low GenTrain scores [Close et al., 2009]). The model with two populations was selected for all CAP data-sets based on the ad hoc statistic (\( \Delta K \)) proposed by Evanno et al. (2005). We estimated \( \Delta K \) based on 10 repeats of 5000 simulations after a 5000 burn-in for 1 to 10 populations. The estimated Q matrix was then obtained for two groups with 100,000 simulations after a 10,000 burn-in period. The kinship matrix (K matrix) was obtained with the EMMA approach proposed by Kang et al. (2008). Association mapping was conducted for each trait in each environment using TASSEL (Bradbury et al., 2007).

Since the data set contained some missing phenotypic values, further adjustment was conducted for minor allele frequencies and missing marker data for
specific phenotypes. As a result the average number of individuals and markers used varies somewhat. Significant markers at $p < 0.05$ were selected after a false discovery rate multitest adjustment (Benjamini and Hochberg, 1995). Significant marker-trait associations were aligned with (i) annotations for the unigenes on which the BOPA SNPs are based, (ii) known genes that affect malt and beer production, and (iii) previously mapped malting quality QTL summarized in Szűcs et al. (2009).

RESULTS

There was phenotypic variability for all the traits in all the environments of the OSU Barley CAP study (Fig. 2). Variability was still present in each succeeding germplasm set (i.e., CAP II and CAP III).

There was not a single best model for all the traits and environments (Fig. 3). Models with the SPAGeDi (Hardy and Vekemans, 2002) kinship matrix performed poorly (data not shown), and models with kinship matrix from TASSEL (Bradbury et al., 2007) performed similar to EMMA (Kang et al., 2008) models (data not shown). Based on $p$-value distribution in CAP I, naïve and fixed-effects models did not perform well in most of the situations, having a distribution skewed toward significance (Fig. 3). The PK.E model performed poorly at Pendleton and for ME at Corvallis. The MK.E model was the best model at Pendleton, while QK.E was among the best models in most of the situations (Supplemental Table S1). Based on the distribution of $p$ values, number of consistently significant markers and coincidence with previously reported QTL and known genes in CAP I, we selected the QK.E mixed model to conduct all the analyses in CAP I through CAP III.

A total of 21 marker-trait associations for ME were found on chromosomes 1H, 2H, 3H, 4H, 5H, and 7H (Table 2; Fig. 4). Most of them were detected in only one of the data sets, although four of the significant markers were detected in two of the three sets. The approximate centimorgan positions are as follows (the exact positions of all SNPs in the consensus map can be found in Supplemental Tables S2 and S3): 110 cM on chromosome 3H in CAP III (SNP 11_20009) and CAP II (SNP 11_20168), 130 cM on chromosome 3H in CAP III (SNP 12_20369) and CAP I (SNP 12_30062), 160 cM on chromosome 5H in CAP I (SNP 12_20369) and CAP III (SNP 11_10842), 160 cM on chromosome 5H in CAP I (SNP 12_30642), and 130 cM on chromosome 7H in CAP I (SNP 11_10861) and CAP III (SNP 11_10182). Most of the QTL for ME were detected in one or two of the environments (Supplemental Table S3). The highest number of
coincident significant associations in the two environments was in the CAP III population and the favorable allele state was the same in both environments (Supplemental Table S3). There were 17 marker-trait associations for BG located on all chromosomes (Table 2; Fig. 5). For CAP I, significant associations were found only at Pendleton and only at Corvallis for CAP II and CAP III. Two marker-BG trait associations were detected independently in two data sets: 110 cM in chromosome 5H in CAP I (SNP 11_20850) and CAP III (SNP 12_30056) and 150 cM in chromosome 7H in CAP II (SNP 12_30593) and CAP III (SNP 12_21328). Four marker-trait associations—on chromosomes 1H, 5H, and 7H—were found for AA (Table 2; Fig. 6). The only significant associations were detected in CAP II grown at Corvallis. Six marker-trait associations were detected for DP on chromosomes 1H, 4H, 6H, and 7H (Table 2; Fig. 7). All of the significant associations were detected in CAP III: in Corvallis data at 50 cM in chromosome 1H (SNP 11_20660) and at three other SNPs with lower GenTrain scores (Close et al., 2009) (Supplemental Table S3) and in Pendleton data at 64 cM (SNP 11_20432). We did not find any marker to be significantly associated with GPC in any population or environment. Although most of the markers with a significant trait association in CAP I had significantly different allelic frequencies in CAP III population, not all of them were fixed in CAP III (Table 3). Furthermore, not all the alleles fixed between CAP I and CAP III were favorable.

