2013

Genome-wide Association Study for Beta-glucan Concentration in Elite North American Oat

Franco G. Asoro
Iowa State University

Mark A. Newell
Iowa State University

M. Paul Scott
Iowa State University, pscott@iastate.edu

William D. Beavis
Iowa State Univeristy, wdbeavis@iastate.edu

Jean-Luc Jannink
United States Department of Agriculture

Published February 6, 2013

RESEARCH

Genome-wide Association Study for Beta-glucan Concentration in Elite North American Oat

Franco G. Asoro, Mark A. Newell, M. Paul Scott, William D. Beavis, and Jean-Luc Jannink*

ABSTRACT

Genome-wide association studies (GWAS) can be a useful approach to detect quantitative trait loci (QTL) controlling complex traits in crop plants. Oat (*Avena sativa* L.) β-glucan is a soluble dietary fiber and has been shown to have positive health benefits. We report a GWAS involving 446 elite oat breeding lines from North America genotyped with 1005 diversity arrays technology (DArT) markers and with phenotypic data from both historical and balanced 2-yr data. Association analyses accounting for pairwise relationships and population structure were conducted using single-marker tests and least absolute shrinkage and selection operator (LASSO). Single-marker tests yielded six and 15 significant markers for the historical and balanced data sets, respectively. The LASSO method selected 24 and 37 markers as the most important in explaining β-glucan concentration for the historical and balanced data sets, respectively. Comparisons of genetic location showed that 15 of the markers in our study were found on the same linkage groups as QTL identified in previous studies. Four of the markers colocalized to within 4 cM of three previously detected QTL, suggesting concordance between QTL detected in our study and previous studies. Two of the significant markers were also adjacent to a β-glucan candidate gene in the rice (*Oryza sativa* L.) genome. Our findings suggest that GWAS can be used for QTL detection for the purpose of gene discovery and for marker-assisted selection to improve β-glucan concentration in elite oat.

F.G. Asoro, M.A. Newell, and W.D. Beavis, Dep. of Agronomy, Iowa State Univ., Ames, IA 50011; M.P. Scott, USDA-ARS, Corn Insects and Crop Genetics Research Unit, Dep. of Agronomy, Ames, IA 50011; J.-L. Jannink, USDA-ARS, R.W. Holley Center for Agriculture and Health, Dep. of Plant Breeding and Genetics, Cornell Univ., Ithaca, NY 14853. Received 20 Jan. 2012. *Corresponding author (jeanluc.jannink@ars.usda.gov).

Abbreviations: AIC, Akaike Information Criterion; BLAST, basic local alignment search tool; BLUP, best linear unbiased prediction; CesA2, cellulose synthase A catalytic subunit 2; Csl, cellulose synthase–like; DArT, diversity arrays technology; FDR, false discovery rate; GWAS, genome-wide association studies; LASSO, least absolute shrinkage and selection operator; LD, linkage disequilibrium; OPN, oat performance nurseries; PC, principal component; PCA, principal component analysis; PK, population structure + kinship model; QTL, quantitative trait loci.

Crop improvement for increased nutritional value is an important objective for breeding programs. In oat (*Avena sativa* L.), breeding for mixed-linkage-(1,3;1,4)-β-D-glucan (referred to as β-glucan) concentration has been an objective for more than two decades in North America (Peterson, 1991). Beta-glucan is a soluble fiber component that is found in endosperm and in the aleurone layer of oat groats (Butt et al., 2008). Food agencies from Sweden, United Kingdom, Finland, and the Netherlands have approved the claim that β-glucan reduces blood cholesterol levels, whereas the U.S. Food and Drug Administration (2010) approved the claim that β-glucan decreases the risk of coronary heart disease (Tiwari and Cummins, 2009; FDA Health Claim 21CFR101.81). The reports on the positive health implications of oats when consumed as a whole grain are happening as plant breeding technologies are also rapidly evolving. Foremost are rapid and high-density genotyping technologies (e.g.,

Published in Crop Sci. 53:542–553 (2013).
doi: 10.2135/cropsci2012.01.0039

diversity arrays technology [DArT] markers; Tinker et al., 2009) and new statistical approaches to analyze the large amount of data that are being generated. The availability of high-density marker data enables high-resolution mapping of quantitative trait loci (QTL) controlling complex traits like β-glucan. Although traditional QTL mapping for β-glucan has been conducted in biparental oat populations (Kianian et al., 2000; De Koeyer et al., 2004), genome-wide association studies (GWAS) have yet to be implemented for QTL detection in elite oat germplasm.

Genome-wide association studies detect associations due to gametic phase disequilibrium between a marker allele and the causative QTL allele. Gametic phase disequilibrium, also known as linkage disequilibrium (LD), between loci is the nonindependence of alleles at two loci (Falconer and Mackay, 1996). The high-resolution mapping potential in GWAS relies on the availability of high-density genotyping technology and less LD in panels of unrelated lines than in biparental families. However, differential genetic relationships between individuals, in the form of different pedigree relationships or of subpopulation structure, can lead to false positives in GWAS. Thus, accounting for population structure and polygenic effects in association tests is important (Yu et al., 2006; Stich et al., 2008).

Single-marker tests for GWAS like the unified mixed-model approach (Yu et al., 2006; Zhao et al., 2007) have been successfully implemented with a suitable correction for multiple testing. Given that complex traits are controlled by multiple QTL in concert, an alternative strategy would be to include all markers in a regression model. However, since the number of markers is usually larger than number of observations, applying an ordinary multiple regression model for variable selection would be impossible due to an insufficient number of degrees of freedom. One solution is to use penalty parameters in a linear model by employing a method such as the least absolute shrinkage and selection operator (LASSO; Tibshirani, 1996). Briefly, a penalized regression minimizes a function with two components: (i) the squared deviation between the phenotype and its prediction, and (ii) a penalty that increases with the magnitude of the regression coefficients. The LASSO solves the L1-norm penalized regression meaning that the penalty is the sum of the absolute values of the regression coefficients weighted by a parameter lambda: the larger the value of lambda the more markers with zero effect will be in the model. Therefore, LASSO can both do shrinkage and marker selection. This makes LASSO an attractive approach for GWAS since markers with small effects are shrunk to zero, resulting in a sparse model whereas only markers with large effects are retained (Wu et al., 2009). One criticism of LASSO is that it may overshrink markers with large effects, resulting in reduced prediction accuracy. However, this should not be a problem when the objective is merely to identify the associations of marker variability with trait variability (Ayers and Cordell, 2010).

Newell et al. (2010) explored the genome-wide LD in a world collection of oat germplasm, and results suggested that GWAS in oats is feasible for QTL detection. Application of GWAS where elite germplasm is used as the association panel provides immediate inference for cultivar development programs (Breseghello and Sorrells, 2006). Consequently, markers that are identified can readily serve as a basis for selection in cultivar development (Bernardo, 2008). The use of breeding lines and cultivars also leads to the opportunity to use phenotypic data routinely collected for plant breeding purposes. However, the data from such programs are highly unbalanced given that a new set of lines is entered every year and only a few lines overlap between years. Although linear mixed models are robust to this kind of situation, it would be beneficial to determine if a balanced data set from limited environments would be useful for GWAS. Using oat cultivars and breeding lines from the United States and Canada as GWAS panel, our objectives were to:

1. Assess population structure of elite oat lines as it relates to β-glucan concentration.
2. Apply GWAS using single- and multiple-marker model tests for β-glucan concentration.
3. Compare significant associations to previous β-glucan QTL studies and to rice (*Oryza sativa* L.) candidate genes to develop a prioritized list for further study.

MATERIALS AND METHODS

Genotype Data, Kinship, and Population Structure

Oat lines studied were entered into the Uniform Oat Performance Nurseries and the Quaker Uniform Oat Nurseries from 1994 to 2007. In addition, 18 lines with phenotypic data from previous research conducted at Iowa State University (Colleoni-Sirghie et al., 2004; Chernyshova et al., 2007), two lines with phenotypic data from North Dakota State University, and seven cultivars with phenotypic data from the National Plant Germplasm System (http://www.ars-grin.gov/npgs/acc/acc_queries.html [accessed 20 July 2012]) were included for a total of 470 lines. Seed was obtained from breeding programs in the United States and Canada and planted in the greenhouse in January 2008. Plants were grown and leaf samples were collected from a single plant for each entry for DNA extraction according to recommended protocols (Diversity Arrays Technology, 2012). Then seeds harvested from the DNA plant were grown in the field as increase hills from April to July 2008 at the Agronomy Farm, near Ames, IA.

