Aberystwyth University

Population Genomics Related to Adaptation in Elite Oat Germplasm

Esvelt Klos, Kathy; Huang, Yung-fen; Bekele, Wubishet A.; Obert, Don E.; Babiker, Ebrahiem; Beattie, Aaron D.; Bjørnstad, Åsmund; Bonman, J. Michael; Carson, Martin L.; Chao, Shiaoman; Gnanesh, Belaghihalli N.; Griffiths, Irene; Harrison, Stephen A.; Howarth, Catherine J.; Hu, Gongshe; Ibrahim, Amir; Islamovic, Emir; Jackson, Eric W.; Jannink, Jean-luc; Kolb, Frederic L.

Population Genomics Related to Adaptation in Elite Oat Germplasm

Kathy Esvelt Klos,* Yung-Fen Huang, Wubishet A. Bekele, Don E. Obert, Ebrahiem Babiker, Aaron D. Beattie, Åsmund Bjørnstad, J. Michael Bonman, Martin L. Carson, Shiaoman Chao, Belaghihalli N. Gnanesh, Irene Griffiths, Stephen A. Harrison, Catherine J. Howarth, Gongshe Hu, Amir Ibrahim, Emir Islamovic, Eric W. Jackson, Jean-Luc Jannink, Frederic L. Kolb, Michael S. McMullen, Jennifer Mitchell Fetch, J. Paul Murphy, Herbert W. Ohm, Howard W. Rines, Brian G. Rossnagel, Jessica A. Schlueter, Mark E. Sorrells, Charlene P. Wight, Weikai Yan, and Nicholas A. Tinker

Abstract

Six hundred thirty five oat (*Avena sativa* L.) lines and 4561 single-nucleotide polymorphism (SNP) loci were used to evaluate population structure, linkage disequilibrium (LD), and genotype–phenotype association with heading date. The first five principal components (PCs) accounted for 25.3% of genetic variation. Neither the eigenvalues of the first 25 PCs nor the cross-validation errors from $K = 1$ to 20 model-based analyses suggested a structured population. However, the PC and $K = 2$ model-based analyses supported clustering of lines on spring oat vs. southern United States origin, accounting for 16% of genetic variation ($p < 0.0001$). Single-locus F-statistic (F_{ST}) in the highest 1% of the distribution suggested linkage groups that may be differentiated between the two population subgroups. Population structure and kinship-corrected LD of $r^2 = 0.10$ was observed at an average pairwise distance of 0.44 cM (0.71 and 2.64 cM within spring and southern oat, respectively). On most linkage groups LD decay was slower within southern lines than within the spring lines. A notable exception was found on linkage group Mrg28, where LD decay was substantially slower in the spring subpopulation. It is speculated that this may be caused by a heterogeneous trans-location event on this chromosome. Association with heading date was most consistent across location-years on linkage groups Mrg02, Mrg12, Mrg13, and Mrg24.

Core Ideas

- An oat association-mapping panel contributed by active breeding programs worldwide.
- Characterized population structure and found subdivisions related to adaptation
- Characterized genome-wide and chromosome-specific linkage disequilibrium
- Performed association-mapping and post hoc modeling of heading date
- Found several consistently associated QTL

K. Esvelt Klos, E.M. Babiker, J.M. Bonman, and G. Hu, USDA–ARS, Small Grains and Potato Germplasm Research Unit, 1691 South 2700 West, Aberdeen, ID 83210, USA; Y.F. Huang, Dep. of Agronomy, National Taiwan Univ., Taipei, Taiwan Republic of China; W.A. Bekele, C.P. Wight, W. Yan, and N.A. Tinker, ECORC, AAFC, Ottawa, ON K1A 0C6, Canada; A.D. Beattie and B.G. Rossnagel, Univ. of Saskatchewan, Crop Development Centre, 51 Campus Dr, Saskatoon, SK S7N 5A8, Canada; D.E. Obert, Limagrain Cereal Seeds, 4846 East 450 North, Lafayette, IN 47905, USA; Å. Bjørnstad, Norwegian Univ. of Life Sciences, Dep. of Plant and Environmental Sciences, P.O. Box 5003, N-1432 Ås, Norway; M.L. Carson, USDA–ARS, Cereal Disease Laboratory, 1551 Lindig Ave, St. Paul, MN 55108, USA; S. Chao, USDA–ARS Cereal Crops Research Unit, Fargo, ND 58102, USA; B.N. Gnanesh and J. Mitchell Fetch, AAFC, Cereal Research Centre, 101 Route 100, Morden R6M 1Y5, MB, Canada; I. Griffiths and C.J. Howarth, Institute of Biological, Environmental and Rural Sciences, Aberystwyth Univ., Gogerddan, Aberystwyth, Ceredigion, SY23 3EE, UK; S.A. Harrison, Louisiana State Univ., 221 M.B. Sturgis, Baton Rouge, LA 70803, USA; A. Ibrahim, Texas A&M Univ. Dep. of Crop and Soil Sciences, 430C Heep Center, College Station Texas 77843, USA; E. Islamovic, BASF, 26 Davis Dr, Research Triangle Park, NC 27709, USA; E.W. Jackson, General Mills Inc., Manhattan, KS 66502, USA; J.L. Jannink, USDA–ARS, Plant, Soil and Nutrition Research, Robert W. Holley Center, Tower Road, Ithaca, NY 14853, USA; F.L. Kolb, Univ. of Illinois, Dep. of Crop Sciences, AE-120 Turner

Hall, 1102 S. Goodwin, Urbana, IL 61801, USA; M.S. McMullen, North Dakota State Univ., Plant Sciences, P.O. Box 6050, Fargo, ND 58108, USA; J.P. Murphy, North Carolina State Univ., Dep. of Crop Science, Campus Box 7620, Raleigh, NC 27695, USA; H.W. Ohm, Dep. of Agronomy, Purdue Univ., 915 West State Street, West Lafayette, IN 47907, USA; H.W. Rines, Univ. of Minnesota, Dep. of Agronomy and Plant Genetics, 1991 Upper Buford Circle, St. Paul, MN 55108, USA; J.A. Schlueter, Univ. of North Carolina, College of Computing and Informatics, 9201 University City Blvd., Charlotte, NC 28223; M.E. Sorrells, Cornell Univ., Plant Breeding and Genetics, 240 Emerson Hall, Ithaca, NY 14853, USA. K. Esvelt Klos and Y.F. Huang contributed equally to the formation of this work. Received 22 Oct. 2015. Accepted 4 Feb. 2016. *Corresponding author (Kathy. klos@ars.usda.gov).

CULTIVATED OAT ($2n = 6x = 42$, AACCDD) has formed a part of the human diet in many cultures for thousands of years (Murphy and Hoffman 1992). Today, the acknowledged health benefits of including oat in a balanced diet include reduction in blood cholesterol levels and mediated risk of cardiovascular disease (Anderson et al., 2009; Othman et al., 2011). Oat is often produced in areas where other temperate cereals fail to perform. This versatile ecological adaptation of cultivated oat may be due to polyploidy and subsequent divergence and specialization. Among the many traits related to adaptation, heading date (equated here with flowering time) is one of the most important. Heading date is under control of quantitative genetic factors as well as qualitative factors such as genes mediating response to vernalization and photoperiod (Holland et al., 1997, 2002; Locatelli et al., 2006; Nava et al., 2012). Through its effect on the timing of reproduction and grain filling, and because of its interaction with environmental cues, heading date can have profound effects on traits such as yield, disease susceptibility, and grain quality.

Genomics and population genomics in oat has lagged compared with many other crops due, in part, to a lack of known genetic variants with physically anchored genomic positions (Tinker et al., 2009). With the recent development of a physically anchored oat consensus map (Oliver et al., 2013) as well as the current availability of an improved higher-density consensus map (Chaffin et al., 2016) based on high-throughput SNP platforms (Huang et al., 2014; Tinker et al., 2014), the stage is set for a population-level genomics work. Understanding the genome-wide patterns of genetic diversity in elite cultivated oat may help to identify genomic regions that have influenced oat improvement in the past few decades. In genomic regions where genetic diversity has been reduced through stringent selection, targeted introgression of alleles via marker-assisted selection can be undertaken.

The genome-wide association study (GWAS) is the method of choice for accurate identification of genomic regions that are in LD with quantitative trait loci (QTL) influencing traits of interest in natural populations or breeding germplasm. One concern in GWAS is the tendency toward spurious evidence of association caused by population structure. The most common causes of spurious associations in crop species are the merging of two or more subgroups with divergent allele frequency distributions and the shared ancestry that often exists among individuals. Both of these factors violate the assumptions underlying the null distribution against which evidence of association is evaluated (Price et al., 2010). The assessment of genetic diversity within a crop species can provide the information necessary to compensate for the tendency toward false positive association in GWAS of structured samples (Price et al., 2006; Weir et al., 2006).