DISCUSSION

Choice of Statistical Model

With so many different model options that are dependent on the specific structure present in the data and on the specific trait being analyzed, there is no obvious way of selecting the best model. To evaluate the alternatives, we compared 16 different models using the CAP I data set for all traits in all environments. For the purpose of simplicity, we show results from 8 of the 16 models (Fig. 3; Supplemental Table S1). The naïve model produced a number of apparent false positive or spurious associations that have been attributed to population structure and genetic relatedness (Malosetti et al., 2007), with a distribution skewed toward significance. This is even more pronounced when small populations are used (Bernardo, 2004). The Q, P, and M fixed-effect models produced many nominally significant markers, presumably false positives based on their distribution. The kinship matrix estimated using SpaGeDi (K.S) and all the mixed models containing the K.S matrix produced somewhat different results than the models with kinship matrix estimated using TASSEL (Bradbury et al., 2007) (K.T) and K.E. The mixed models including Q, P, and M produced similar results, with some inconsistencies for significant trait associations with P and M and some nonsignificant associations for markers at previously reported QTL or known genes that the Q models were able...
to detect. The **PK.E** model performed poorly using Pendleton data and for ME at Corvallis. The **MK.E** model was the best model for Pendleton while **QK.E** was among the best models for most traits. Most of the QTL positions were significant in more than 13 out of the 16 models tested. Some markers were significant only in a few models, and those models were usually naïve, **Q, QK.T,** and **QK.S.** Based on the cumulative distribution of \( p \) values, the number of consistent significant markers, and previously reported QTL or known genes, we selected the **QK.E** mixed model.

Our approach to selecting the best model has some caveats. First, we did not perform a formal hypothesis test to select the best model. Some of the model comparison strategies based on likelihoods are inappropriate in this case because we are using restricted maximum likelihood estimates and the models have different fixed effects, which is the main point of the comparison (Verbeke and Molenberghs, 2000). Second, an assertion bias could be generated by selecting the best model based on consistency and then proposing it as a good model because it picks the markers that are more consistent. Additionally, a significant marker-trait association in several models does not mean that it is a true QTL; it could still be a false positive. However, most models will capture the few important QTL. The differences between them will be for the smaller-effect QTL. The model selected is the mainstream model used for MAS (Yu et al., 2006) and we used it to illustrate the results. Additionally, Bradbury et al. (2011) compared several models for a broader set of the Barley CAP and concluded that either **QK** or **PK.E** performs well. Furthermore, models that best fit the data are not necessary the best models; there could be challenge interpretation in terms of the known biochemistry and biology of the malting process.

We used a false discovery rate correction (Benjamini and Hochberg, 1995) at an \( \alpha \) level of 0.05 to correct for the increase in type I error rate with multiple comparisons. Although this procedure, like the Bonferroni test, is conservative because the tests are not all independent, the loss in power to detect true QTL is not as large as in the Bonferroni test (Benjamini et al., 2001). Additionally, the distribution of corrected \( p \) values in our study tends to be uniform (Fig. 3). Model diagnostics show appropriate residuals distribution (i.e., residuals follow normal distributions and homogenous variances).

**Coordinated Agricultural Project Malting Quality Quantitative Trait Loci and Prior Reports**

We detected several chromosome regions important for malting quality in barley on 1H (ME, DP, BG, and AA), 2H (ME and BG), 3H (ME), 5H (ME, BG, and AA), and 7H (ME, DP, AA, and BG). These regions were close to known genes that are important for malt and beer production and to previously mapped QTL for malt quality (see Szűcs et al. (2009) for a review; Fig. 4 through 7; Table 2; Supplemental Tables S2 and S3). We were able to detect QTL consistent with previous knowledge. Very few of the positions are new discoveries, but they do serve to validate the effectiveness of GWA for these traits and in this germplasm.