DNA samples of the 470 lines were submitted to Diversity Arrays Technology Pty Ltd (Yarralumla, ACT, Australia) for genotyping, of which 446 produced high-quality genotypic data. The DArT marker redundancies were removed as described in Asoro et al. (2011). The genotypic data were used to compute the kinship matrix (**K**), defined as the proportion of common alleles

shared by any oat line using the emma.kinship function in the emma (Kang et al., 2008) package implemented in the R software (R Development Core Team, 2011). A matrix estimating population structure, denoted as **P**, was calculated using principal components analysis (PCA) on the marker data and retaining the first five components through scree test (Cattell, 1966).

Analysis of Data from Uniform Performance Trials in North America

Phenotypic data for β-glucan from the Uniform Oat Performance Nurseries and the Quaker Uniform Oat Nurseries stored in the Graingenes 2.0 database (Carollo et al., 2005) was used for analysis. Data from the same lines with different listed names were merged under one entry name and confirmed through the Pedigree of Oat Lines database (Tinker and Deyl, 2005). In total, the data consisted of 450 lines (446 genotyped lines plus four long-term checks used for the phenotypic analysis) based on 2909 observations and 129 environment combinations of test years (1994–2007) and locations in the United States and Canada. The four long-term checks were lines used in oat performance nurseries (OPN) and were included to provide overlap across environments.

One common strategy to analyze highly unbalanced data sets is to employ a mixed-model approach and use the best linear unbiased prediction (BLUP) for each line as the response variable for GWAS (Zhang et al., 2009). However, BLUP values are shrunken toward the mean and the amount of shrinkage is dependent on the number of data points per individual. Because our data were highly unbalanced, differential shrinkage means that the trait would in effect be measured on a different scale for each observation, leading to reduced power and higher effect estimation error (Garrick et al., 2009). To avoid these shortcomings, we first fitted our data with the following mixed model:

$$y = \text{mean} + \text{environment} + \text{oat lines} + \text{error}$$

where y are the β-glucan observations (expressed in %), population *mean* and *environment* were considered fixed effects, and *oat lines* were considered random effects. The covariance matrix of *oat lines* was assumed proportional to the kinship matrix computed above. The mixed model was fitted using the kinship.BLUP function in rrBLUP package (Endelman, 2011). Raw phenotypes (y) were corrected for the fixed effects estimated from the model to derive the values for the observations corrected for environment. Finally, the sample mean of the corrected observations for each oat line was computed and used as the phenotypic value for GWAS (denoted $y^{\star}$). This value is referred to here as the OPN value. It measures β-glucan concentration without differential shrinkage despite large differences in replication across lines.

Analysis of Beta-glucan Data from Ames 2009 and 2010

Balanced data sets for the elite lines came from field experiments that were conducted at the Agronomy Farm, Iowa State University, from April to July 2009 and April to July 2010. For 2009 and 2010, each hill plot consisted of seed collected from the 2008 field season. The source for each line in the 2008 field season was from the original seed source that was genotyped in January 2008 in the greenhouse. A total of 475 oat lines consisting of the 470 lines mentioned above plus checks were planted in two replicates using an incomplete block design where each incomplete block was composed of 25 hills arranged in a 5 × 5 grid. Heads were manually harvested and threshed after 1 wk of drying in the field. Oats were dehulled using a Codema Laboratory dehuller (Codema LLC) and milled into flour in 15-ml polycarbonate vials containing two 9.5-mm ball bearings (OPS Diagnostics LLC, Lebanon, NJ) using a reciprocating shaker (Talboys HT Homogenizer, Troemner, Thorofare, NJ). Beta-glucan concentration (as percent on a dry-weight basis) was then measured using the streamlined mixed-linkage β-glucan enzymatic laboratory kit from Megazyme (Megazyme Inc., Wicklow, UK) that was improved for high-throughput analysis in a 96-well plate (M.A. Newell, H.J. Kim, A. Moran-Lauter, P.J. White, unpublished data, 2011).

To correct for fixed effects due to plate differences, the samples from a given incomplete block were analyzed on the same plate, thus confounding plate and incomplete block effects. The observations from 2 yr were combined using the fixed-effects model:

$$y = \text{mean} + \text{year} + \text{replication} + \text{incomplete block}\ (\text{replication} \times \text{year}) + \text{oat lines} + \text{error}$$

Statistical analysis was done using PROC MIXED in SAS Version 9.2 (SAS Institute, 2010) and least square means of *oat lines* were treated as the phenotypic values ($y^{\star}$) and referred to as the Ames values.

Population Structure + Kinship Model for Single-Marker Association Analysis

The mixed model for each marker in the association analysis (Yu et al., 2006) is as follows:

$$\mathbf{y}^{\star} = \mu + \text{marker} + \text{population structure} + \text{polygenic effect of oat line} + \text{error}$$

where $\mathbf{y}^{\star}$ is a vector of adjusted phenotypic data from either the OPN or the Ames data source, μ represents the population mean, *marker* is the fixed marker effect, *population structure* fixed effects are the first five PCA scores, *polygenic effect of oat line* is a random effect, and *error* is the random residual error. The variance of the polygenic effect is assumed to be equal to $\mathbf{K}V_A$, where $\mathbf{K}$ is the kinship matrix of oat lines and V_A is the additive variance due to polygenic effects. The mixed linear model for association analysis was implemented by modifying the GWA function within the rrBLUP package. The modification was done to include output for p-values, R^2, and marker effects. The false discovery rate (FDR; Benjamini and Hochberg, 1995) for multiple testing was applied to the p-values for marker effects from the PK model. We used a relaxed FDR of 0.33 to identify more markers that we subsequently filtered based on other criteria such as basic local alignment search tool (BLAST) homology and comparison to previous biparental mapping studies for β-glucan concentration (Kianian et al., 2000; De Koeyer et al., 2004). Lastly, LD (measured as r^2) among the significant markers was calculated to determine if the significant markers were likely capturing effects from the same causal locus.

Mixed-model LASSO

A mixed-model LASSO method proposed by Wang et al. (2010) for GWAS in plants was applied in this study. The R function "amltest" (Dong Wang, personal communication,

2011) was modified so that marker effects would not be weighted. The objective of the mixed model LASSO is to estimate the marker and population structure effects that minimize the following equation:

$$\left(\mathbf{y}^{\star}-\mathbf{X}^{\mathrm{T}}\beta\right)^{\mathrm{T}}\mathbf{V}^{-1}\left(\mathbf{y}^{\star}-\mathbf{X}^{\mathrm{T}}\beta\right)+\lambda\sum_{j=1}^{p}|\beta|$$

where $\mathbf{X}$ is the matrix containing all markers and first five principal component axes as predictors, β is a vector of predictor effects, $\mathbf{y}^{\star}$ is the β-glucan data (response variable), and lambda (λ) is the penalty parameter. The $\mathbf{V}^{-1}$ is defined as $\sigma_g^2 ZKZ^T+\sigma_e^2 I$, where $\mathbf{Z}$ is the design matrix for observations, $\mathbf{K}$ is the kinship of all lines described above, and σ_g^2 and σ_e^2 are the genetic variance of oat lines and residual variance, respectively. As a note, the ordinary LASSO does not contain the V^{-1} term (Tibshirani, 1996).

