Population Structure in the Model Grass Brachypodium distachyon Is Highly Correlated with Flowering Differences across Broad

Published April 29, 2016

ORIGINAL RESEARCH

Population Structure in the Model Grass *Brachypodium distachyon* Is Highly Correlated with Flowering Differences across Broad Geographic Areas

Ludmila Tyler, Scott J. Lee, Nelson D. Young, Gregory A. Delulio, Elena Benavente, Michael Reagon, Jessica Sysopha, Riccardo M. Baldini, Angelo Troìa, Samuel P. Hazen, and Ana L. Caicedo*

Abstract

The small, annual grass *Brachypodium distachyon* (L.) Beauv., a close relative of wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.), is a powerful model system for cereals and bioenergy grasses. Genome-wide association studies (GWAS) of natural variation can elucidate the genetic basis of complex traits but have been so far limited in *B. distachyon* by the lack of large numbers of well-characterized and sufficiently diverse accessions. Here, we report on genotyping-by-sequencing (GBS) of 84 *B. distachyon*, seven *B. hybridum*, and three *B. stacei* accessions with diverse geographic origins including Albania, Armenia, Georgia, Italy, Spain, and Turkey. Over 90,000 high-quality single-nucleotide polymorphisms (SNPs) distributed across the Bd21 reference genome were identified. Our results confirm the hybrid nature of the *B. hybridum* genome, which appears as a mosaic of *B. distachyon*-like and *B. stacei*-like sequences. Analysis of more than 50,000 SNPs for the *B. distachyon* accessions revealed three distinct, genetically defined populations. Surprisingly, these genomic profiles are associated with differences in flowering time rather than with broad geographic origin. High levels of differentiation in loci associated with floral development support the differences in flowering phenology between *B. distachyon* populations. Genome-wide association studies combining genotypic and phenotypic data also suggest the presence of one or more photoperiodism, circadian clock, and vernalization genes in loci associated with flowering time variation within *B. distachyon* populations. Our characterization elucidates genes underlying population differences, expands the germplasm resources available for *Brachypodium*, and illustrates the feasibility and limitations of GWAS in this model grass.

Core Ideas

- Genotyping diverse *Brachypodium* accessions expands research tools for grasses.
- The *B. hybridum* genome is a mosaic of *B. distachyon*- and B. *stacei*-like sequences.
- Three distinct, genetically defined populations of *B. distachyon* were identified.
- Flowering time, more than geography, distinguishes *B. distachyon* populations.
- Results support the feasibility of genome-wide association studies in a model grass.

L. Tyler, Biochemistry and Molecular Biology Dep., Univ. of Massachusetts, Amherst, MA, 01003; S.J. Lee, Biology Dep., Univ. of Massachusetts, Amherst, MA, 01003 and Plant Biology Graduate Program, Univ. of Massachusetts, Amherst, MA, 01003; N.D. Young, S.P. Hazen, and A.L. Caicedo, Biology Dep., Univ. of Massachusetts, Amherst, MA, 01003; G.A. Delulio, Biochemistry and Molecular Biology Dep., Univ. of Massachusetts, Amherst, MA, 01003 and Plant Biology Graduate Program, Univ. of Massachusetts, Amherst, MA, 01003; E. Benavente, Dep. of Biotechnology, ETSIA, Technical Univ. of Madrid, 28040 Madrid, Spain; M. Reagon, Dep. of Biology, The Ohio State Univ.–Lima, Lima, Ohio, 45804; J. Sysopha, Biology Dep., Univ. of Massachusetts, Amherst, MA, 01003; R.M. Baldini, Dep. of Biology, Univ. of Florence, 50121 Florence, Italy; A. Troìa, Dipartimento STEBICEF, Sezione di Botanica ed Ecologia vegetale, Università di Palermo, Palermo, 90133, Italy. Received 25 Aug. 2015. Accepted 29 Dec. 2015. *Corresponding author (caicedo@bio.umass.edu).

Abbreviations: *F*-statistic, F_{ST}; GBS, genotyping-by-sequencing; GLM, general linear model; GO, gene ontology; GWAS, genome-wide association studies; LD, linkage disequilibrium; MLM, mixed linear model; MTA, marker–trait association; PCA, principal component analysis; PCR, polymerase chain reaction; SNP, single nucleotide polymorphism; SSR, simple-sequence repeat.