Han et al. (1997a) identified two genome regions as (i) the most consistent across populations and environments and (ii) good candidates for MAS for malt quality: on 4H at 37 to 59 cM and on 7H between 6 and 21cM. Szűcs et al. (2009) also reported a QTL for malt quality in this region of 4H, which is close to several known malt- and beer-related genes: an endoglucanase at 40 cM and a protein disulfide-isomerase at 48 cM. On 7H there is a putative \( \beta \)-glucosidase at position 25 cM (Hayes et al., 2003). We did not detect any QTL in either region, although we did identify markers associated with malt quality traits on 4H proximal to cM 37: 11_20145 at 1.6 cM for ME, 11_10738 at 19.5 cM for BG, and 11_21418 at 26.2 cM for DP. In chromosome 1H we found two significant marker-trait associations for ME that were 30 cM apart: 12_20187 at 101.4 cM and 11_10782 at 131.9 cM. Quantitative trait loci were previously reported in three different populations for ME and other malting quality traits on 1H between the two QTL that we detected (Szűcs et al., 2009; Supplemental Table S2). We detected two significant associations on chromosome 6H: 11_20552 at 42.4 cM for DP and 11_20889 at 75.2 cM for BG. There are few QTL for malting quality reported on chromosome 6H (Szűcs et al., 2009) and close to the positions we report (Supplemental Table S2).
Figure 3. (previous page) Cumulative distribution function (cdf) of \( p \) values in genome-wide scans for the CAP I barley array for all traits and environments. A close-up of the critical region is shown for each plot. The different curves correspond to different models compared: Naive, marker regression without correction for population structure; \( Q \), posterior probabilities matrix inferred from software STRUCTURE (Pritchard et al., 2000); \( P \), fixed-effects matrix from principal component analysis; \( M \), fixed-effects matrix from nonmetric multidimensional scaling; \( K \), mixed models using kinship matrix as implemented by efficient mixed model association [EMMA (Kang et al., 2008)]; \( PK \), mixed models with \( K \) matrix as fixed effects and kinship matrix as random effects; \( PK,E \), mixed models with \( P \) matrix as fixed effects and \( K \) matrix as random effects; and \( MK \), mixed models with \( M \) matrix as fixed effects and \( K \) matrix as random effects. ME, malt extract; CO, Corvalis location; BG, wort \( \beta \)-glucan; AA, \( \alpha \)-amylase; DP, diastatic power; GPC, grain protein content; PE, Pendleton location.

Table 2. Number and chromosome locations of single nucleotide polymorphisms (SNPs) showing significant [false discovery rate (FDR); \( p < 0.05 \)] associations with four malting quality traits measured in three germplasm sets (CAP I, CAP II, and CAP III), each phenotyped in two environments. The significant associations are classified as follows: SNP, number of significant marker-trait association; Reported gene, the SNP is within genes previously reported to be associated with malting and/or brewing quality or within a 10 cM window of such a gene; Reported quantitative trait loci (QTL), the SNP is within a 10 cM window of a previously reported QTL associated with malting and brewing quality. The complete data from which this table is derived is presented in Supplemental Table S2. Summary tables for each germplasm set and environment combination are presented in Supplemental Table S3.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>ME(^1) SNP</th>
<th>ME(^1) QTL</th>
<th>BG(^1) SNP</th>
<th>BG(^1) QTL</th>
<th>AA(^1) SNP</th>
<th>AA(^1) QTL</th>
<th>DP(^8) SNP</th>
<th>DP(^8) QTL</th>
<th>Total SNP</th>
<th>Total QTL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1H</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>2H</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>3H</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>4H</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5H</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>6H</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7H</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>8</td>
<td>17</td>
<td>17</td>
<td>7</td>
<td>13</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>48</td>
</tr>
</tbody>
</table>

\(^{1}\text{ME}, \text{malt extract.}\)
\(^{2}\text{BG, wort \( \beta \)-glucan.}\)
\(^{3}\text{AA, \( \alpha \)-amylase.}\)
\(^{4}\text{DP, diastatic power.}\)