The mixed-model LASSO procedure proposed by Wang et al. (2010) was applied in our study using the following:

1. The algorithm started from an ordinary LASSO on $y^{\star}$ and the $\mathbf{X}$ matrix to reduce the number of variables. The first 50 predictors that entered into the LASSO solution path denoted as X_q were chosen.
2. One marker was added every iteration based on the order in the LASSO solution path, starting from a model with no markers up to the model with 50 markers, by doing the following:
   i. using the estimates of fixed effects from the previous iteration, variance components were calculated by maximum likelihood, the $\mathbf{V}^{-1}$ matrix was obtained, and the Akaike Information Criterion (AIC) value was calculated.
   ii. $y^{\star}$ is adjusted such that $\tilde{y}=V^{-1/2}y^{\star}$ and the $\mathbf{X}$ matrix is adjusted in a similar fashion: $\tilde{X}=V^{-1/2}X$, where $\tilde{y}\sim N\left(\tilde{X}\beta,I_n\right)$. The ordinary LASSO was then applied to $\tilde{y}$ and $\tilde{X}$ using the LARS algorithm of Efron et al. (2004).
3. Lastly, the AIC was used to determine the final model and the algorithm was rerun with only the final set of markers to determine the marker effects. As a consequence of LASSO, the entry order of markers becomes important since once the marker enters the model it usually remains in the model (Wu et al., 2009). Entry order can thus be used for ranking the markers (Sung et al., 2009).

Comparison to Previous QTL Studies and BLAST Homology Search

Markers associated in this study were compared to previous QTL studies conducted in biparental populations. The location of the significant markers from this study and previously identified markers linked to β-glucan QTL from Kianian et al. (2000), De Koeyer et al. (2004), and Groh et al. (2001) were compared based on the updated 'Kanota' × 'Ogle' map (Tinker et al., 2009).

To determine whether our study and previous studies picked up some of the same causal loci, we tested whether the positions of our associated markers were more likely than a random sample of positions to fall on the same linkage groups as QTL from previous studies. The updated map is 1989 cM long, whereas linkage groups on which β-glucan QTL have been identified total 828 cM. We therefore compared the fraction of our associated markers that were on linkage groups with previous β-glucan QTL with the binomial distribution with success probability 828/1989 = 0.416.

The nucleotide sequences of all oat DArT markers (Tinker et al., 2009) significant in this study plus the sequences of markers in perfect LD with those markers (Supplementary Table 1) were used in a BLASTn analysis against rice annotated sequences (Ouyang et al., 2007). The search was limited to an E-value cutoff of 1×10^{-15}. The location of all rice cellulose synthase and cellulose synthase–like genes were also determined. Then the location of the DArT marker sequences homologs and the rice candidate genes for β-glucan were compared.

To establish a threshold value for proximity, a point was chosen at random in the rice genome (~370,000 kb) and the distance in kilobytes between that point and the nearest rice candidate gene was determined. This process was conducted 1 million times to construct a distribution of distances under the null hypothesis that DArT marker homolog positions were random relative to rice candidate genes. The distance at the 5% quantile of this null distribution was 247 kb and was taken as the threshold value for adjacency to a rice candidate gene.

RESULTS

Beta-glucan Concentration Data and Population Structure in Elite Oat

Phenotypic values for β-glucan concentration used for the association analysis from two data sets (OPN and Ames) were significantly correlated (r = 0.71; Table 1). The two data sets had the same standard error of the mean of 0.03. The oat line variance was higher in Ames (0.45) than for the OPN data set (0.19), but the two had comparable residual variances (0.25–0.26). The broad-sense heritability was therefore higher in the Ames than in the OPN data source (0.63 and 0.43, respectively).

The clustering method proposed by M.A. Newell, D. Cook, and H. Hofmann (unpublished data, 2011) resulted in five clusters using the *k*-means method. The number of

Table 1. Summary statistics of beta-glucan concentration (%) for two data sets.

Descriptive data	Data set	
	OPN†	Ames
Mean	5.06	4.17
Minimum	3.15	2.32
Maximum	7.62	7.76
SE of mean	0.03	0.03
Phenotypic SD	0.56	0.69
Oat line variance‡	0.19	0.45
Residual variance‡	0.25	0.26
H^2	0.43	0.63
Correlation of OPN and Ames	0.71	

† Oat performance nurseries.

‡ Computed from original observed data where oat lines and residuals are the only random effects and both are assumed independently and identically distributed. The variances were significantly different from zero (p = <0.0001) based on Wald Z test.

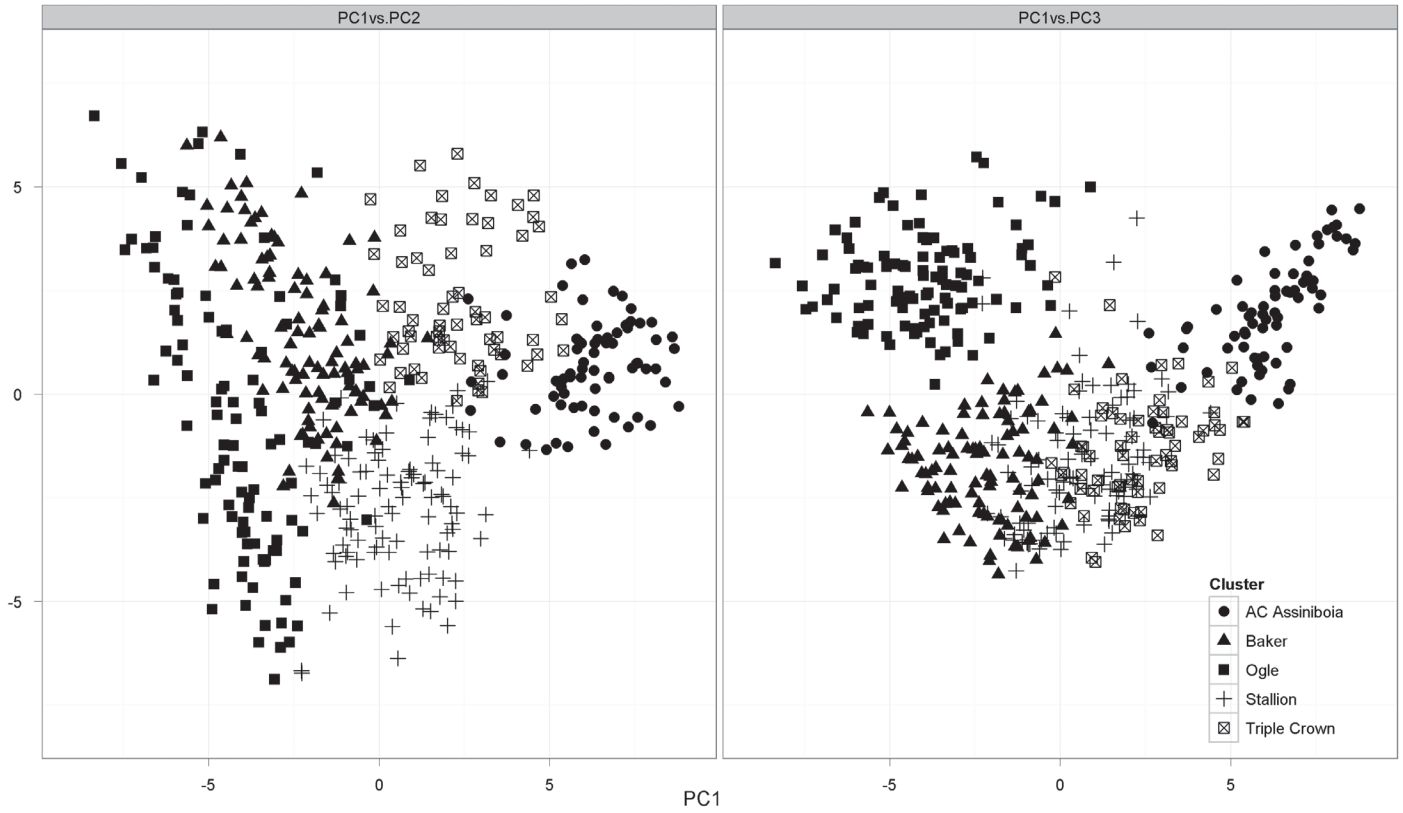

Figure 1. Scatterplot of principal components (PCs) on the marker data where lines are assigned to various *k*-means clusters labeled according to a popular cultivar in each cluster, PC1 vs. PC2 (left panel) and PC1 vs. PC3 (right panel).

lines for each cluster ranged from 66 oat lines in the 'Triple Crown' Cluster to 105 in the Ogle Cluster (Table 2). Principal component analysis showed that the first five principal components (PCs) explained 23% of the marker variation (data not shown). Visualization of the clusters in a scatterplot of PC1 vs. PC2 and PC1 vs. PC3 showed distinct separation of clusters with minimal overlap (Fig. 1).