The Collaborative Oat Research Enterprise (CORE) was a global partnership spanning ~30 sites and collaborating institutions including funding organizations acknowledged in this work. The goals of the CORE were to develop high-throughput marker technologies, a new consensus map, and baseline diversity and association data required to support new initiatives in oat molecular breeding. The present study addresses the assembly and analysis of a germplasm diversity panel contributed by all active public oat breeding programs in North America, several programs beyond North America, as well as an additional set of lines with global or historical relevance. In this work, we evaluate genetic diversity, characterize population structure, examine the extent of pairwise LD, and present GWAS results for heading date. This work builds on two marker platforms developed by the CORE (Huang et al., 2014; Tinker et al., 2014).

Materials and Methods

Plant Materials and Phenotypic Observations

A set of 682 oat lines was nominated by breeders for inclusion in the CORE study (Supplemental Table S1). This included 109 lines chosen based on geographic area of origin to comprise a world-diversity panel, 433 lines nominated by breeders of predominantly spring oat (spring), and 140 lines nominated by breeders in the southern United States and South American oat producing regions (southern). The latter group includes oat varieties developed for both winter-grown as well as winter-dormant production systems and was delineated partly because it was not known which lines would require or respond to vernalization. After eliminating some duplicate nominations, lines that may have been subject to sample handling error, and lines with poor quality genotypes from one or both genotyping methods, 635 lines were used in these analyses (Supplemental Table S1).

Seeds from a single panicle of each line were sown in headrows in Aberdeen, ID, in 2009. Leaf tissue was used for DNA extraction and the harvested seed was used in all subsequent evaluations. DNA was extracted using the cetyl trimethyl ammonium bromide protocol of Anderson et al. (1992) with modification including use of a FastPrep homogenizer (MP Biomedical) for 5 min at 25 strokes s^{-1} to homogenize tissue.

Details of planting and harvest of each reported field test are deposited with all primary data in the public database T3/oat (http://triticeaetoolbox.org/oat/). In the 2010 and 2011 field seasons, lines were planted in April, May, or June using 50 g of seed in unreplicated four- to five-row plots of 1.2 to 2.2 m length at Aberdeen, ID, Fargo, ND, Ithaca, NY, and Tetonia, ID, in the United States; at Saskatoon, SK, in Canada; and at Ås in Norway. Lines were planted in 3.25- to 4.6-m² plots at Lacombe, AB, and Ottawa, ON, in Canada. Lines were planted in March, April, or May as hill plots using 15 to 20 seed per hill with two replicates at Saint Paul, MN, in the United States and at Aberystwyth in the United Kingdom. Plots were grown using agronomic practices standard to each region and cooperating breeding program. Days to heading was recorded as the Julian calendar day.

Genotyping

A total of 4975 SNP assays were interrogated using an oat iSelect 6K-beadchip array (Illumina) at the USDA–ARS Genotyping Laboratory at Fargo, ND, as described by Tinker et al. (2014). Multiallelic and monomorphic SNPs and those with poor genotype calls resulting from weak signal or ambiguous clustering were eliminated and the remaining biallelic SNPs were filtered to retain those with relative minor allele frequency (MAF) $\geq$ 0.01, missing data < 0.05, and heterozygosity $\leq$ 0.05. Genotypes for additional SNPs were acquired through genotyping-by-sequencing (GBS) using methods described by Huang et al. (2014). Single-nucleotide polymorphism genotype calls were made using the UNEAK pipeline (Lu et al., 2013) with the *PstI-MspI* restriction enzyme combination, 3×10^8 as the maximum reads per sequence, 9×10^8 as the maximum merged tag count, merging of multiple samples per line, and maximum error tolerance of 0.02. Population-based filtering was applied to identify GBS SNPs with the same call rate, MAF, and heterozygosity as the beadchip SNPs. A further filter was applied to eliminate GBS SNPs for which the heterozygous state was called in excess of the rare homozygous state, since these calls may represent alleles at duplicate loci. ANOVA, in SAS 9.4 (SAS Institute, 2013), was used to test mean differences in the proportion of lines used by region of origin and to estimate marker statistics by genotyping platform, oat line nomination panel, and linkage group.

Genetic Diversity, Population Structure, and Linkage Disequilibrium

Principal component analysis (PCA) was performed in TASSEL v.4.0 (Bradbury et al., 2007) using a subset of 1341 markers without missing data. The percentage of genotypic variation accounted for by the first five PCs is reported here to enable comparison with reports from other oat studies. Genetic diversity within and among primary and secondary regions of geographic origin was evaluated using analysis of molecular variance (AMOVA) in the GenAlex software package (Peakall and Smouse, 2006; Peakall and Smouse, 2012).

Model-based structure analysis was performed for K = 1 to 20 ancestral populations using ADMIXTURE v1.23 (Alexander et al., 2009) under the default parameters. To comply with the assumption of unlinked markers in linkage equilibrium, we used PLINK v1.07 (Purcell et al., 2007; http://pngu.mgh.harvard.edu/purcell/plink/) to create a subset of 340 SNPs by removing SNPs within a 20-SNP sliding window, advanced with a step of five SNPs, with r^2 > 0.1. Models were evaluated by plotting the cross-validation error against the number of subpopulations (K).

Differentiation between spring type and southern-origin subpopulations was investigated using F_{ST} where lines were designated as spring, southern, or admixed based on the model-based population structure analysis at K = 2 and an 80% cut-off of the proportion of the genome belonging to the inferred ancestral population. Admixed lines were discarded and F_{ST} was estimated between spring and southern groups using the R package HIERFSTAT (Goudet, 2005). The SNP genomic position was taken as that reported in the current community-based oat consensus map (Chaffin et al., 2016) using the chromosome labels Mrg01 through Mrg33, with Mrg standing for merged linkage group, since the consensus map was produced by merging 12 biparental linkage maps. A permutation test was performed by randomly assigning lines to two groups and estimating F_{ST} for each SNP. The minimum and maximum F_{ST} estimates were retained, and this process was repeated 100 times to create a null distribution of F_{ST} estimates.

Pairwise LD was estimated as the squared allele frequency correlation (r^2) based on raw genotype data using an optimized version of LD.Measure (Stéphane Nicolas, personal communication, 2014) implemented in the R package LDcorSV (Mangin et al., 2012). Because of suggested clustering of lines by region of origin and breeding program, and potential cryptic relatedness among lines, LD was also estimated as r^2_{sk} using a model that incorporated the first three PCs to account for population structure and a kinship matrix to account for cryptic relatedness. The LD was estimated separately within spring and southern groups as r^2_k using a model that included a kinship matrix but not the PC factor. Kinship matrices were calculated using TASSEL. The relationship between LD and genetic distance was modeled by nonlinear regression taking into account recombination, drift, low level of mutation, and sample size (Hill and Weir, 1988). The critical r^2 value as evidence of linkage was calculated by resampling of unlinked markers, corrected for structure and kinship, according to Breseghello and Sorrells (2006). The modified versions of LDit.r source codes were used to plot the nonlinear regression model of r^2 against genetic distance (Remington et al., 2001).

Association Mapping of Heading Date

Mixed linear model association analysis was performed in TASSEL under the default settings. Models incorporating the first two to four PCs and a kinship matrix were evaluated by examining their quantile–quantile plots.

Based on this, population structure was accounted for by incorporating the first three PCs, and cryptic relatedness among lines was modeled using a kinship matrix calculated using the scaled identity-by-state method. Statistical significance was taken as $p \leq 1.69 \times 10^{-5}$ to achieve a Bonferroni-corrected $\alpha = 0.05$. Post hoc associations between heading date and specific SNPs or combinations of SNPs were examined by testing general linear models with genotypes as fixed independent variables.

Results

Geographic Origins of Oat Lines

The geographic origins of oat lines nominated to the CORE and used in these analyses are summarized in Supplemental Table S1. Lines originating from a specific breeding program, regardless of ancestry, were deemed to originate from the geographic location of the program. The majority of lines originated in North America (84%) largely as breeding lines or cultivars developed at publically funded universities or government research units. Seven percent of nominated lines originated in Europe, while 3% originated in South America. Overall, 93% of nominated lines were used in these analyses, the remainder being filtered from analysis for reasons of data consistency and integrity.