Our results support the idea that malting quality is a complex phenotype composed of multiple component traits, each determined by many genes interacting with the environment. While QTL studies of biparental crosses have incorporated epistatic interaction (Jannink et al., 2001; Malosetti et al., 2004), the subject is not as well developed in association mapping. The small population sizes we had available compromise the approaches proposed by Jannink (2007), and those implemented by Cuesta-Marcos et al. (2010) and Wűrschum et al. (2011). Given the marker coverage and the location and number of the genes known to affect malt and beer production, one would expect that 32% of markers chosen at random would lie within a known QTL or near a known gene that affects malting and/or beer quality. Nonetheless, more than 50% of the significant associations we detected are coincident with or near a gene and/or QTL for malting and/or brewing quality (Supplemental Table S3). Additional evidence that at least some of the QTL we detected are not due to spurious associations is provided by the genetic dissection of the presumed donor of malting quality alleles in this germplasm array (Muñoz-Amatriain et al., 2010a).

The Source of Favorable Alleles for Malting Quality

Due to the complexity of malting quality traits and differences in the same parameters specified by different segments of the industry (Sparrow, 1971; Briggs, 1978), it is not always straightforward to classify alleles as favorable or unfavorable for malting quality. 88AB536 is a germplasm accession with good malting quality with favorable alleles tracing to the high quality spring malt variety Morex (Marquez-Cedillo et al., 2000). This accession was extensively characterized using structural and functional genomics tools by Muñoz-Amatriain et al. (2010a), and it is in the pedigree of most germplasm accessions favoring favorable alleles for AA trace to Morex via 88AB536 (Supplemental Table S3). However, 88AB536 was the source of both favorable and unfavorable alleles for AA, ME, BG, and DP (Supplemental Table S3). Clearly, there are other sources of high malting quality alleles besides 88AB536. Favorable alleles provided by unadapted germplasm, even when they were masked by unfavorable phenotypes, were found by other authors (Tanksley et al., 1982; de Vicente and Tanksley, 1993).
Germplasm Specificity of Quantitative Trait Loci

Most of the marker-trait associations detected in this study are germplasm specific. That is, they were significant in only one of the three germplasm arrays. When QTL were detected in more than one germplasm array, it was usually in the case of CAP I and CAP III. This was not entirely unexpected because most of CAP III lines are derived from CAP I lines, while the CAP II population contains somewhat different germplasm. In the following discussion, we are assuming that different significant markers at the same position or very close to the same position are responsible for the same QTL.

Since the development of CAP III lines was based on phenotypic selection of CAP I lines for malting quality and other traits, including the influx of new alleles contributed by Doyce (Fig. 1), we found several explanations for detecting different marker-trait associations in that population. First, if selection was acting on favorable alleles detected in CAP I, we would expect to have some of these to be approaching fixation in the CAP III...
population. Indeed, alleles at some markers were fixed (i.e., genotypic frequencies smaller than 15%) in the CAP III population (Table 2). However, it was surprising to find that some were fixed for “high malting quality” alleles while others were fixed for “low malting quality” alleles. This could be explained by selection (i.e., fixation of favorable alleles), random genetic drift due to small population sizes, extreme allele frequencies already present in the CAP I population (Table 2), and unmasking of new effects of close markers separable by recombination from the actual QTL or gene. Finally, Doyce-related lines were driving the association in some markers (Supplemental Table S3; Supplemental Fig. S1e; Supplemental Table S2).

Since the CAP I population contains subsets of lines not fully accounted for by population structure (the Q matrix, Fig. 1), we studied whether that structure was responsible for some of the marker-trait associations. We found several patterns: first, parallel responses between Oregon intermated lines (OIL) and non-Oregon intermated lines (NL) lines for some marker-trait associations.

Figure 5. Profile of significance levels for wort β glucan (BG) content with all the markers in the structured association mixed model in all populations for all the environments. QTL, quantitative trait loci.
environments indicating that population structure was not responsible for the association (Supplemental Fig. S1a) and second, associations driven by NL lines. In these cases, very little change in the average effect was detected for the different allelic states in the OIL while an important difference in mean effect was found for the different allelic states of the NL. It is, of course, possible that spurious associations are the cause of significance in such cases (Jannink et al., 2001). Third, we found genetic-background interaction where genotypic states in OIL produce “high” malting quality while the same genotypic state in NL produces “low” malting quality. Again, we cannot rule out spurious associations.