Means for β-glucan concentration per cluster (Table 2) showed significant differences based on ANOVA ($p < 0.0001$). The AC Assiniboia Cluster had the lowest β-glucan (3.79 for OPN and 4.78 for Ames), whereas the Ogle Cluster had the highest (4.46 for OPN and 5.39 for Ames). The correlations of β-glucan concentration with PC1, PC2, and PC5 scores were significant at $p < 0.05$. For the OPN data set, the correlations with the PCs were −0.32, −0.16, and 0.25, respectively, for the significant PCs. For the Ames data set, the correlations were −0.33, −0.11, and 0.22, respectively, for these PCs.

Table 2. Beta-glucan (BG) summary by cluster. Each cluster was named according to an oat cultivar in it.

Cluster	No. of lines	Mean % BG† (SD)	
		OPN‡	Ames
Baker	101	4.29 (0.56)	5.09 (0.46)
Ogle	105	4.46 (0.77)	5.39 (0.66)
AC Assiniboia	70	3.79 (0.45)	4.78 (0.32)
Stallion	104	4.09 (0.73)	5.00 (0.59)
Triple Crown	66	4.07 (0.65)	4.88 (0.39)

† Means of clusters are significantly different from each other based on ANOVA ($p < 0.0001$).

‡ Oat performance nurseries.

Population Structure + Kinship Single Test Association

The comparison of $-\log_{10}(p\text{-value})$ from the two data sets showed that there are more significant markers in the Ames data set for any nominal *p*-value cutoff (Supplementary Fig. 1). An FDR of 0.33 resulted in six significant markers for the OPN data set and 15 for the Ames data set (Table 3). Out of 21 markers identified as significant from the two data sets, there were four common markers, oPt.12985, oPt.14067, oPt.16436, and oPt.18130. These four common markers have consistent direction of marker effects across data sets. The fraction of the phenotypic variance explained (R^2) by the significant markers ranged from 2.1 to 2.8% for the OPN data and from 1.7 to 3.2% for the Ames data (data not shown), whereas the absolute marker effects ranged from 0.30 to 0.39 for the OPN data set and 0.26 to 0.47 for the Ames data set (Table 3).

Pairwise LD (r^2) values >0.50 between significant markers were observed in the following marker pairs for the Ames data set: oPt.17611 and oPt.12985 with 0.89, oPt.11737 and oPt.14067 with 0.76, oPt.11737 and oPt.3063 with 0.73, oPt.14067 and oPt.3063 with 0.86, oPt.9329 and oPt.2635 with 0.86, and oPt.12704 and oPt.16436 with 0.65 (Supplementary Fig. 3). There were

no pairwise LD values >0.50 between significant markers for the OPN data set.

Mixed-model LASSO Association

For model selection in mixed-model LASSO, the lowest AIC corresponded to a model with the first 24 markers in the OPN data set and the first 37 markers in the Ames data set (Supplementary Fig. 2 and 4, Table 4). There were 13 markers in common between the OPN and Ames data sets using the mixed-model LASSO. The absolute effect of markers that were included in the model ranged from 0.003 to 0.18 for OPN and 0.004 to 0.17 for Ames. The marker with the largest and most consistent effect was oPt.18130 (0.18 and 0.17, respectively, for OPN and Ames). Furthermore, LD relationships among the markers identified using mixed-model LASSO indicated that only one pair of markers, oPt.3063 and oPt.14067, was in high LD ($r^2 = 0.86$), which occurred only for the Ames data set (Supplementary Fig. 5).

Markers across Models and Data Sets

All of the significant markers identified in the PK OPN were also significant in the mixed-model LASSO OPN. For Ames, 11 out of the 15 significant markers identified in the PK association were also significant in the mixed-model LASSO (Tables 3 and 4, Fig. 2). Only three markers were consistently detected across all data sets and models (oPt.14067, oPt.12985, and oPt.18130). It was also observed that there were 10, 16, and 3 markers that were unique to the mixed-model LASSO OPN, mixed-model LASSO Ames, and PK Ames, respectively. Altogether, the two models and data sets resulted in 51 unique markers that were further explored using various independent filters.

Table 3. Significant markers from single-marker test using the population structure + kinship model for oat performance nurseries (OPN) and Ames data sets at false discovery rate of 0.33. Underlined markers are common between the two data sets.

OPN			Ames		
Marker	*p* values	Marker effects	Marker	*p* values	Marker effects
oPt.18130	0.0004	−0.38	oPt.12985	0.0001	0.45
oPt.16436	0.0005	0.39	oPt.2635	0.0002	0.45
oPt.11819	0.0006	−0.39	oPt.14067	0.0006	0.47
oPt.12985	0.0014	0.30	oPt.3063	0.0009	0.44
oPt.14067	0.0017	0.34	oPt.18130	0.0018	−0.41
oPt.11728	0.002	−0.34	oPt.17611	0.0022	0.38
			oPt.2590	0.0024	0.29
			oPt.14317	0.0031	−0.37
			oPt.16436	0.0033	0.41
			oPt.9329	0.0034	−0.34
			oPt.12704	0.0037	−0.39
			oPt.6974	0.0039	0.40
			oPt.11737	0.0042	−0.34
			oPt.1505	0.0049	0.26
			oPt.16158	0.0049	−0.36

Comparison to Beta-glucan QTL Mapping Studies and BLASTn Homology Search

To compare to previous QTL mapping studies, 24 out of the 51 markers significantly associated with β-glucan in this study were present in the updated Kanota × Ogle map (Tinker et al., 2009; Wight et al., 2003). None of the remaining 27 markers were in high LD (cutoff of $r^2 = 0.75$) with any of the mapped markers (data not shown), so we did not seek to place them using LD. The 24 markers covered

Table 4. Selected markers from mixed-model LASSO based on Akaike Information Criterion, sorted based on their entry order in the model. Underlined markers are common between the two data sets.

OPN†			Ames		
Marker	Rank	Marker effects	Marker	Rank	Marker effects
oPt.11819	1	−0.165	oPt.18130	1	−0.172
oPt.18130	2	−0.175	oPt.14067	2	0.125
oPt.11728	3	−0.160	oPt.6926	3	0.122
oPt.8249	4	0.153	oPt.11728	4	−0.103
oPt.16436	5	0.140	oPt.17220	5	0.109
oPt.8758	6	0.071	oPt.11849	6	−0.097
oPt.14067	7	0.115	oPt.12985	7	0.129
oPt.11149	8	−0.106	oPt.14317	8	−0.102
oPt.0732	9	−0.078	oPt.6974	9	0.113
oPt.15994	10	0.053	oPt.11149	10	−0.116
oPt.17024	11	0.036	oPt.16158	11	−0.037
oPt.11699	12	0.056	oPt.2590	12	0.056
oPt.1661	13	−0.042	oPt.17024	13	0.051
oPt.8247	14	0.039	oPt.1505	14	0.061
oPt.10545	15	0.019	oPt.3063	15	0.025
oPt.9990	16	−0.020	oPt.15994	16	0.065
oPt.17018	17	0.015	oPt.7556	17	−0.047
oPt.5064	18	−0.018	oPt.2635	18	0.093
oPt.9120	19	−0.012	oPt.12704	19	−0.065
oPt.10823	20	0.011	oPt.4358	20	−0.025
oPt.0894	21	0.013	oPt.8751	21	−0.062
oPt.12985	22	0.008	oPt.11359	22	−0.037
oPt.7556	23	−0.004	oPt.14778	23	−0.032
oPt.17670	24	0.003	oPt.16618	24	0.022
			oPt.5671	25	0.019
			oPt.16444	26	0.022
			oPt.13088	27	0.019
			oPt.0077	28	−0.027
			oPt.0233	29	0.014
			oPt.11819	30	−0.014
			oPt.17018	31	0.012
			oPt.8247	32	0.011
			oPt.10545	33	0.006
			oPt.8249	34	0.008
			oPt.12279	35	−0.008
			oPt.7652	36	0.006
			oPt.9209	37	−0.004

† Oat performance nurseries.