Marker Characteristics, Genetic Diversity and Population Structure

After population-based filtering, a total of 1926 high-quality biallelic SNPs were called from the 6K Illumina platform, and 2635 biallelic SNPs were called based on GBS analysis (Table 1). Call rate, heterozygosity, and MAF distributions for the three oat nomination panels were remarkably similar regardless of genotyping platform (Supplemental Fig. S1). Table 1 also summarizes the numbers and properties of SNPs by genotyping platform and nomination panel. Oat nomination panels did not differ for evaluated characteristics ($P > 0.05$). GBS produced a greater proportion of rare SNP alleles than did the 6K-array genotyping platform ($P < 0.0001$). Also, GBS SNPs had lower average call rate, higher average proportion of heterozygous calls (which may represent alleles at duplicate loci), and lower average MAF ($P < 0.0001$). Genotyping-by-sequencing SNPs polymorphic in only one of the nomination panels made up 0.2% of all SNPs. A location on the new oat consensus map was available for 3012 SNPs from the filtered dataset (Supplemental Table S2). The average distance between SNPs on this map was 2.17 cM (range 0.1–30.1 cM). Most (18 of 21) linkage groups had at least one gap of >10 cM between adjacent SNPs. However, 40% of SNPs were mapped to the same position as at least one other SNP. Linkage groups differed for mean MAF ($P < 0.0001$), and also for the proportion of rare SNPs ($P < 0.0001$).

The first five PCs accounted for 25.3% of genetic marker variation. A plot of the Eigenvalues of the first 25 PCs revealed a smooth curve with no clear change

Table 1. The number of single-nucleotide polymorphisms (SNPs) used in these analyses by genotyping platform and oat line nomination panel, the number with 0.01 < minor allele frequency (MAF) < 0.05, the number polymorphic in only a single panel, the mean proportion of complete genotype calls (call rate), the mean proportion of heterozygous calls, and MAF.

Platform†	Panel‡	SNPs	Rare§	Private¶	Call rate	Heterozy-gosity	MAF
6K array	All	1926	116	NA	1.0	0.002	0.25
	WDP	1925	129	0	1.0	0.001	0.24
	Spring	1926	238	0	1.0	0.002	0.23
	Southern	1876	277	0	1.0	0.003	0.21
GBS	ALL	2635	800	NA	0.97	0.011	0.15
	WDP	2581	782	0	0.97	0.014	0.15
	Spring	2628	1024	7	0.99	0.010	0.13
	Southern	2535	711	4	0.98	0.015	0.17

† 6K array, Infinium iSelect 6K oat beadchip; GBS, genotype-by-sequencing.

‡ WDP, world diversity panel, lines selected to represent the geographic diversity of oat lines; spring, lines nominated by spring oat breeders; southern, lines nominated by winter oat breeders.

§ Rare SNPs are those with MAF < 0.05.

¶ Private SNPs are those with alleles observed in only one panel.

in slope to suggest an optimum number of population subgroups (not shown). Similarly, a plot of the cross-validation errors from model-based analysis had no clear change in slope (not shown). Scatter plots of the first three PCs, indicated a separation of lines by spring type vs. southern origin on PC 1 (Fig. 1); assignment of lines by ADMIXTURE at $K = 2$ also corresponded closely with this classification (not shown). The spring vs. southern classification accounted for 16% of genetic variation ($P < 0.0001$). Within these groups, there appeared to be some clustering based on geographic origin (Fig. 2). AMOVA indicated that region of origin accounted for 13% of the variation within spring lines and for 10% of the variation within southern lines ($P < 0.0001$). Spring oat breeding programs in this sample may be more genetically differentiated than southern oat breeding programs, but they were also more numerous.

As illustrated in Supplemental Fig. S2, uncorrected values of r^2 indicated a high degree of association among many unlinked markers as well as a high degree of association among loosely-linked markers within some chromosomes especially on Mrg02, Mrg11, Mrg15, and Mrg28. Adjustment of LD values for structure and kinship provided substantial reduction in the associations among unlinked or loosely linked makers, with the kinship adjustment providing the greatest benefit. Uncorrected associations among unlinked and loosely linked markers were highest in the Southern subset as was the benefit of adjusting for kinship. Incorporation of population structure and kinship estimates into the model was considered necessary for all further work and for providing meaningful estimates of LD decay that were comparable to previous estimates in oat (Newell et al., 2011).
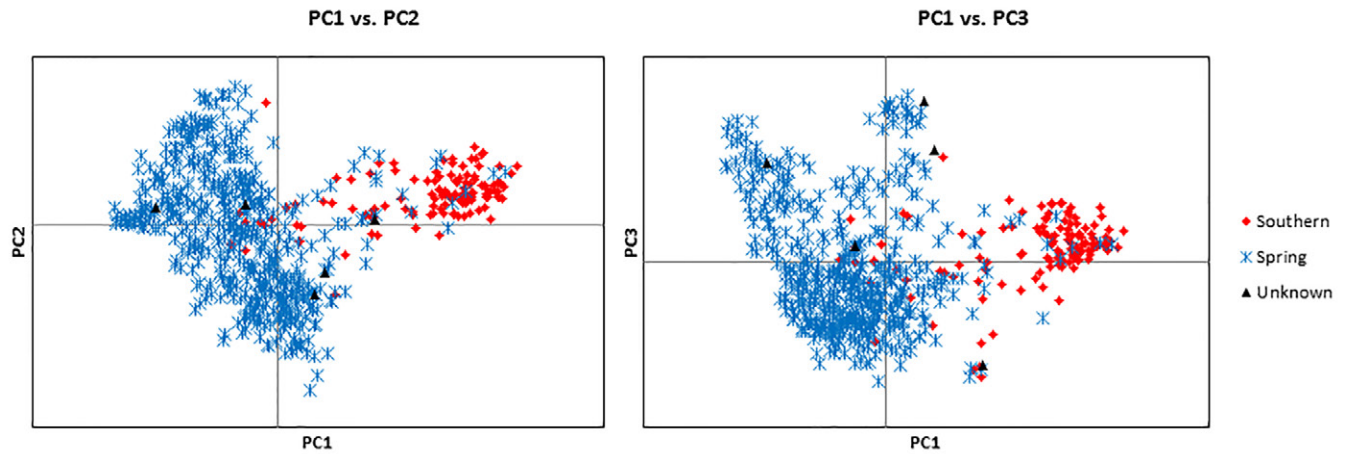
Fig. 1. Scatterplots of principal components (PCs) on the marker data with lines labeled according to spring growth habit or origin in the southern oat breeding region.
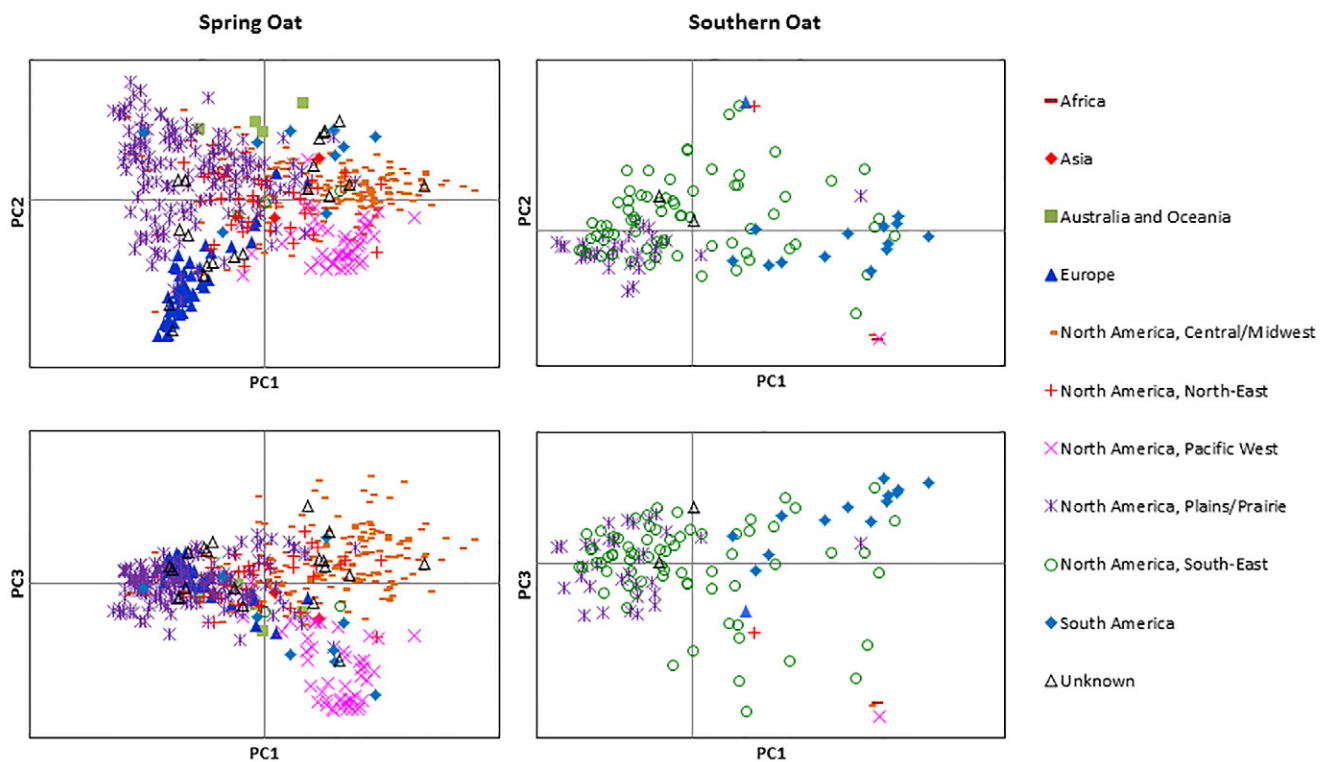

Fig. 2. Scatterplots of principal components (PCs) on the marker data within lines grouped according to spring habit or southern origin and labeled according to their geographic subregion of origin.