Although the response of Oregon CAP I-related lines and lines unrelated to Oregon CAP I germplasm (UC1) are not always parallel in the Barley CAP II population, the UC1 are not driving the associations for ME, BG, or DP (Supplemental Fig. S1b). The only markers for which UC1 could be driving the associations are 11_20432 at 1H-64.3 in DP at Corvallis, OR (COR), and 11_21418 at 4H-26.2 in

---

**Figure 6. Profile of significance levels for α-amylase (AA) content with all the markers in the structured association mixed model in all populations for all the environments. QTL, quantitative trait loci.**

---

Known genes that affect malt and beer production

— Previously reported QTL for alpha-amylase (AA)

○ CORVALLIS

■ PENDLETON
DP at Pendleton, OR (PEN). We found favorable genetic background interactions for sister lines with a sib in CAP I lines for ME and DP and from Idaho lines not used as parents of CAP I, Nebraska lines not used as parents of CAP I, and Oregon hooded lines that share one parent (Kold) with CAP I while the other parent (Hoody) is not a parent of CAP I lines for BG (Supplemental Fig. S1c). The Bu27 introgressions may have introduced some unfavorable alleles for BG, increasing the frequency of high BG content on 1H-61.5, 2H-53.5, and 5H-151.4 (Supplemental Fig. S1d).

Environment-Specific Quantitative Trait Loci

Most of the QTL are environment specific; that is, they are detected only in one environment. Even though the locations do not represent the same environments for all populations because the populations were phenotyped during different years, there is a consistent pattern. We detected more marker-trait associations at Corvallis than at Pendleton (Supplemental Table S3). In many cases, associations found at Pendleton were also found at Corvallis. This behavior has been found in other species as well (Li

Figure 7. Profile of significance levels for diastatic power (DP) content with all the markers in the structured association mixed model in all populations for all the environments. QTL, quantitative trait loci.
et al., 2011). Pendleton receives an average of approximately 750 mm precipitation per year whereas Corvallis receives approximately 1,200 mm precipitation per year. Lower rainfall, rotation, and crop residue management at Pendleton could lead to greater field heterogeneity and as a consequence lower heritability. However, Pendleton is not a location that would be considered a high drought-stress site, compared to other regions throughout the world where barley is grown. Although phenotyping is a crucial element of a QTL study, most of the efforts in QTL detection have not acknowledged the importance of the environment in finding marker-trait associations and rely on one or very few environments (Zhu et al., 2008). When multi-environment trials are used, methods are available for appropriate testing of marker effects (Malosetti et al., 2004; Piepho, 2005; Boer et al., 2007; Mohringer and Piepho, 2009). Comparatively speaking, phenotyping has remained statistically whereas progress in genotyping is dazzling. As an unfortunate consequence, most resources are usually devoted to genotyping and phenotyping is taken for granted.

The Consequences of Small Population Sizes

Small population sizes are usually not recommended in QTL studies because either they do not have sufficient power to find significant marker-trait associations or they produce many false positives. Small population sizes (i.e., <500 progenies) are not recommended for QTL studies in biparental lines because of reduced power for detecting QTL (Rebai and Goffinet, 1993; Melchinger et al., 1998), overestimation of QTL effects (Beavis, 1994), imprecise location of QTL (van Ooijen, 1992; Visscher et al., 1996), and increased false discovery rates (Lande and Thompson, 1990; Bernardo, 2004). In association mapping, population size is also important and the choice of population size depends on the relatedness of the individuals and the extent of linkage disequilibrium, the type of study (i.e., candidate-gene or genome-wide), and the number of markers (Zhu et al., 2008). In self-pollinated crop breeding, assembling a population of 500 homozygous accessions optimized for association mapping may be a challenge. Furthermore, in many cases increasing population sizes implies increasing diversity by including germplasm that is not representative of the breeding program (Wong and Bernardo, 2008). Accordingly, many recent association mapping studies conducted on plants have used small population sizes (i.e., smaller than 100 individuals) and Bradbury et al. (2011) demonstrated that these are sufficient to detect QTL using GWA. There is usually a trade-off between the number of genotypes and the quality of the phenotyping because extensive phenotyping is expensive and time consuming. Malting quality is a particularly expensive trait to measure phenotypically and increasing population size would limit the number of environments. The dilemma is that we have shown that it is important to be able to evaluate the genotypes in several environments. Therefore, population sizes should be carefully designed to balance power and phenotyping capacity. Using small germplasm arrays of approximately 100 accessions we were able to find marker-trait associations consistent with previously mapped genes and QTL that affect malt and beer production.