Published in Plant Genome
Volume 9. doi: 10.3835/plantgenome2015.08.0074

Studying inherited natural variation can elucidate the genetic basis of complex, multigenic traits. In particular, next-generation genotyping has enabled GWAS, which examine large, diverse populations for correlations between alleles at individual markers and phenotypes of interest. The GWAS approach has been successful in a number of plant species, including rice (*Oryza sativa* L.), maize (*Zea mays* L.), barley, the bioenergy grass *Miscanthus ×giganteus* J. M. Greef & Deuter ex Hodk. & Renvoize, and the dicotyledenous model *Arabidopsis thaliana* (L.) Heynh. (Atwell et al., 2010; Huang et al., 2010; Tian et al., 2011; Pasam et al., 2012; Slavov et al., 2014). With the increasing demand for plants as sources of food, feed, and fuel, there is a growing need for tools to improve our understanding of agronomically important traits. Whether the investigations are performed in crops or model species, the success of GWAS depends on the availability of extensive, diverse, and well-characterized germplasm collections.

Here, we expand the germplasm resources available for the model grass *B. distachyon*. *Brachypodium distachyon* was first proposed as a model research organism nearly 15 yr ago (Draper et al., 2001) and has rapidly developed into a powerful system with a wealth of research tools (Brkljacic et al., 2011; Mur et al., 2011; Girin et al., 2014). As a small, diploid annual with a compact, sequenced genome (International Brachypodium Initiative, 2010) and amenability to both crossing and genetic transformation (Vain et al., 2008; Vogel and Hill, 2008; Alves et al., 2009; Bragg et al., 2012), *B. distachyon* has many of the advantages of the preeminent plant model, *A. thaliana* (Koornneef and Meinke, 2010). However, unlike *A. thaliana*, *B. distachyon* is in the Pooideae subfamily of the grasses, together with wheat, oat (*Avena sativa* L.), and barley and is thus an especially valuable model for the temperate cereals (Mochida and Shinozaki, 2013). Because it has the type II cell walls typical of grasses, *B. distachyon* is also being used to investigate feedstock properties for bioenergy production (Gomez et al., 2008; Lee et al., 2012; Marriott et al., 2014; Tyler et al., 2014). Despite all these advantages, germplasm collections of *B. distachyon* remain limited. Although the native range of *B. distachyon* stretches across the Mediterranean region (Garvin et al., 2008; López-Alvarez et al., 2015), material from only two countries, Spain and Turkey, has been characterized in detail. This limitation has hindered assessment of structure and other population parameters in the species, as well as development of a diverse germplasm collection for GWAS.

Until recently, the diploid *B. distachyon*, with a haploid chromosome number of 5, was grouped in the same species with morphologically similar *Brachypodium* plants that have haploid chromosome numbers of 10 and 15. These latter two cytotypes are now considered separate species: *B. stacei* (n = 10) and *B. hybridum* (n = 15) (Catalán et al., 2012). Fluorescence and genomic in situ hybridization have elegantly demonstrated that *B. hybridum* is an allotetraploid arising from interbreeding between diploid, *B. distachyon*-like and *B. stacei*-like parents (Hasterok et al., 2004; Idziak et al., 2011; Catalán et al., 2012). Flow cytometry has shown that the genome of *B. hybridum* is approximately twice the size of the *B. stacei* or *B. distachyon* genome (Catalán et al., 2012). Furthermore, DNA barcoding and polymerase chain reaction (PCR)-based analysis of simple-sequence repeats (SSRs) have emerged as methods for *Brachypodium* species identification (Giraldo et al., 2012; López-Alvarez et al., 2012). While *B. stacei* and *B. hybridum* are useful for investigating speciation and ploidy, *B. distachyon* has remained the premier model system within this genus, and it is crucial to correctly distinguish the three species when developing germplasm resources for GWAS.

In this study, we undertook the collection and characterization of 54 new *Brachypodium* accessions, together with 40 previously described lines, originating from the northern arc of the Mediterranean region and the Middle East. Our collection includes novel material from Spain, Italy, Albania, Georgia, and Armenia. We performed reduced-representation sequencing to generate a large collection of polymorphic markers, which allowed us to both differentiate between species and elucidate the population structure within *B. distachyon*. Interestingly, we identified flowering time as a major factor correlated with population structure. Despite this correlation, and the occurrence of strong structure in the species, we were able to use flowering time as an example to illustrate the feasibility of GWAS and limits to its resolution in our extended germplasm panel.