After adjustment for population structure and cryptic relatedness, a full set of adjusted LD decay plots were prepared (Supplemental Fig. S3). Using the nonlinear prediction, an LD of $r^2 > 0.10$ was observed within linkage groups at an average distance between markers of 0.44 cM (Supplemental Table S3). This distance ranged from 0.10 cM on linkage groups Mrg01 and Mrg24 to 1.10 cM on linkage group Mrg28. Genome-wide LD of $r^2 > 0.10$ was observed at an average distance between SNPs of 0.71 cM in the spring lines and 2.64 cM in the southern lines, indicating that LD tended to decay more rapidly within the spring growth habit group. Based on

resampling of unlinked markers, it was determined that the critical adjusted LD value for both the full set and the spring subset was 0.089, which was close to the value of 0.1 tested above. However, by the same methods, the critical value for the southern subset of lines was estimated at 0.171, suggesting that part of the reason for a slower LD decay in the winter subsets was due to a higher level of spurious LD. Even after accounting for this difference in critical value, it appeared that the winter lines showed a slower rate of LD decay on most chromosomes as typified by Mrg20 in Fig. 3. The notable exception to this
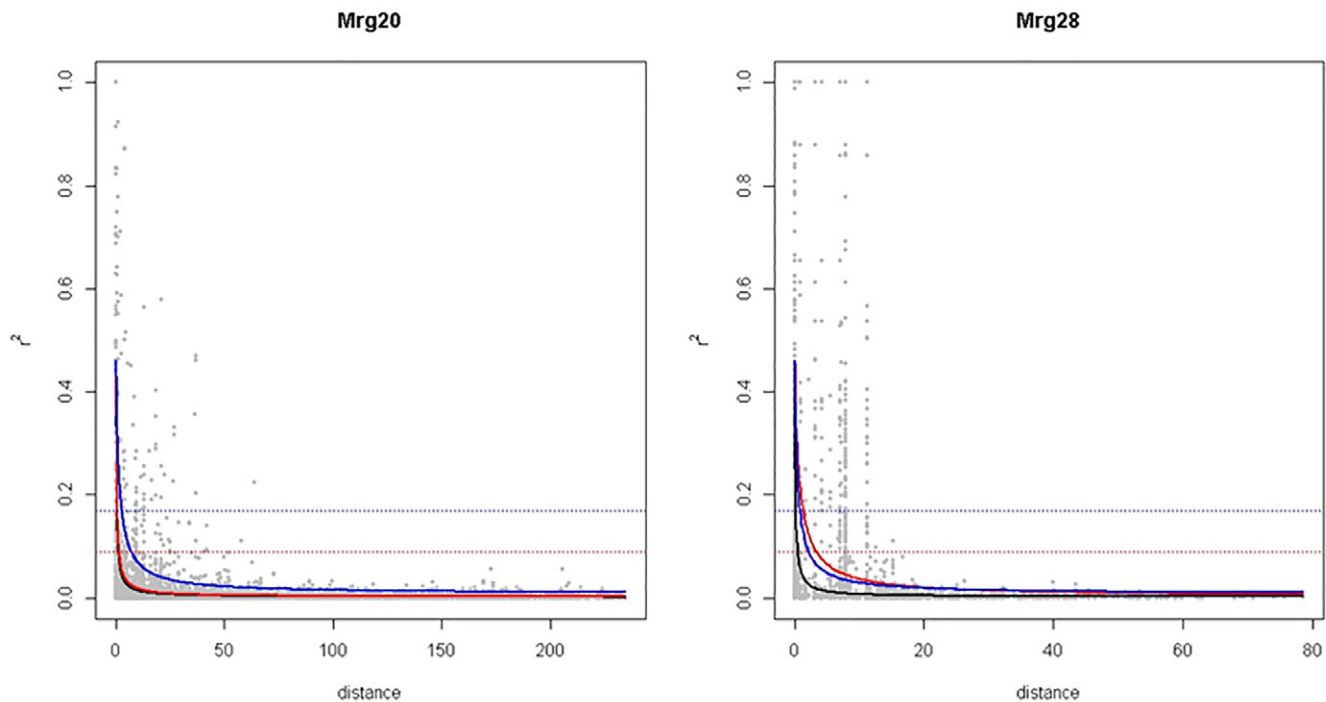
Fig. 3. Linkage disequilibrium (LD) decay rate in linkage groups Mrg20 and Mrg28. Gray dots represent adjusted r^2_{sk} coordinates for the full population based on marker pairs on these chromosomes. The black lines show the nonlinear regression curve of the r^2_{sk} values of the full set ($n = 635$), red lines show the spring r^2_k ($n = 499$), and blue lines show r^2_k of the southern lines ($n = 124$). Critical LD values are shown as horizontal broken lines at 0.089 and 0.17 for the full or spring set (red) and southern set (blue) respectively.

was on Mrg28, where spring lines showed a substantially slower rate of LD decay than southern (Fig. 3).

Genetic differentiation between spring oat and southern-origin oat was investigated using F_{ST} for individual loci. Based on the permutation test, expectation under the null distribution was 95% of $F_{ST} < 0.0039$. However, 37% of observed values met this threshold, so a more conservative level of the most extreme 1% was chosen to identify genomic regions with evidence of differentiation between population subgroups. Clusters of markers with F_{ST} in the most extreme 1% of the distribution (Fig. 4, colored red) were observed on linkage groups Mrg11 (63 SNPs at 45.6–77.8 cM), Mrg12 (four SNPs at 58.5–63.5 cM), Mrg20 (10 SNPs at 98.8–156.7 cM), and Mrg21 (six SNPs at 111.3–181.1 cM). On these linkage groups, the majority of SNPs with extreme F_{ST} values were within 10 cM of a single position (53, 144, and 115 cM on Mrg11, Mrg20, and Mrg21, respectively). A five-SNP sliding window average F_{ST} was also evaluated (not shown) and results were similar except that extreme F_{ST} values on Mrg11, Mrg20, and Mrg21 were more prominent than the highest values observed for SNPs on other linkage groups.

Association Mapping Results

Data on heading date were available for one or more germplasm panel for up to 15 location-years (Supplemental Table S4). Overall, the strongest evidence of association was observed on linkage group Mrg02 from 29.1 to 34.1 cM (Table 2). The SNP GMI_ES_CC4504_192

at 30.1 cM is especially notable for the fact that p-values were below the Bonferroni-adjusted threshold in eight of 15 location-years. The SNPs where the Bonferroni-adjusted threshold was achieved in at least one location-year, and p-values < 0.0001 in additional location-years, included GMI_ES03_c14909_90 on Mrg12 at 42 cM, avgbs_244127 on Mrg13 at 30.3 cM, and avgbs_53470 on Mrg24 at 64.6 cM (Table 2). Other evidence of association was more specific to location-year.