Implications for Marker-Assisted Selection

Although at first glance there seem to be too many candidate markers for a MAS program built on our findings, if we discard markers based on presumed spurious association or that are found only in specific germplasm and rely on those coincident with previous information, a set of 10 to 20 markers is generated that could be targeted.

Table 3. Evolution of allelic frequencies from CAP I to CAP III of markers with a significant marker-trait association in the CAP I array of barley germplasm; frequency of α-amylase alleles in CAP I and CAP III population with their standard error are provided, as well as the difference in frequency between them (Delta), p value for the test of delta not different from zero, the effect of the change and whether a marker close to that position was significant in CAP III or not.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Position</th>
<th>Marker</th>
<th>Trait</th>
<th>CAP I SE</th>
<th>CAP III SE</th>
<th>Delta SE</th>
<th>p value</th>
<th>Effect</th>
<th>Other markers §</th>
</tr>
</thead>
<tbody>
<tr>
<td>1H</td>
<td>101.4</td>
<td>12_20187</td>
<td>ME</td>
<td>0.429</td>
<td>0.055</td>
<td>0.123</td>
<td>0.044</td>
<td>0.305</td>
<td>0.0001</td>
</tr>
<tr>
<td>1H</td>
<td>131.9</td>
<td>11_10782</td>
<td>ME</td>
<td>0.691</td>
<td>0.051</td>
<td>0.951</td>
<td>0.029</td>
<td>−0.26</td>
<td>0.0003</td>
</tr>
<tr>
<td>2H</td>
<td>125.5</td>
<td>11_10446</td>
<td>ME</td>
<td>0.547</td>
<td>0.056</td>
<td>0.474</td>
<td>0.069</td>
<td>0.073</td>
<td>0.0577</td>
</tr>
<tr>
<td>3H</td>
<td>131.6</td>
<td>11_10842</td>
<td>BG</td>
<td>0.2</td>
<td>0.045</td>
<td>0.372</td>
<td>0.065</td>
<td>−0.172</td>
<td>0.0048</td>
</tr>
<tr>
<td>4H</td>
<td>1.6</td>
<td>11_20145</td>
<td>ME</td>
<td>0.357</td>
<td>0.054</td>
<td>0.367</td>
<td>0.064</td>
<td>−0.01</td>
<td>0.2831</td>
</tr>
<tr>
<td>5H</td>
<td>26.3</td>
<td>12_30167</td>
<td>ME</td>
<td>0.158</td>
<td>0.041</td>
<td>0.075</td>
<td>0.035</td>
<td>0.083</td>
<td>0.0414</td>
</tr>
<tr>
<td>5H</td>
<td>151.4</td>
<td>12_30062</td>
<td>ME</td>
<td>0.14</td>
<td>0.038</td>
<td>0</td>
<td>0</td>
<td>0.14</td>
<td>0.0096</td>
</tr>
<tr>
<td>7H</td>
<td>133.8</td>
<td>11_10861</td>
<td>ME</td>
<td>0.593</td>
<td>0.056</td>
<td>0.481</td>
<td>0.068</td>
<td>0.112</td>
<td>0.0227</td>
</tr>
<tr>
<td>2H</td>
<td>150.7</td>
<td>11_10791</td>
<td>BG</td>
<td>0.842</td>
<td>0.041</td>
<td>0.825</td>
<td>0.05</td>
<td>0.017</td>
<td>0.2233</td>
</tr>
<tr>
<td>5H</td>
<td>102.1</td>
<td>12_20850</td>
<td>ME</td>
<td>0.298</td>
<td>0.051</td>
<td>0.16</td>
<td>0.049</td>
<td>0.138</td>
<td>0.0104</td>
</tr>
<tr>
<td>6H</td>
<td>73.2</td>
<td>11_20889</td>
<td>BG</td>
<td>0.556</td>
<td>0.056</td>
<td>0.641</td>
<td>0.065</td>
<td>−0.085</td>
<td>0.0421</td>
</tr>
</tbody>
</table>