15 linkage groups (Table 5). From the 24 markers, 15 were found on the same linkage groups as previously identified QTL (Table 5). The probability of 15 or higher from a binomial distribution with success probability of 0.416 and 24 trials is 0.03. We therefore rejected the null hypothesis that our associations were random relative to previous QTL identifications. Five of the 15 markers were located on linkage group 22_44_18. Three of those 15 markers mapped to within 1 cM of QTL found in Kianian et al. (2000).

A BLASTn homology search was conducted for all significant markers identified in this study. Thirteen out of 51 unique markers from all methods and data sets were found to have homology to a total of 34 rice genes. Six homologs were found in chromosome 1 of rice, four in chromosome 3, eight in chromosome 4, one each for chromosome 5, 6, 8, and 12, two on chromosome 7, and lastly 10 hits were found in chromosome 11. The search showed that none of the markers reported in this study had a direct homology to any cellulose synthase gene families. To further filter the hits, the location of all cellulose synthase were searched in the database and compared to the location of DArT marker homologs (Fig. 3). The comparison showed that closest distance was 63 kbp, which was between the homolog (LOC_ Os03 g59480) of oPt.12704 and cellulose synthase A catalytic subunit 2 (CesA2) (LOC_Os03 g59340). This is followed by a homolog of oPt.8758 (LOC_Os03 g58910) at 235 kb away from the same CesA2 gene. The distances of homologs of oPt.12704 and oPt.8758 to one of the candidate genes can be considered adjacent given that only 5% of random positions in the rice genome are within 247 kb of a candidate gene.

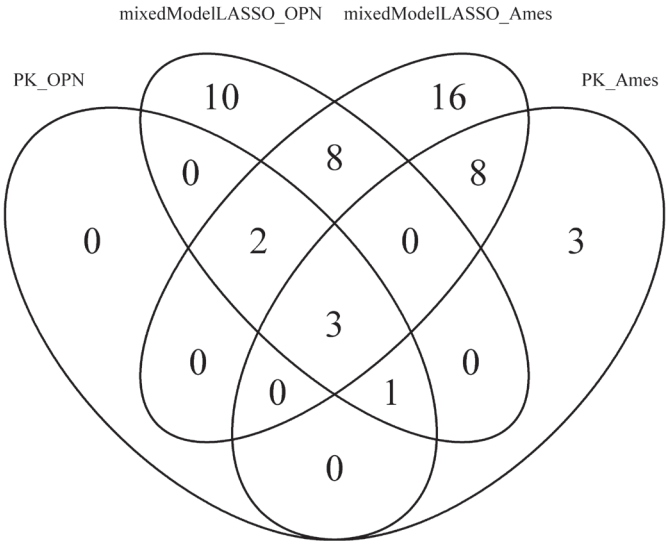

Figure 2. Venn diagram of markers identified in the population structure + kinship (PK) and mixed-model least absolute shrinkage and selection operator (LASSO) models for two data sets, oat performance nurseries (OPN) and Ames.

Table 5. Concordant genomic regions between beta-glucan important mapped markers in elite oat association study and markers from published biparental quantitative trait loci (QTL) mapping studies. The genetic position is based on framework markers in the updated 'Kanota' × 'Ogle' map (K × O; Tinker et al., 2009).

Marker	K × O linkage group	Position	Distance from previous β-glucan QTL	Reference
		cM		
oPt.12985	1_3_38_break	0.5	0.4 cM from cdo346A	Kianian et al., 2000
oPt.17611	1_3_38_break	1	0.1 cM from cdo346A	Kianian et al., 2000
oPt.5671	1_3_38_X3	25		
oPt.17024	4_12_13	54	21.7 cM from cdo549B	Kianian et al., 2000
oPt.11819	5_30	107.5		
oPt.9990	6	15.6	70.5 cM from cdo82	Kianian et al., 2000
oPt.10823	6	90	3.9 cM from cdo82	Kianian et al., 2000
oPt.6974	7_10_28	71.5	5.3 cM from acacac236	Groh et al., 2001
oPt.6926	15	27		
oPt.16444	15	3		
oPt.2635	16_23	42.5		
oPt.9329	16_23	42.5		
oPt.0732	17	23	15.5 cM from cdo1340	Kianian et al., 2000
oPt.4358	17	38.5	0 cM from cdo1340	
oPt.17220	21_46_31_40	61		
oPt.14317	22_44_18	105.6	11.6 cM from cdo484A	De Koeyer et al., 2004
oPt.12704	22_44_18	106.5	12.5 cM from cdo484A	De Koeyer et al., 2004
oPt.5064	22_44_18	148.5	54.5 cM from cdo484A	De Koeyer et al., 2004
oPt.16618	22_44_18	73.5	20.5 cM from cdo484A	De Koeyer et al., 2004
oPt.16436	22_44_18	114	20 cM from cdo484A	De Koeyer et al., 2004
oPt.8249	24_26_34	53.4	31.4 cM from b-glucanase	Yun et al., 1993
oPt.1661	32	30	25 cM from cdo395A	De Koeyer et al., 2004
oPt.0233	36	20		
oPt.15994	37	11.4		