Post Hoc Modeling of Single-Nucleotide Polymorphisms Affecting Heading Date

The map locations with the largest and most consistent GWAS effects on heading date on Mrg02 (29–34 cM) and Mrg12 (42 cM) were chosen for further inspection by post hoc combined model analysis. On Mrg02, a single haplotype combining the two SNPs with the most consistent association (GMI_ES_LB_11316 and GMI_ES_CC4504_192) was evaluated. The explained variance (adjusted r^2) of general linear models fitting single SNPs and haplotypes (Table 3) showed that GMI_ES_LB_11316 consistently explained more variance than GMI_ES_CC4504_192, while the haplotype containing both of these SNPs consistently increased the adjusted r^2 beyond that of GMI_ES_LB_11316 alone. Adding GMI_ES03_c14909_90 from Mrg12 to a combined model increased the adjusted r^2 to approximately the sum of the best models from Mrg02 and Mrg12, suggesting that genetic factors on these two linkage groups are acting additively.
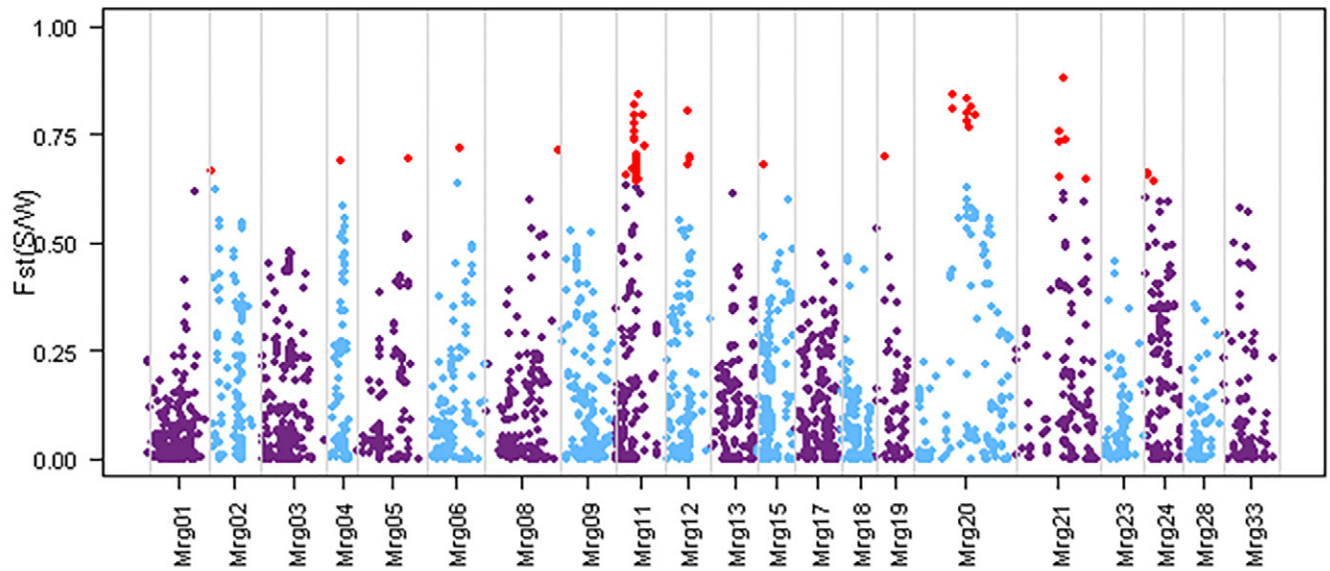
Fig. 4. Single-locus F_{ST} estimates contrasting the proportion of within-group variance for lines with spring or southern origin and plotted according to genomic location. Markers with F_{ST} values in the highest 1% of the full distribution are shown in red.

Table 2. Strength of SNP associations with heading date in 10 locations over 2 yr.

Marker	Linkage group	cM	Aberdeen		Aberystwyth		Ås	Fargo		Ithaca		Lacombe	Ottawa		Saskatoon	St. Paul	Tetonia
			2010	2011	2010	2011	2011	2010	2011	2010	2011	2010	2010	2011	2010	2011	2011
avgbs_202898	Mrg01	39.3		**	**	**	*	*		**	****	*		**	*	*	
GMI_ES_CC4504_192	Mrg02	30.1	***	****	****	***	**	****		****	****	***	****	**	****	**	****
GMI_ES15_c1210_347	Mrg02	30.1	**	**	**	**	*	**		**	*	**	**		**		****
GMI_DS_LB_6375	Mrg02	30.6		*	**	**	**	**		***	**	*	**	**	****	*	**
GMI_ES_LB_11316	Mrg02	34.1		**	****	***	**	**		**		***	****		**		***
avgbs_223032	Mrg02	71.6			**			**	****			**					*
avgbs_1676	Mrg02	87.3		*	**	*		**	**	****			**		*		**
GMI_ES22_c11766_112	Mrg02	85.2		*	**	*		**	*	****		*		**	*		**
GMI_GBS_24858	Mrg06	45.2		**		****	*			*	*				*		**
avgbs_15433	Mrg06	67.6				**		****									
avgbs_222856	Mrg11	32.4							**		**				*	****	
GMI_GBS_98189	Mrg12	35.3						****	*								
GMI_DS_LB_5810	Mrg12	40.9		*	**	**		****	**	*	*		**	*	**	**	**
GMI_ES03_c14909_90	Mrg12	42		**	****	**		****	****	****	***	***	***	**	****	**	**
avgbs_231664	Mrg12	56.7						****		**	**	**	**	*			**
avgbs_244127	Mrg13	30.3	**				**		**	**	****				**	****	*
GMI_ES15_c6576_330	Mrg13	33									**				*	****	
avgbs_85626	Mrg13	35.9	*							*	**			**	**	*	****
GMI_ES_LB_11438	Mrg17	110.4		****	*					*							
avgbs_221360	Mrg17	110.4		****	**				**	**	*	*	*	*	*		
avgbs_87322	Mrg20	72.1		*					****	**			*	*	*	*	*
GMI_GBS_114417	Mrg20	226.6				*	*	****	*				**	*	**		
avgbs_53470	Mrg24	64.6		*	**		****	**				*				*	****

* Significant at the 0.05 probability level.

** Significant at the 0.01 probability level.

*** Significant at the 0.001 probability level.

**** Significant at the 1.7×10^{-5} probability level.

Table 3. Adjusted r^2 values from post hoc linear model analyses.

Chromosomes	Mrg02	Mrg02	Mrg02	Mrg12	Mrg02 × Mrg12
Markers or haplotypes	SNP_192†	SNP_11316‡	Haplotype§	SNP_90¶	Coined haplotype#
Aberdeen 2010	0.16	0.18	0.19	0.04	0.24
Aberdeen 2011	0.17	0.18	0.20	0.05	0.24
Aberystwyth 2010	0.09	0.13	0.15	0.03	0.21
Aberystwyth 2011	0.09	0.12	0.12	0.02	0.17
Fargo 2010	0.17	0.2	0.21	0.06	0.27
Fargo 2011	0.14	0.17	0.19	0.09	0.26
Ithaca 2010	0.13	0.2	0.22	0.05	0.29
Ithaca 2011	0.18	0.19	0.22	0.08	0.31
Lacombe 2010	0.12	0.17	0.18	0.05	0.25
Norway 2011	0.14	0.19	0.19	0.13	0.28
Ottawa 2010	0.26	0.28	0.30	0.11	0.38
Ottawa 2011	0.12	0.14	0.17	0.06	0.21
Saint Paul 2011	0.08	0.12	0.13	0.04	0.17
Saskatoon 2010	0.22	0.24	0.27	0.11	0.37
Tetonia 2011	0.17	0.23	0.25	0.04	0.31
Mean	0.15	0.18	0.20	0.07	0.27

† SNP_192, GMI_ES_CC4504_192

‡ SNP_11316, GMI_ES_LB_11316

§ Haplotype, GMI_ES_CC4504_192 × GMI_ES_LB_11316

¶ SNP_90, GMI_ES03_c14909_90

Combined haplotype, GMI_ES_CC4504_192 × GMI_ES_LB_11316 × GMI_ES03_c14909_90

Discussion

In this study, we characterize genetic diversity, population structure, and genomic patterns of LD in a sample of elite oat germplasm using SNPs derived from two high-throughput genotyping platforms. This information was applied using association mapping to identify QTL contributing to heading date.

The 6K SNP array and GBS genotyping technologies produced SNP sets differing in MAF distribution. Rare SNPs were notably lacking in the 6K-array platform. This lack of rare alleles demonstrates the effectiveness of deliberate selection for common alleles in the development of this platform (Oliver et al., 2013; Tinker et al., 2014) but may also be interpreted as ascertainment bias, resulting in a lack of rare variants representative of the population. Genome-wide association study approaches have proven effective at identifying common variants with moderate- to large-effect sizes on quantitative traits but less effective at identifying rare variants or those with small or context-dependent effects on phenotype (Brachi et al., 2011). Enrichment for common SNPs may have the benefit of focusing genotyping resources where GWAS is most effective. In addition, the enrichment for common alleles in the 6K array provides cross-applicability of the genotyping platform, a feature that has greatly increased the density of SNPs with map positions from numerous populations (Tinker et al., 2014).

The population genetic parameters observed in the CORE lines were consistent with those previously reported. In the largest studies, using diversity array technology markers, evidence supported a weak population structure for oat with the first three PCs accounting for the major population subgroupings as overlapping and diffuse clusters (Asoro et al., 2013; Newell et al., 2011, 2012; Tinker et al., 2014). We consider the CORE sample to be only weakly structured as well. The first three PCs accounted for 23.8% of the genetic variation, which is within the range of 14 to 24.3% found in previous reports (Achleitner et al., 2008; Asoro et al., 2013; Huang et al., 2014; Newell et al., 2011, 2012). The suggested population subgroupings of the CORE lines were also consistent with previously reported patterns. The primary division by PC or model-based *K*-means clustering was into overlapping groups of lines differentiated by adaptation to subtropical growth after fall sowing. Souza and Sorrells (1991) and Newell et al. (2011) both observed this pattern and also noted the confounding of adaptation with the proportion of red oat (*A. sativa* ssp. *byzantina* K. Koch) ancestry. This clustering of oat lines could be a result of recent admixture with red oat, typified by the emergence of the Red Rust-proof landrace in the southern United States in the 1860s and its subsequent use in southern oat breeding programs (Stanton, 1955). Alternately, the binary clustering observed in the CORE could be based on the tendency of breeders to avoid crosses between oat lines known or perceived to have widely differing growth habit or adaptation.