1ME: malt extract; BG: wort β-glucan.
2Effect: whether the change was in the direction of increasing malt quality (+), decreasing it (−), or nonsignificant (NS).
3Other markers: position and marker name if there was a significant marker in CAP III within 10 cM of the one identified in CAP I.
for MAS (Supplemental Table S3). Additionally, we have shown that some of the markers have already been fixed between the CAP I and CAP III populations. Alternatively, GS could be used instead of MAS to advance germplasm to higher levels of malting quality (Wong and Bernardo, 2008; Heffner et al., 2009). This approach could be advantageous given the small population sizes that will be necessary given the costs of phenotyping malting quality traits (Wong and Bernardo, 2008).

CONCLUSIONS

We were able to detect QTL for malting quality traits—malt extract, β glucan, diastic power, and α-amylase—in relatively small germplasm arrays of winter barley using a strict approach to the analysis. Most of the QTL are germplasm specific; we only detected them in one of the three populations. Furthermore, most of the QTL are environment specific; we only detected them in one of the environments studied. Finally, most of the QTL detected were close to previously mapped genes relevant for malt and beer production and to previously reported QTL for malting quality related traits. These results essentially validate prior findings in barley, a crop with a long history of qualitative and quantitative genetics research. The broader implication of these results is that GWA applied to relatively small populations may be useful for detecting QTL with large effects. This could be useful in genetic dissection of novel traits in well-characterized crops and in generating new knowledge in new and/or orphan crops.

Supplemental Information Available

Supplemental material associated with this manuscript is located at http://www.crops.org/publications/tpg.

Supplemental Table S1. Number of significant single nucleotide polymorphisms (SNPs) [false discovery rate (FDR)-correction, p < 0.05] detected with eight of the models for five malting quality traits measured in two environments (Corvallis, OR and Pendleton, OR) in the CAP I array of barley germplasm. The models compared are Naive, marker regression without correction for population structure; Q, posterior probabilities matrix inferred from software STRUCTURE (Pritchard et al., 2000); P, fixed-effects matrix from principal component analysis; M, fixed-effects matrix from nonmetric multidimensional scaling; K, mixed model using kinship matrix as implemented by efficient mixed model association (EMMA [Kang et al., 2008]); Q.K.E, mixed models with Q matrix as fixed effects and kinship matrix as random effects; P.K.E, mixed models with P matrix as fixed effects and K matrix as random effects; and MK, mixed models with M matrix as fixed effects and K matrix as random effects.

Supplemental Table S2. Single nucleotide polymorphism (SNP) marker list with chromosome position, reported genes, candidate genes, and previously reported quantitative trait loci (QTL) and whether or not the marker was significant in our study.

Supplemental Table S3. Summary statistics for the significant markers in malt extract, wort β glucan, α-amylase, and diastic power in Corvallis, OR (COR), and Pendleton, OR (PEN). The population in which the marker-trait association was found, chromosome, position, name and GenTrain score (Close et al., 2009) of the significant markers are provided as well as effect of allelic substitution, allelic state of potential donors of good malting quality (88AB536 and Morex), number of significant markers at the same location, and other possible causes of the association.

Acknowledgments

We would like to thank A. Corey (Oregon Barley Project), S. Petrie, and K. Rhinhart (Columbia Basin Agricultural Research Center) for assistance with field phenotyping. This research was supported by USDA-CSREES-NRI Grant No. 2006-55606-16722, “Barley Coordinated Agricultural Project: Leveraging Genomics, Genetics, and Breeding for Gene Discovery and Barley Improvement”, CSIC-UDELAR, and FONTAGRO-0617.

References


Cuesta-Marcos, A., P. Szucs, T.J. Close, T. Filichkin, G.J. Muehlbauer, K.P. Smith, and P.M. Hayes. 2010. Genome-wide SNPs and re-sequencing of growth habit and inflorescence genes in barley: Implications for...