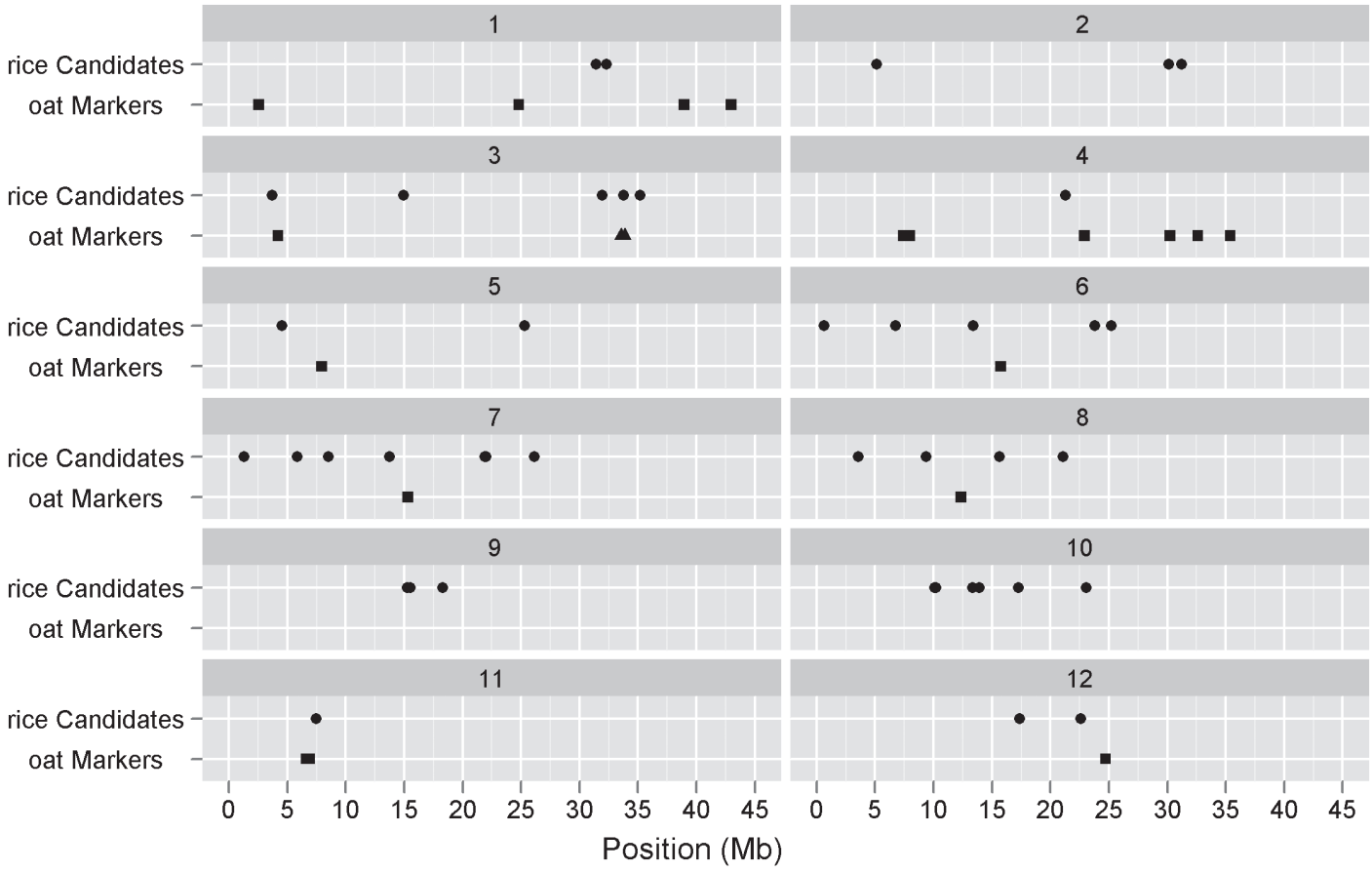

Figure 3. Comparison of locations of cellulose synthase gene families (circles) and locations of diversity arrays technology (DArT) marker homologs (squares and triangles) along the rice genome (*x* axis). The triangle in chromosome 3 indicates that homologs are significantly close to rice candidate genes. The panels are named according to rice chromosome number. The *x* axis is expressed as position in the rice genome in megabase pair (Mb).

DISCUSSION

Beta-glucan Concentration and Population Structure in the Elite Oat Association Panel

The inbred lines used in this study represent the wide spectrum of elite oat germplasm in North America. In particular, the association panel is composed of lines that were tested from 1994 to 2007 in Uniform Oat Performance Nurseries representing 15 breeding institutions in the United States and Canada. Analysis of phenotypic data from historical (OPN) and balanced (Ames) data sets showed that these two data sets are highly correlated but show heterogeneity of variances; therefore, data from OPN and Ames were analyzed separately. The standard deviations (0.56, 0.69 for OPN and Ames, respectively) for β-glucan in this study were comparable to values found by Peterson et al. (2005) of 0.44 to 0.59 for a group of elite cultivars. The level of heritability found in this study was similar to previous results (Holthaus et al., 1996).

The smaller broad-sense heritability in the OPN relative to the Ames data can be attributed to genotype × environment interaction. Variance components for this interaction were not estimable due to the OPN's unbalanced nature, so no direct comparison between the data sets could be made. It seems reasonable to assume, however, that Ames data came from more homogeneous environments. In this sense, data from the OPN represents *broad-adaptation β-glucan concentration* which has lower genetic variance, whereas Ames represents *narrow-adaptation β-glucan concentration*.

We note that seed obtained for DNA does not correspond exactly with seed historically submitted to the OPN for phenotypic evaluation. The latter seed would have had higher genetic heterogeneity, and some genetic drift may have occurred within the experimental line between OPN phenotyping and random sampling of a single plant for DNA extraction. We do not believe that this issue would bias our study in any systematic way other than contributing noise to the OPN analysis. It may, therefore, also have been a factor in reducing the number of significant associations between genotype and the OPN relative to the Ames data.

Association studies have long been known to be sensitive to population structure (Kennedy et al., 1992). To explore population structure, we used both cluster analysis and PCA on the marker data. The PC scatterplots classified with the clustering results indicated that the clusters

were distinctly separated with little overlap. For example, PC1 separated the AC Assiniboia Cluster from the Baker and the Ogle clusters, whereas PC3 separated the Ogle Cluster from the remaining clusters. It was also shown that the identified clusters differed for their β-glucan concentration means, indicating the potential of structure to generate false-positive associations in the GWAS. The correlation of β-glucan concentration with a subset of the PCs also implies that the population structure effects explain some of the variation in β-glucan and can account for confounding effects.

Kinship among oat lines was also used in all GWAS models in our study to account for fine-grained relationship (Yu et al., 2006). Accounting for kinship among lines, measured as the covariance among observations, has been known to reduce false positives and eliminate bias in marker effects by accounting for the genetic background effects (Kennedy et al., 1992). One general cause of population structure confounding is that single-factors models are used to identify associations for traits that are multigenic (Atwell et al., 2010). Explicitly multigenic models therefore make sense, and we evaluated their impact by contrasting a single-marker test (Yu et al., 2006) and mixed-model LASSO (Wang et al., 2010).

Single-marker Tests

The 13 unique markers from OPN and Ames data sets had relatively low R^2 values ranging between 2 and 3%. In a previous QTL mapping study, six putative QTL were identified in the Kanota × Ogle population (137 recombinant inbred lines) in which the five markers explained 2 to 5% and one marker explained 12% variation in phenotypic data (Kianian et al., 2000). In general, the low R^2 value and many significant markers indicate that β-glucan is controlled by multiple loci with small additive effects (Holthaus et al., 1996). The high R^2 in biparental QTL mapping studies than GWAS panel may be explained by the higher LD and the reduced overall genetic variability in the former than in the latter. On the other hand, the magnitude of individual marker effects in our study was comparable to Kianian et al. (2000). For example, the marker with the largest effect in this study (oPt.12985) can increase β-glucan concentration by 0.30 to 0.45%, similar to the 0.35% for a large-effect marker identified by Kianian et al. (2000).

Pairwise linkage disequilibrium, measured as the r^2 between markers, indicated that some of the significant markers from the PK model for the Ames data set were in high LD. For example, the high LD between oPt.12985 and oPt.17611 can be explained by the fact they are located on the same region of linkage group 1_3_38_break in the Kanota × Ogle population (Table 5). The high LD relationships among oPt.12704, oPt.16436, and oPt.14317, likewise, are explained by their close proximity on linkage group 22_14_18 (Tinker et al., 2009).

Mixed-model LASSO

In this study, we used mixed-model LASSO as an alternative approach for QTL detection. In cases where correlation between predictors is present, the algorithm selects the best marker within a group of correlated markers and sets the effect of other predictors to zero (Ayers and Cordell, 2010). We decided to explore this method because theory indicates that traits are controlled by multiple factors acting in concert. Instead of choosing a particular lambda for LASSO, we included an initial number of variables in the model and applied a goodness-of-fit test, the AIC, to decide the optimal number of markers in the model. Based on the AIC, the best model included the first 24 and 37 markers that entered into the model for the OPN and Ames data sets, respectively.

The distribution of absolute marker effects comprised many markers with zero effect, markers with near zero, and few markers with large effects (Fig. 3). The range of nonzero marker effects (0.003–0.18) suggested that the magnitude of effects can be used to determine markers that are likely associated with the trait. The low pairwise LD among the significant markers identified by LASSO confirms previous conclusions that LASSO results identify markers that are more independent (Sung et al., 2009).

Markers from Different Models and Data Sets

The results in this study provide examples of advantages and disadvantages for single-marker PK and mixed LASSO analyses. The single-marker PK test is a popular method with many studies, confirming its application in gene discovery and marker-assisted selection programs. However, this method will generate multiple hits for a single QTL when markers included are in high LD. The simplicity of application for the single-marker test makes it a good initial method to explore associated markers. A major difference between the PK and mixed LASSO analyses is that the objective of the former is to perform hypothesis tests on every marker, whereas that of the latter is to identify the best subset of markers in a model selection process. The two analyses are therefore complementary.

We also found that marker effects were heavily shrunk in mixed-model LASSO compared to individual effects in the PK model. First, this is explained by the fact the LASSO method shrinks effects regardless of the dimension of the data (Tibshirani, 1996). Second, the fact that the LASSO model had more markers than the PK model may mean that in the PK model each marker may be capturing more than one QTL, whereas markers in the LASSO model captured unique QTL (Ayers and Cordell, 2010). A positive outcome of such algorithm is that the markers with small effects can still be detected to be important to model β-glucan concentration. As a general recommendation, we propose to use both PK single-marker test and mixed-model LASSO to identify markers in genome-wide studies.