Within the spring and southern oat groups, the CORE lines appeared to be clustered according to geographic origin, a pattern also observed in previous studies. Apparent clustering by breeding programs should be viewed with caution given the known tendency for family relatedness to generate artifacts in PCA (Price et al., 2010). However, regional clustering and the apparent proximity of breeding programs that address similar environments and production systems probably do reflect germplasm exchange and recent gene flow. The oat community in North America is dominated by publically funded breeding programs that actively exchange material. A typical crossing program will incorporate a mixture of parents from three sources, listed in order of decreasing frequency: (i) predominantly elite breeding lines from within the program (for recurrent improvement), (ii) recent cultivars or exchanged material from other programs in a similar target environment, and (iii) exotic material for the introduction of new traits such as disease resistance.

Extensive long-range LD was observed when estimated without adjustment for population structure and kinship. As expected, based on visual examination of PC clustering, kinship was sufficient to control for this within growth habit groups, while a model incorporating both PC and kinship performed better across the full population. This correction eliminated most major blocks of LD that spanned multiple chromosomes; however, blocks of regional LD remained and were observed most prominently on linkage groups Mrg02, Mrg11, and
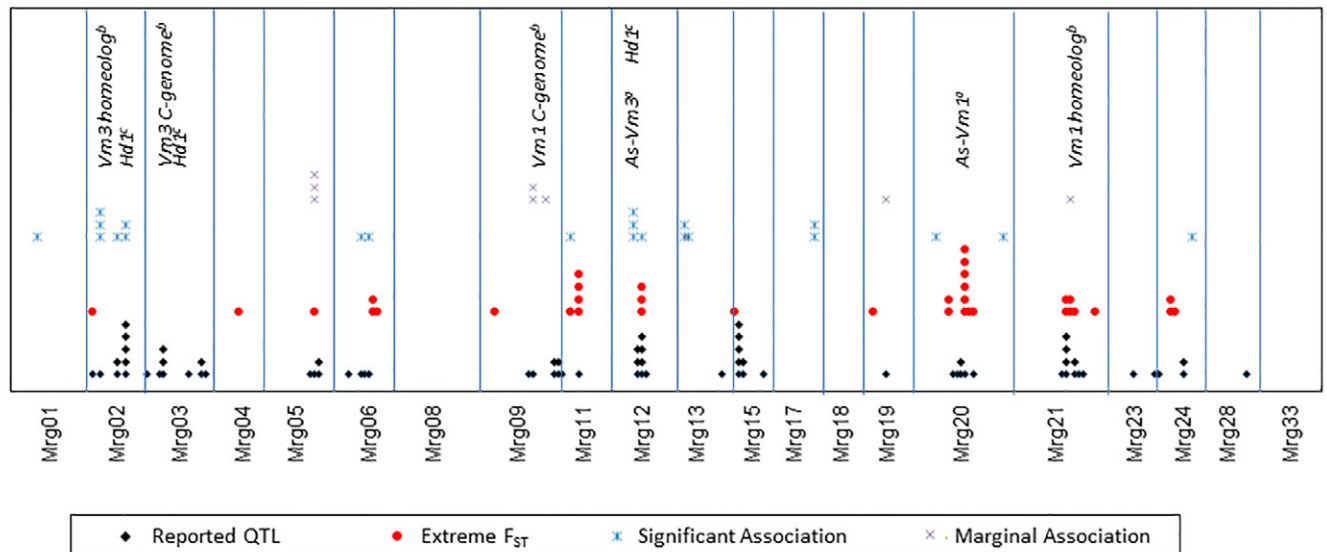
Fig. 5. The relative locations (in 10-cM bins) of positive evidence for quantitative trait loci (QTL) influencing adaptation in oat. Reported QTL were for days to heading or days to maturity in the field under standard management or after early or late sowing, days from emergence to heading under controlled conditions with or without prior vernalization, and vernalization response (Holland et al., 1997, 2002; Nava et al., 2012; Tanhuanpää et al., 2012). Extreme F_{ST} observed for single-nucleotide polymorphisms mapped to the same genomic position are summarized by a single point. Significant association refers to p-values $\leq 1.69 \times 10^{-5}$ in this study. Marginal evidence of association ($p < 0.0001$) is reported only within 10 cM of QTL reported in the literature. Candidate gene locations are taken (a) from Nava et al. (2012), inferred (b) based on homeologous matches, or based (c) on blast hits to oat from rice genome sequence (Locatelli et al., 2006).

Mrg15. Such blocks increase the likelihood that sufficient LD will exist to detect QTL within the block but present problems for fine-scale localization of QTLs. This problem may arise because the most significant p-value from among several in a block of SNPs in high LD may not be at the SNPs closest to the QTL but rather at the SNPs with the best allele–frequency match. There were also regions of the oat genome with very low levels of LD. The QTL within these regions may not be in sufficient LD with any marker for detection in an association analysis.

Linkage disequilibrium decayed more slowly in the lines nominated to the southern panel than the full set of lines or to those in the spring panel (Supplemental Table S3; Supplemental Fig. S3). This may reflect differences in breeding history between two distinct oat subpopulations. It may also be the result of a higher degree of relatedness among southern lines, since only three breeding programs nominated more than half of the lines in the southern panel. A third possibility is that estimates of LD extent are inflated in the southern panel because of physical deviations from the consensus map. The current oat consensus map was developed using predominantly spring–spring oat crosses, so marker order and spacing will be optimized for spring lines wherever differences exist (Chaffin et al., 2016). Given the previously noted differences in translocation frequency among traditional *A. sativa byzantina* and *A. sativa*-type lines (Jellen and Beard, 2000), inconsistencies in marker order and spacing between the consensus map and the majority of southern lines could inflate distances at which significant LD occurs. Additional possibilities leading to a slower LD decay in southern germplasm

could include differences in admixture or differences in selection practice (Flint-Garcia et al., 2003).

Patterns of low genetic diversity and high LD have been observed in genomic regions subjected to recent selection (Palaisa et al., 2004). Regions with high F_{ST} values between spring and southern groups may reflect genomic regions that have undergone selective pressure for phenotypes related to adaptation. The most notable linkage groups for extreme F_{ST} were Mrg11, Mrg20, and Mrg21. The genomic positions (summarized within 10-cM bins) of adaptation-related candidate genes and QTL reported in the literature, and of associations from this study, are shown relative to extreme F_{ST} in Fig. 5. Mrg11 may correspond to linkage group OT7, on which a QTL influencing growing degree days to flowering was identified when the OT population was tested at an early sowing date in Iowa (Holland et al., 2002). In addition, higher than average long-range LD was observed in this region, suggesting potential for adaptation-related selective pressure. Mrg20 and Mrg21 represent homeologous chromosomes (Chaffin et al., 2016). Nava et al. (2012) mapped an ortholog of the vernalization gene *Vrn1* to markers now placed on Mrg20, and QTL for vernalization response have also been mapped to this linkage group (Holland et al., 2002). The Mrg21 linkage group contains markers mapped to KO24 and OT31, to which vernalization response QTL were also mapped (Holland et al., 1997, 2002; Nava et al., 2012). On the other hand, long-range LD was not observed on these linkage groups after population structure and kinship adjustment. A smaller cluster of SNPs with extreme F_{ST} values was also

observed on Mrg12, notable because of the location overlap with vernalization response gene *As-vrn3* (Nava et al. 2012) and consistently strong evidence of association with heading date in the CORE.

As a note of caution, structural genomic rearrangements can also result in localized genetic differentiation between populations because they serve as a barrier to gene flow. Chaffin et al. (2016) evaluated alternate marker placements based on information specific to the biparental mapping populations available to their consensus map. They concluded that clusters of markers with consistent alternate placement suggested the potential for numerous translocations in oat, many of which were heterologous among mapping populations. Population-specific alternate marker placements were observed on regions of Mrg20 and Mrg21, overlapping with the regions of extreme F_{ST} observed in this study. In particular, evidence of a Mrg09–Mrg20 translocation was observed in two spring–spring crosses (IL86-1156 × Clintland 64 and AC Assiniboia × MN841801), which could influence F_{ST} estimates if this translocation were not present in southern lines. Additionally, F_{ST} is not a formal test of differential selection between groups, and thus the occurrence of corresponding vernalization response QTL may be coincidental.