Materials and Methods

Plant Materials and Growth Conditions

Between 2008 and 2012, we harvested seeds from individual plants in the field to develop five new lines from Albania, four from Armenia, 10 from Italy, and 16 from Spain. Seeds collected in the country of Georgia were used to generate 18 new lines; however, because seeds from more than one plant at each harvest site were pooled, lines from the same site in Georgia could be siblings. Also included in this study are two reference *B. hybridum* lines, ABR100 and ABR113; one reference *B. stacei* line, ABR114 (Draper et al., 2001; Hasterok et al., 2004); and five previously characterized inbred *B. distachyon* lines, Bd21, Bd21-3, Bd3-1, Bd30-1, and Bd1-1 (Vogel et al., 2006; Vogel and Hill, 2008; Schwartz et al., 2010; Ream et al., 2014). Eighteen previously characterized inbred *B. distachyon* lines from Turkey (Filiz et al., 2009; Vogel et al., 2009) and 14 additional accessions from Spain (Giraldo et al., 2012) were included as well. We derived a Ukrainian line by single-seed descent from material deposited at the USDA National Plant Germplasm System under stock number PI 639818. Although D. Garvin used the same stock to generate inbred line Bd29-1 (Ream et al., 2014), the Ukrainian line used in the current study (Ukr-Nvk1) and Bd29-1 are distinct. A 95th line, BsUPM_AL1.1 (Giraldo et al., 2012), which

had been identified as *B. stacei*, was also submitted for GBS. However, because the GBS results and chromosome counting indicated that the submitted sample was contaminated with DNA from a second accession, BsUPM_AL1.1 was removed from subsequent analyses. To maintain consistency with the previous nomenclature, while also promoting the clear identification of lines in this study, three-letter prefixes indicating geographic origin were combined with the accession name assigned by the original collector. A three-letter country code (e.g., Ita for Italy)—and in some cases, a three-letter locality code (e.g. Sic for Sicily)—was added to the beginning of each accession name. For consistency, prefixes were added even to the names of well-characterized reference lines; for example, Irq-Bd21 indicates Bd21, the line that originated in Iraq and was the source of the *B. distachyon* reference genome sequence (International Brachypodium Initiative, 2010). Supplemental Table S1 contains detailed information about the final 94 lines used.

To synchronize germination and promote flowering, seeds were cold treated before planting. Seeds were first sterilized by removing the lemmas, washing the seeds for 4 min in a solution of 15% bleach plus 0.1% Triton-X 100 (Sigma-Aldrich), and rinsing the seeds twice with sterile water. Seeds were placed on damp paper towels for stratification at 4°C before sowing.

For DNA extraction and seed bulking, seeds were sown in a 3:1 mixture of potting mix (#2, Conrad Fafard Inc.) and Turface particles (Profile Products). Plants were grown in a greenhouse with a 20-h photoperiod including supplemental lighting. Temperature varied somewhat with the outside temperature at our location (Amherst, MA, 42.392°N, 72.525°W). For phenotypic characterizations, seeds were sown in potting mix (Pro-Mix BX, Hummert International), and plants were grown in an environmentally controlled greenhouse under 16-h light vs. 8-h dark conditions, with day vs. night temperatures of 24 and 20°C, respectively.

Seeds for all accessions that proved amenable to bulking are being deposited into the USDA National Plant Germplasm System/Germplasm Resources Information Network (NPGS/GRIN, http://www.ars-grin.gov/npgs/).

DNA Extraction and Genotyping-by-Sequencing

DNA was extracted from leaf tissue of individual plants or inbred lines using a DNeasy Plant Mini Kit (Qiagen) with the following modifications: Buffer AP1 was heated to 65°C before use; a TE solution of 10 mM Tris, 0.1 mM ethylenediamine tetraacetic acid (EDTA), pH 8 was used for elution and was also heated to 65°C.

DNA concentration was quantified using a Qubit 2.0 fluorometer and Quanti-iT dsDNA HS Assay Kit (Invitrogen). The quality of the genomic DNA was confirmed by running samples on a 1% agarose gel for comparison to a λ-DNA *Hind*III size standard (New England Biolabs) and by digesting a subset of samples with *Hind*III (New England Biolabs).

The DNA samples were submitted to the Cornell University Institute for Genomic Diversity for GBS (Ithaca, NY), as described previously (Elshire et al., 2011). The GBS library with 96 unique barcodes represented one blank and 95 genomic DNA samples (Supplemental Table S1). Briefly, samples were digested with *Ape*KI, ligated to barcoding and common adaptors, and amplified via PCR. The resulting fragments were tested for size and sequenced in a single lane on an Illumina HiSeq 2000 using 100 nucleotide single-end reads. The mean number of reads obtained per individual was 2.1 million. Raw reads were submitted to the National Center for Biotechnology Information Short Read Archive under experiment SRX796779.