Comparison to QTL Studies

The most comprehensive genetic map in oat, Kanota × Ogle (Tinker et al., 2009), included only 24 out of the 51 markers identified in this study. These 24 markers were scattered across 15 of total 31 linkage groups of the Kanota × Ogle map, thus supporting the multigenic nature of β-glucan concentration in oat (Kianian et al., 2000; Orr and Molnar, 2008). The genetic location of 15 out of those 24 DArT markers corresponded to the same linkage groups of markers for β-glucan QTL identified by Kianian et al. (2000), De Koeyer et al. (2004), and Groh et al. (2001), a significantly greater number than expected by chance, indicating that our study detected some of the same signal as in biparental populations.

The genomic regions of four DArT markers in this study corresponded to the same regions of three QTL (cdo346A, cdo82, cdo1340) identified by Kianian et al. (2000). Two of the markers identified here (oPt.12985 and oPt.17611) colocalized within <1 cM of cdo346A—the marker with the largest effect QTL in the Kanota × Ogle population. This implies that oPt.12985 and oPt.17611 might be detecting the same QTL given that these markers are also in high LD. Another associated marker, oPt.10823, mapped within 4 cM of a previously identified QTL (cdo82).

Five associated DArT markers (oPt.14317, oPt.12704, oPt.5064, oPt.16618, and oPt.16436) are close to a QTL from the Terra × Marion population De Koeyer et al. (2004). Three of these markers (oPt.14317, oPt.12704, and oPt.16436) had high LD with each other and mapped within 10 to 20 cM of cdo484A, the marker explaining the most variance in Terra × Marion (De Koeyer et al., 2004).

The rest of the markers in the study were >20 cM distant from previously detected QTL. At 20 cM, the expected LD in elite oat decays already to less than $r^2 = 0.05$, indicating that these markers will probably not be able to capture sufficient variance of β-glucan QTL to be identified (Newell et al., 2010). Therefore, those markers may be detecting separate QTL.

Rice BLAST Homologies

The *CslF* and *CslH* gene families have been previously shown to affect β-glucan synthesis in various species within the grass family (Burton et al., 2006; Doblin et al., 2009). Given the shared evolutionary history of species within the grasses, it is possible to identify the same gene families through comparative genomic methods. In this study, none of the significant markers were directly homologous to *CslF* or *CslH* gene families in rice. However, these gene families are not the only participants in β-glucan synthesis given that they interact with the whole carbohydrate synthesis network (Fincher, 2009). Therefore, we cannot rule out the possibility that the markers reported in this study could lead to QTL controlling components of that metabolic network. For example, two of the significant markers (oPt.12704 and oPt.8758) in our study are adjacent to cellulose synthase A catalytic subunit 2 (*CesA2*), a gene that was identified to be coexpressed with *CslF6* in transcriptional studies for barley (*Hordeum vulgare* L.) (Burton and Fincher, 2009).

Implications for Marker-assisted Selection

There is still a high discrepancy between QTL studies and application of these studies in marker-assisted selection (Xu and Crouch, 2008). Attempts to breed for high β-glucan concentration using marker-assisted selection has been initiated based on early QTL mapping studies. Orr and Molnar (2008) developed markers for β-glucan based on QTL identified in the Kanota × Ogle population (Kianian et al., 2000) and in the Terra × Marion population (De Koeyer et al., 2004). Because these populations were developed from parents chosen to be highly distinctive for their phenotype, it is possible that these QTL will be population specific and therefore less useful in the context of breeding programs (Bernardo, 2008). Since we used elite oat, the QTL that we found have higher probability of being valid across elite populations. Finally, we note that our results confirm that β-glucan concentration is a polygenic trait (Holthaus et al., 1996; Chernyshova et al., 2007). For such traits, genomic selection, a method that predicts breeding values using all markers (Meuwissen et al., 2001), may be employed in lieu of traditional marker-assisted selection programs to increase β-glucan in oat (Asoro et al., 2011).

Our study can serve as an additional resource in understanding genetic mechanisms for this trait and enhancing the marker-assisted selection efforts toward the development of new oat cultivars with increased β-glucan concentration. The FDR cutoff of 0.33 and the LASSO model both use relaxed marker detection thresholds. However, we implemented further independent filters and we can prioritize the important QTL using these criteria. For the consistency of significance across methods and data sets, the important markers were oPt.14067, oPt.12985, and oPt.18130. For close proximity to previous QTL, the important markers were oPt.12985, oPt.17611, oPt.10823, oPt.4358, oPt.6974, oPt.14317, and oPt.12704. Finally, oPt.12704 and oPt.8758 were adjacent to β-glucan candidate genes. Since oPt.12985 and oPt.12704 were important for two criteria, they rise to the top of the list as candidates for further research.

Acknowledgments

This research was funded by the United States Department of Agriculture, National Institute of Food and Agriculture, Grant 2008-55301-18746. We thank Adrienne Moran Lauter for laboratory work and George Patrick for field work.

References

Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott, and J.-L. Jannink. 2011. Accuracy and training population design for genomic selection in elite North American oats. Plant Genome 4:132–144. doi:10.3835/plantgenome2011.02.0007

Atwell, S., Y.S. Huang, B.J. Vilhjalmsson, G. Willems, M. Horton, Y. Li, D. Meng, A. Platt, A.M. Tarone, T.T. Hu, R. Jiang, N.W. Muliyati, 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

Ayers, K.L., and H.J. Cordell. 2010. SNP selection in genome-wide and candidate gene studies via penalized logistic regression. Genet. Epidemiol. 34:879–891. doi:10.1002/gepi.20543

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.

Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Sci. 48:1649–1654. doi:10.2135/cropsci2008.03.0131

Breseghello, F., and M.E. Sorrells. 2006. Association analysis as a strategy for improvement of quantitative traits in plants. Crop Sci. 46:1323–1330. doi:10.2135/cropsci2005.09-0305

Burton, R.A., and G.B. Fincher. 2009. (1,3;1,4)-β-D-Glucans in cell walls of the Poaceae, lower plants, and fungi: A tale of two linkages. Mol. Plant 2:873–882. doi:10.1093/mp/ssp063

Burton, R.A., S.M. Wilson, M. Hrmova, A.J. Harvey, N.J. Shirley, A. Medhurst, B.A. Stone, E.J. Newbigin, A. Bacic, and G.B. Fincher. 2006. Cellulose synthase-like *CslF* genes mediate the synthesis of cell wall (1,3;1,4)-β- D-glucans. Science 311:1940–1942. doi:10.1126/science.1122975

Butt, M.S., M. Tahir-Nadeem, M.K.I. Khan, R. Shabir, and M.S. Butt. 2008. Oat: Unique among the cereals. Eur. J. Nutr. 47:68–79. doi:10.1007/s00394-008-0698-7

Carollo, V., D.E. Matthews, G.R. Lazo, T.K. Blake, D.D. Hummel, N. Lui, D.L. Hane, and O.D. Anderson. 2005. GrainGenes 2.0. An improved resource for the small-grains community. Plant Physiol. 139:643–651. doi:10.1104/pp.105.064485

Cattell, R.B. 1966. The scree test for the number of factors. Multivar. Behav. Res. 1:245–276.

Chernyshova, A.A., P.J. White, M.P. Scott, and J.-L. Jannink. 2007. Selection for nutritional function and agronomic performance in oat. Crop Sci. 47:2330–2339. doi:10.2135/cropsci2006.12.0759

Colleoni-Sirghie, M., J.-L. Jannink, and P.J. White. 2004. Pasting and thermal properties of flours from oat lines with high and typical amounts of beta-glucan. Cereal Chem. 81:686–692. doi:10.1094/CCHEM.2004.81.6.686

De Koeyer, D.L., N.A. Tinker, C.P. Wight, J. Deyl, V.D. Burrows, L.S. O'Donoughue, A. Lybaert, S.J. Molnar, K.C. Armstrong, G. Fedak, D.M. Wesenberg, B.G. Rossnagel, and A.R. McElroy. 2004. A molecular linkage map with associated QTLs from a hulless × covered spring oat population. Theor. Appl. Genet. 108:1285–1298. doi:10.1007/s00122-003-1556-x

Diversity Arrays Technology. 2012. Plant DNA extraction protocol for DArT. Diversity Arrays Technology, Pty Ltd., Yarralumla, Australia. http://www.diversityarrays.com/faq.html (accessed 1 Oct. 2012).