We speculated that adaptation to different planting times and different flowering cues could underlie genetic structure. Hence, we selected heading date for GWAS as an indicator trait for local adaptation. Numerous regions of the genome were associated with heading date within location-years, even using the extremely conservative Bonferroni adjustment for multiple testing. Several of these associations coincide with candidate gene locations or validate QTL previously reported based on studies of biparental populations (Fig. 5). The QTL contributing to variation in flowering time on linkage group Mrg02 have been mapped in the 'Kanota' × 'Ogle', Ogle × 'Tam-O-301', 'Terra' × 'Marion', and UFRGS8 × P68/5*Starter populations (De Koeyer et al., 2004; Holland et al., 1997, 2002; Locatelli et al., 2006; Siripoonwiwat et al., 1996). For the most part, these QTL overlap with the markers associated in the CORE at 71.6 to 85.2 cM. However, the QTL mapped in Kanota × Ogle covered a large region of Mrg02 between 6 and 108 cM when days to flowering in the greenhouse was evaluated with or without vernalization. This finding overlaps with associations reported here at 29.1 to 34.1 cM as well as those more distal. Candidate genes for heading date associations located on both Mrg02 as well as Mrg12 include *Vrn3* [equivalent to *HD1* in rice (*Oryza sativa* L.) or *FT* in *Arabidopsis thaliana* (L.) Heynh.] and the linked regulatory gene Constans (*CO*). A strong association between heading date and orthologs of *Hd1* have been observed in many cereals (Liu et al. 2015). Nava et al. (2012) cloned a partial ortholog of *Vrn3* in oat and mapped this to the region of a Kanota × Ogle map that corresponds to position 40.9 cM of Mrg12. Chaffin et al. (2016) demonstrated that this region is highly homeologous with two possible regions

on Mrg12 at 30 or 85 cM. It is interesting that both of these regions, as well as the region on Mrg12, show strong associations with heading date. Further evidence for the location of these candidate genes was provided by the identification of both *HD1* and *CO* from rice at these same locations using the syntenic matches to rice presented by Chaffin et al. (2016). A third homeologous region was identified on Mrg03, where QTL have been located in other studies (Fig. 5). Locatelli et al. (2006) previously suggested *CO* as a candidate gene for the strong day-length-dependent QTL that they mapped to Kanota × Ogle group 17. Since this region corresponds to the same regions above, *CO* and the inferred day-length-insensitivity gene of oat (*Di1*) can also be considered as candidate genes for these associations.

The QTL on Mrg06 and Mrg11 in this study were in chromosomal regions with QTL reported by Holland et al. (2002) for growing degree days to heading in Ames, IA, after early seeding in the field. Those QTL were not observed after late seeding or growth in the greenhouse, suggesting effects dependent on environment. That may explain the lack of consistency of association across location-years in the CORE. The QTL on Mrg12, overlapping with the consistent associations here, were also found under diverse growing conditions in Kanota × Ogle, Ogle × Tam-O-301, and 'Aslak' × 'Matilda' populations (Holland et al., 1997, 2002; Siripoonwiwat et al., 1996; Tanhuanpää et al., 2012). In this region, lies the vernalization gene *As-vrn3* (Nava et al. 2102), a match validated by synteny with the rice genome (Chaffin et al. 2016). The QTL on Mrg11 and Mrg20 did not overlap with genomic regions containing SNPs with extreme F_{ST} estimates.

Heading date or flowering time is known to be controlled by several QTL, especially in temperate cereals such as barley (*Hordeum vulgare* L.) and wheat (*Triticum aestivum* L.) (Cockram et al., 2007; Muñoz-Amatriaín et al., 2014; Zanke et al., 2014). Here, we demonstrated that linear models using only two significant genomic regions explained almost 30% of phenotypic variance for heading date. In winter wheat, models using five significant SNPs explained more than 37% of the phenotypic variance (Zanke et al., 2014), while in spring wheat it was found that 10 to 15 SNPs explained up to 70% of phenotypic variation (Bogard et al., 2014). Similar results were obtained in rice (Gao et al., 2014). Because of the inconsistency among environments of other genomic regions, as well as the risk of model overfitting, we did not evaluate more complex models. We suggest, however, that these two major regions (Mrg02 and Mrg12) would be ideal for targeted molecular breeding, while genomic selection methods could potentially use many more of these QTL depending on intensity and on the presence of multiple traits in the selection index.

In summary, this study provides a characterization of the genetic diversity, population structure, and genome-wide patterns of LD based on the largest and most diverse sample of elite oat lines available to date. The first three PCs were sufficient to differentiate among

subgroups of this weakly structured population, and patterns of clustering were consistent with our understanding of the history of oat breeding and the complexity of the oat genome. The assessments of LD and population structure provided here can be used to aid the design of GWAS experiments in the CORE sample as demonstrated by successful identification of QTL for heading date. Markers for these QTL may prove useful to breeding programs attempting to optimize regional adaptation and productivity.

Supplemental Information Available

Supplemental Table S1. Origin of lines nominated to this study (and included in the analyses) by nomination panel and growth habit as reported by nominating breeders.

Supplemental Table S2. The number of SNPs per linkage group (with call rate ≥ 0.95, MAF ≥ 0.01, heterozygosity ≤ 0.05); the number of rare SNPs, and the mean SNP call rate, heterozygosity, MAF; and the largest observed gap (in cM) between adjacent SNPs.

Supplemental Table S3. The mean pairwise distance in cM between markers at which $r^{2\ddagger} > 0.10$ or $r^2 >$ critical value[††] for all oat lines, and within groups of lines with spring growth habit or southern origin.

Supplemental Table S4. Summary of mean heading date by panel for 10 locations and 2 yr.

Supplemental Fig. S1. SNP call rate, heterozygosity, and minor allele frequency distributions by genotyping platform and oat line nomination panel.

Supplemental Fig. S2. Heat maps of pairwise LD calculated as raw r^2, and after linear adjustment for population structure using PC (r^2_s), cryptic relatedness using a kinship matrix (r^2_k) and both (r^2_{sk}) for the full set of oat lines and within the spring and southern nomination panels. Uncorrected r^2 values are shown below the diagonal.

Supplemental Fig. S3. Pairwise linkage decay (LD) plots for 21 consensus chromosome representations in oat. Gray dots represent adjusted r^2_{sk} coordinates for the full population based on marker pairs on these chromosomes. The black lines show the nonlinear regression curve of the r^2_{sk} values of the full set ($n = 635$), red lines show the spring set r^2_k ($n = 499$) and blue lines show r^2_k of the southern lines ($n = 124$). Critical LD values are shown as horizontal broken lines at 0.089, 0.089 and 0.17 for the full set (black), spring set (red) and southern set (blue), respectively.

References

Achleitner, A., N.A. Tinker, E. Zechner, and H. Buerstmayr. 2008. Genetic diversity among oat varieties of worldwide origin and associations of AFLP markers with quantitative traits. Theor. Appl. Genet. 117:1041–1053. doi:10.1007/s00122-008-0843-y

Alexander, D.H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664. doi:10.1101/gr.094052.109

Anderson, J.A., Y. Ogihara, M.E. Sorrells, and S.D. Tanksley. 1992. Development of a chromosomal arm map for wheat based on RFLP markers. Theor. Appl. Genet. 83:1035–1043. doi:10.1007/BF00232969

Anderson, J.W., P. Baird, R.H. Davis, Jr., S. Ferreri, M. Knudtson, A. Koraym, et al. 2009. Health benefits of dietary fiber. Nutr. Rev. 67:188–205. doi:10.1111/j.1753-4887.2009.00189.x

Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott, N.A. Tinker, and J.L. Jannink. 2013. Genomic, marker-assisted, and pedigree-BLUP selection methods for β-glucan concentration in elite oat. Crop Sci. 53:1894–1906. doi:10.2135/cropsci2012.09.0526

Bogard, M., C. Ravel, E. Paux, J. Bordes, F. Balfourier, S.C. Chapman, et al. 2014. Predictions of heading date in bread wheat (*Triticum aestivum* L.) using QTL-based parameters of an ecophysiological model. J. Exp. Bot. 65:5849–5865. doi:10.1093/jxb/eru328

Brachi, B., G.P. Morris, and J.O. Borevitz. 2011. Genome-wide association studies in plants: The missing heritability is in the field. Genome Biol. 12:232. doi:10.1186/gb-2011-12-10-232

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:10.1093/bioinformatics/btm308

Breseghello, F., and M.E. Sorrells. 2006. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177. doi:10.1534/genetics.105.044586