Filtering Sequencing Data for Single-Nucleotide Polymorphism Calling and Quality Control

Single-nucleotide polymorphism calling and initial filtering for the raw sequence reads were performed at Cornell University, as described by Elshire et al. (2011), using the GBS pipeline in TASSEL Version 3.0.138 (1 Nov. 2012). In summary, reads were checked for the presence of a barcode and at least 72 bases of good sequence, trimmed to 64 bases, and aligned to the Bd21 v1.0 reference genome sequence, designated Bd21-1 (http://www.brachypodium.org/gmod/genomic/accessions/1; file: Bdistachyon.MIPS_1_2.fa.gz, downloaded 8 Nov. 2012) (International Brachypodium Initiative, 2010). Sequence tags mapping to multiple genomic loci were discarded. The maximum number of good reads per lane was set to 4×10^8, the minimum sequence tag count to 5, and the minimum minor allele frequency to 0.001. Heterozygotes were called when each allele was supported by the minimum sequence tag count; loci with more than three alleles were discarded. Minimum site coverage was set to 0.8, that is, SNP data had to be present for at least 80% of the accessions for the SNP to be retained. This analysis yielded a filtered set of 192,974 SNPs in HapMap format.

Next, we discarded SNPs that mapped to the multiple, small, extrachromosomal contigs included in the Bd21 reference genome (International Brachypodium Initiative, 2010). We further filtered SNPs to remove those in or adjacent to mononucleotide repeats longer than 5 bases and SNPs with missing data (*N*) for >10% of the accessions. This additional filtering resulted in 99,872 high-quality SNPs for the full set of 94 *Brachypodium* lines characterized by GBS (Supplemental Data S1). However, our later analyses indicated that 10 of 94 (11%) of the accessions were *B. stacei* or *B. hybridum* (Supplemental Table S1). Repeating the filtering for the 84 putative *B. distachyon* accessions yielded 54,392 high-quality SNPs in a *B. distachyon*-only set (Supplemental Data S14).

Species Assignment via Sequencing of the *trnL-trnF* Locus

Polymerase chain reaction primers flanking a *trnL-trnF* intergenic region diagnostic for *Brachypodium* species identification (López-Alvarez et al., 2012) were used to

amplify part of the plastid *trnLF* locus from the same DNA samples that were sent for GBS. As previously described (Taberlet et al., 1991; López-Alvarez et al., 2012), 'c' (5′ CGAAATCGGTAGACGCTACG 3′) and 'f' (5′ ATTTGAACTGGTGACACGAG 3′) universal primers were used for PCR. Amplification was performed with ExTaq DNA polymerase (Clontech) under the following conditions: 98°C for 30 s; 35 cycles of 98°C for 10 s, 53°C for 30 s, 72°C for 1 min; and 72°C for 10 min. The forward 'c' primer was used for Sanger sequencing of the PCR products. The sequences were manually trimmed for quality and aligned in CLUSTAL W, with default settings, as implemented in MEGA5 (Tamura et al., 2011). A 741-nucleotide region of the resulting alignment was used to generate a neighbor-joining tree with bootstrap analysis, *p*-distance substitution, and pairwise deletion in MEGA5 (Felsenstein, 1985; Saitou and Nei, 1987; Nei and Kumar, 2000; Tamura et al., 2011). In a few cases, sequence at the *trnLF* locus was not determined, either because of a lack of clear PCR or sequencing results (Alb-AL1A, Alb-AL1D, and Geo-30i2) or because the species assignment had been previously established (Irq-Bd3-1 and Por-ABR113). For comparison, an ABR113 sequence (GenBank: JN187658.1) determined by Catalán et al. (2012) was included. New *trnLF* sequences were submitted to Genbank (accessions KU163145-KU163233; Supplemental Table S1).

Population Structure and Relatedness

The high-quality SNP sets were subjected to principal component analysis (PCA) using SmartPCA in the EIGENSOFT package (Patterson et al., 2006; Price et al., 2006). The PCA used the full set of 99,872 or 54,392 SNPs for all-*Brachypodium* or *B. distachyon*-only analyses, respectively. The number of output principal components was set to 10 without removing outliers.

Subsets of the high-quality SNPs were analyzed using version 2.3.3 of the STRUCTURE software (Pritchard et al., 2000; Falush et al., 2003, 2007; Hubisz et al., 2009) as installed on the Bioportal web service at the University of Oslo, Norway, in July of 2013. Sites with ambiguity codes were excluded from the analysis. To limit problems associated with closely linked loci and ease computational burden, SNPs 15 to 16 kb apart were randomly selected, resulting in 10,169 or 9907 loci for the all-*Brachypodium* or *B. distachyon*-only analyses, respectively. Burn in was set to 100,000 and the number of runs to 500,000. Because these *Brachypodium* species are primarily selfing and cannot achieve Hardy–Weinberg equilibrium, ploidy was set to 1. An admixture model was used to test the possibilities that $K = 1$ to $K = 15$ populations with three replicates. The ΔK method (Evanno et al., 2005) was used to determine the number of populations consistent with the SNP data.