Doblin, M.S., F. Pettolino, S.M. Wilson, R. Campbell, R.A. Burton, G.B. Fincher, E. Newbigin, and A. Bacic. 2009. A barley *cellulose synthase-like CSLH* gene mediates (1,3;1,4)-β-D-glucan synthesis in transgenic *Arabidopsis*. Proc. Natl. Acad. Sci. USA 106:5996–6001. doi:10.1073/pnas.0902019106

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani. 2004. Least angle regression. Ann. Stat. 32:407–499.

Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250–255. doi:10.3835/plantgenome2011.08.0024

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman Technical and Scientific, Essex, UK.

Fincher, G.B. 2009. Exploring the evolution of (1,3;1,4)-beta-D-glucans in plant cell walls: Comparative genomics can help! Curr. Opin. Plant Biol. 12:140–147. doi:10.1016/j.pbi.2009.01.002

Garrick, D.J., J.F. Taylor, and R.L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41:55. doi:10.1186/1297-9686-41-55

Groh, S., A. Zacharias, S.F. Kianian, G.A. Penner, J. Chong, H.W. Rines, and R.L. Phillips. 2001. Comparative AFLP mapping in two hexaploid oat populations. Theor. Appl. Genet. 102:876–884. doi:10.1007/s001220000468

Holthaus, J.F., J.B. Holland, P.J. White, and K.J. Frey. 1996. Inheritance of beta-glucan content of oat grain. Crop Sci. 36:567–572. doi:10.2135/cropsci1996.0011183X003600030006x

Kang, H.M., N.A. Zaitlen, C.M. Wade, A. Kirby, D. Heckerman, M.J. Daly, and E. Eskin. 2008. Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723. doi:10.1534/genetics.107.080101

Kennedy, B.W., M. Quinton, and J.A.M. Vanarendonk. 1992. Estimation of effects of single genes on quantitative traits. J. Anim. Sci. 70:2000–2012.

Kianian, S.F., R.L. Phillips, H.W. Rines, R.G. Fulcher, F.H. Webster, and D.D. Stuthman. 2000. Quantitative trait loci influencing β-glucan content in oat (Avena sativa, 2n=6x=42). Theor. Appl. Genet. 101:1039–1048. doi:10.1007/s001220051578

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Newell, M.A., D. Cook, N.A. Tinker, and J.-L. Jannink. 2010. Population structure and linkage disequilibrium in oat (*Avena sativa* L.): Implications for genome-wide association studies. Theor. Appl. Genet. 122:623–632. doi:10.1007/s00122-010-1474-7

Orr, W., and S.J. Molnar. 2008. Development of PCR-based SCAR and CAPS markers linked to β-glucan and protein content QTL regions in oat. Genome 51:421–425. doi:10.1139/G08-026

Ouyang, S., W. Zhu, J. Hamilton, H. Lin, M. Campbell, K. Childs, F. Thibaud-Nissen, R.L. Malek, Y. Lee, L. Zheng, J. Orvis, B. Haas, J. Wortman, and C.R. Buell. 2007. The TIGR Rice Genome Annotation Resource: Improvements and new features. Nucleic Acids Res. 35:D883–D887. doi:10.1093/nar/gkl976

Peterson, D.M. 1991. Genotype and environment effects on oat beta-glucan concentration. Crop Sci. 31:1517–1520. doi:10.2135/cropsci1991.0011183X003100060025x

Peterson, D.M., D.M. Wesenberg, D.E. Burrup, and C. Erickson. 2005. Relationships among agronomic traits and grain composition in oat genotypes grown in different environments. Crop Sci. 45:1249–1255. doi:10.2135/cropsci2004.0063

R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Stat. Comput., Vienna. http://www.r-project.org (accessed 18 Jan. 2012).

SAS Institute. 2010. SAS/STAT® 9.2 user's guide. SAS Inst., Cary, NC.

Stich, B., J. Mohring, H.-P. Piepho, M. Heckenberger, E.S. Buckler, and A.E. Melchinger. 2008. Comparison of mixed-model approaches for association mapping. Genetics 178:1745–1754. doi:10.1534/genetics.107.079707

Sung, Y.J., T.K. Rice, G. Shi, C.C. Gu, and D.C. Rao. 2009. Comparison between single-marker analysis using Merlin and multi-marker analysis using LASSO for Framingham simulated data. BMC Proc. 3(Suppl. 7):S27. doi:10.1186/1753-6561-3-s7-s27

Tibshirani, R. 1996. Regression shrinkage via the lasso. J.R. Stat. Soc. 58:267–288.

Tinker, N.A., and J.K. Deyl. 2005. A curated internet database of oat pedigrees. Crop Sci. 45:2269–2272. doi:10.2135/cropsci2004.0687

Tinker, N.A., A. Kilian, and C.P. Wight, K. Heller-Uszynska, P. Wenzl, H.W. Rines, A. Bjornstad, C.J. Howarth, J.L. Jannink, J.M. Anderson, B.G. Rossnagel, D.D. Stuthman, M.E. Sorrells, E.W. Jackson, S. Tuvesson, F.L. Kolb, O. Olsson, L.C. Federizzi, M.L. Carson, H.W. Ohm, S.J. Molnar, G.J. Scoles, P.E. Eckstein, J.M. Bonman, A. Ceplitis, and T. Langdon. 2009. New DArT markers for oat provide enhanced mapcoverage and global germplasm characterization. BMC Genomics 10:39. doi:10.1186/1471-2164-10-39

Tiwari, U., and E. Cummins. 2009. Factors influencing β-glucan levels and molecular weight in cereal-based products. Cereal Chem. 86:290–301. doi:10.1094/CCHEM-86-3-0290

U.S. Food and Drug Administration. 2010. Health claims: Soluble fiber from certain foods and risk of coronary heart disease. U.S. Food and Drug Administration, Silver Spring, MD. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=101.81 (accessed 19 May 2011).

Wang, D., K.M. Eskridge, and J. Crossa. 2010. Identifying QTLs and epistasis in structured plant populations using adaptive mixed LASSO. J. Agric. Biol. Environ. Stat. 16:1–15.

Wight, C.P., N.A. Tinker, S.F. Kianian, M.E. Sorrells, L.S. O'Donoughue, D. Hoffman, S. Groh, G.J. Scoles, C.D. Li, F.H. Webster, R.L. Philips, H.W. Rines, S.M. Livingston, K.C. Armstrong, G. Fedak, and S.J. Molnar. 2003. A molecular marker map in 'Kanota' × 'Ogle' hexaploid oat (*Avena* spp.) enhanced by additional markers and a robust framework. Genome 46:28–47. doi:10.1139/g02-099

Wu, T.T., Y.F. Chen, T. Hastie, E. Sobel, and K. Lange. 2009. Genome-wide association analysis by lasso penalized logistic regression. Bioinformatics 25:714–721. doi:10.1093/bioinformatics/btp041

Xu, Y., and J.H. Crouch. 2008. Marker-assisted selection in plant breeding: From publications to practice. Crop Sci. 48:391–407.

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 accounting for multiple levels of relatedness. Nat. Genet. 38:203–208. doi:10.1038/ng1702

Yun, S.J., D.J. Martin, B.G. Gengenbach, H.W. Rines, and D.A. Somers. 1993. Sequence of a (1–3,1–4) beta-glucanase cDNA from oat. Plant Physiol. 103:295–296. doi:10.1104/pp.103.1.295

Zhang, Z., E.S. Buckler, T.M. Casstevens, and P.J. Bradbury. 2009. Software engineering the mixed model for genome-wide association studies on large samples. Brief. Bioinform. 10:664–675. doi:10.1093/bib/bbp050

Zhao, K., M.J. Aranzana, S. Kim, C. Lister, C. Shindo, C. Tang, C. Toomajian, H. Zheng, C. Dean, P. Marjoram, and M. Nordborg. 2007. An *Arabidopsis* example of association mapping in structured samples. PLoS Genet. 3:e4. doi:10.1371/journal.pgen.0030004