Chaffin, A.S., Y.F. Huang, S. Smith, W.A. Bekele, E. Babiker, B.N. Gnanesh, B.J. Foresman, S.G. Blanchard, J.J. Jay, R.W. Reid, C.P. Wight, S. Chao, R. Oliver, E. Islamovic, F.L. Kolb, C. McCartney, J.W. Mitchell Fetch, A.D. Beattie, Å. Bjørnstad, J.M. Bonman, T. Langdon, C.J. Howarth, C.R. Brouwer, E.N. Jellen, K.E. Klos, J.A. Poland, T.F. Hseih, R. Brown, E. Jackson, J.A. Schlueter, and N.A. Tinker. 2016. A consensus map in cultivated hexaploid oat reveals conserved grass synteny with substantial sub-genome rearrangement. Plant Genome 9. doi:10.3835/plantgenome2015.10.0102

Cockram, J., H. Jones, F.J. Leigh, D. O'Sullivan, W. Powell, D.A. Laurie, et al. 2007. Control of flowering time in temperate cereals: Genes, domestication, and sustainable productivity. J. Exp. Bot. 58:1231–1244. doi:10.1093/jxb/erm042

De Koeyer, D.L., N.A. Tinker, C.P. Wight, J. Deyl, V.D. Burrows, L.S. O'Donoughue, et al. 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

Flint-Garcia, S.A., J.M. Thornsberry, and E.S. Buckler, IV. 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54:357–374. doi:10.1146/annurev.arplant.54.031902.134907

Gao, H., M. Jin, X.M. Zheng, J. Chen, D. Yuan, Y. Xin, et al. 2014. Days to heading 7, a major quantitative locus determining photoperiod sensitivity and regional adaptation in rice. Proc. Natl. Acad. Sci. USA 111:16337–16342. doi:10.1073/pnas.1418204111

Goudet, J. 2005. Hierfstat, a package for R to compute and test hierarchical *F*-statistics. Mol. Ecol. Notes 5:184–186. doi:10.1111/j.1471-8286.2004.00828.x

Hill, W.G., and B.S. Weir. 1988. Variances and covariances of squared linkage disequilibria in finite populations. Theor. Popul. Biol. 33:54–78. doi:10.1016/0040-5809(88)90004-4

Holland, J., H. Moser, L. O'Donoughue, and M. Lee. 1997. QTLs and epistasis associated with vernalization responses in oat. Crop Sci. 37:1306–1316. doi:10.2135/cropsci1997.0011183X003700040047x

Holland, J., A. Portyanko, L. Hoffman, and M. Lee. 2002. Genomic regions controlling vernalization and photoperiod responses in oat. Theor. Appl. Genet. 105:113–126. doi:10.1007/s00122-001-0845-5

Huang, Y.F., J.A. Poland, C.P. Wight, E.W. Jackson, and N.A. Tinker. 2014. Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat. PLoS ONE 9:E102448. doi:10.1371/journal.pone.0102448

Jellen, E. and J. Beard. 2000. Geographical distribution of a chromosome 7C and 17 intergenomic translocation in cultivated oat.

Locatelli, A.B., L. Federizzi, S. Milach, C.P. Wight, S. Molnar, J.T. Chapados, et al. 2006. Loci affecting flowering time in oat under short-day conditions. Genome 49:1528–1538. doi:10.1139/g06-108

Liu, H., H. Liu, L. Zhou, Z. Zhang, X. Zhang, M. Wang, et al. 2015. Parallel domestication of the heading date gene in cereals. Mol. Biol. Evol. 32:2726–2737. doi:10.1093/molbev/msv148

Lu, F., A.E. Lipka, J. Glaubitz, R. Elshire, J.H. Cherney, M.D. Casler, et al. 2013. Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. PLoS Genet. 9:E1003215. doi:10.1371/journal.pgen.1003215

Mangin, B., A. Siberchicot, S. Nicolas, A. Doligez, P. This, and C. Cierco-Ayrolles. 2012. Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. Heredity 108:285–291. doi:10.1038/hdy.2011.73

Muñoz-Amatriaín, M., A. Cuesta-Marcos, J.B. Endelman, J. Comadran, J.M. Bonman, H.E. Bockelman, et al. 2014. The USDA barley core collection: Genetic diversity, population structure, and potential for genome-wide association studies. PLoS ONE 9:E94688. doi:10.1371/journal.pone.0094688

Murphy, J.P., and L.A. Hoffman. 1992. The origin, history, and production of oat. In: H.G. Marshall, and M.E. Sorrells, editors, Oat science and technology. ASA and CSSA, Madison, WI. p. 1–28.

Nava, I.C., C.P. Wight, M.T. Pacheco, L.C. Federizzi, and N.A. Tinker. 2012. Tagging and mapping candidate loci for vernalization and flower initiation in hexaploid oat. Mol. Breed. 30:1295–1312. doi:10.1007/s11032-012-9715-x

Newell, M.A., F.G. Asoro, M.P. Scott, P.J. White, W.D. Beavis, and J.-L. Jannink. 2012. Genome-wide association study for oat (*Avena sativa* L.) beta-glucan concentration using germplasm of worldwide origin. Theor. Appl. Genet. 125:1687–1696. doi:10.1007/s00122-012-1945-0

Newell, M.A., D. Cook, N.A. Tinker, and J.L. Jannink. 2011. 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

Oliver, R.E., N.A. Tinker, G.R. Lazo, S. Chao, E.N. Jellen, M.L. Carson, et al. 2013. SNP discovery and chromosome anchoring provide the first physically-anchored hexaploid oat map and reveal synteny with model species. PLoS ONE 8:E58068. doi:10.1371/journal.pone.0058068

Othman, R.A., M.H. Moghadasian, and P.J. Jones. 2011. Cholesterol-lowering effects of oat beta-glucan. Nutr. Rev. 69:299–309. doi:10.1111/j.1753-4887.2011.00401.x

Palaisa, K., M. Morgante, S. Tingey, and A. Rafalski. 2004. Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep. Proc. Natl. Acad. Sci. USA 101:9885–9890. doi:10.1073/pnas.0307839101

Peakall, R., and P.E. Smouse. 2006. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6:288–295. doi:10.1111/j.1471-8286.2005.01155.x

Peakall, R., and P.E. Smouse. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research: An update. Bioinformatics 28:2537–2539. doi:10.1093/bioinformatics/bts460

Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904–909. doi:10.1038/ng1847

Price, A.L., N.A. Zaitlen, D. Reich, and N. Patterson. 2010. New approaches to population stratification in genome-wide association studies. Nat. Rev. Genet. 11:459–463. doi:10.1038/nrg2813

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. deBakker, M.J. Daly, and P.C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analysis. Am. J. Hum. Genet. 81:559–575. doi:10.1086/519795

Remington, D.L., J.M. Thornsberry, Y. Matsuoka, L.M. Wilson, S.R. Whitt, J. Doebley, S. Kresovich, M.M. Goodman, and E.S. Buckler. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. USA 98:11479–11484. doi:10.1073/pnas.201394398

SAS Institute. 2013. SAS system for Windows. v. 9.4. SAS Inst. Inc., Cary, NC.

Siripoonwiwat, W., L.S. O'Donoughue, D. Wesenberg, D.L. Hoffman, J.F. Barbosa-Neto and M.E. Sorrells. 1996. Chromosomal regions associated with quantitative traits in oat. J. Quant. Trait Loci 2.

Souza, E., and M. Sorrells. 1991. Relationships among 70 North American oat germplasms: II. Cluster analysis using qualitative characters. Crop Sci. 31:605–612. doi:10.2135/cropsci1991.0011183X003100030011x

Stanton, T.R. 1955. Oat identification and classification. Tech. Bull. No. 1516. USDA–ARS, Washington, DC.

Tanhuanpää, P., O. Manninen, A. Beattie, P. Eckstein, G. Scoles, B. Rossnagel, et al. 2012. An updated doubled haploid oat linkage map and QTL mapping of agronomic and grain quality traits from Canadian field trials. Genome 55:289–301. doi:10.1139/g2012-017

Tinker, N.A., S. Chao, G.R. Lazo, R.E. Oliver, Y.F. Huang, J.A. Poland, et al. 2014. A SNP genotyping array for hexaploid oat. Plant Gen. 7.

Tinker, N.A., A. Kilian, C.P. Wight, K. Heller-Uszynska, P. Wenzl, H.W. Rines, et al. 2009. New DArT markers for oat provide enhanced map coverage and global germplasm characterization. BMC Genomics 10:39. doi:10.1186/1471-2164-10-39

Weir, B.S., A.D. Anderson, and A.B. Hepler. 2006. Genetic relatedness analysis: Modern data and new challenges. Nat. Rev. Genet. 7:771–780. doi:10.1038/nrg1960

Zanke, C., J. Ling, J. Plieske, S. Kollers, E. Ebmeyer, V. Korzun, et al. 2014. Genetic architecture of main effect QTL for heading date in European winter wheat. Front. Plant Sci. 5:217.