Relationships among GBS haplotypes of *B. distachyon* and *B. stacei* were assessed using neighbor-joining and maximum likelihood analyses. Heterozygous calls were converted to missing data to decrease the impact of possible misalignment of paralogous regions. Both neighbor-joining and maximum likelihood RAxML (Stamatakis, 2014) trees, based on 93,801 SNPs in total, were generated using iPlant apps (Goff et al., 2011). The neighbor-joining tree was generated using a Kimura two-parameter nucleotide substitution model and 1000 bootstrap replicates to assess branch support. The maximum likelihood tree was generated using default settings and the general time-reversal plus gamma (GTR + G) model. Unrooted trees were visualized with Dendroscope (Huson et al., 2007).

Calculating Timing of Population Divergence

We used DIYABC version 2.0 (Cornuet et al., 2014), which applies approximate Bayesian computation (Beaumont et al., 2002), to infer the time of divergence for the two main *B. distachyon* populations. Sicilian accessions constituting PCA-defined Cluster 3A (see Results section) were excluded from this analysis to reduce the effects of population substructure. We used a subset of 10,000 randomly selected SNPs with a minor allele frequency cutoff of 5%. We assumed a simple demographic model, which simulated a population that split at time t with no subsequent gene flow between the resulting populations. Our model included three parameters: the sizes of the two populations and time of divergence. The analysis using 200,000 simulations was performed on a 64-core computer server at the University of Massachusetts–Amherst. Posterior estimates of the three parameters were determined using the logistic transformation option.

F-Statistic Analysis of Population Differentiation

We estimated SNP-specific F_{ST} between the two main *B. distachyon* populations and searched for outliers using the programs LOSITAN (Beaumont and Nichols, 1996; Antao et al., 2008) and BayeScan 2.1 (Foll and Gaggiotti, 2008). Because of their status as a separate clade, and to keep the analysis to two populations, Sicilian accessions in PCA-defined Cluster 3A were excluded. Sites with a minor allele frequency <5% were also removed. Heterozygous calls were recoded as missing data to decrease the impact of possible misalignment of paralogous regions; however, average F_{ST} values calculated with or without heterozygous calls were very similar. In total, 26,897 SNPs were included. For LOSITAN, 50,000 simulations were run using default parameters and both the neutral-mean-F_{ST} and force-mean-F_{ST} options. Loci outside the 95% confidence interval were considered outliers. For BayeScan, prior odds for the neutral model were set to 1, which assumes that the neutral model is as likely as a model with selection. False discovery rate was set to 0.05, and outlier detection was performed with an *R* function provided by BayeScan. The SNPs identified as outliers or as fixed by both programs were considered the best candidates for underlying population divergence. Putative fixed SNPs further filtered out loci that had any (rare) ambiguity code calls in the original dataset. We identified the *B. distachyon* gene closest to each candidate SNP

and then identified rice homologs from the top hit using BLASTx for protein similarity (Camacho et al., 2009). These rice homologs were examined for gene ontology (GO) term enrichment, as GO term annotation was insufficient for analyses in *B. distachyon*. We used the full MSU7.0 *Oryza sativa* nontransposable-element gene identifier as background for term enrichment analysis in agriGO (Du et al., 2010). Significance was evaluated using a hypergeometric statistical test with a Hochberg false discovery rate multiple comparison correction; the minimum number of mapping entries was set to two.

Calculating Linkage Disequilibrium

Decay of linkage disequilibrium (LD) with distance was calculated separately for the STRUCTURE-defined *B. distachyon* Populations 1 ($n = 57$) and 2 ($n = 23$), excluding Sicilian accessions in PCA Cluster 3A to avoid overestimation of LD resulting from population substructure as well as for the full set of *B. distachyon* ($n = 84$) using TASSEL 4.0 (Bradbury et al., 2007). Ambiguity codes were changed to *N*, and sites with more than 10% missing data were dropped. For the full set and Population 1, SNPs with a minor allele frequency of <5% were discarded. For Population 2, because of the smaller sample size, a minor allele frequency cutoff of 8.5% was used. Between 17,640 and 28,393 SNPs were used for the LD calculations. The LD was evaluated with a sliding window 200 SNPs wide. A mean LD r^2 value was taken from all LD measures in each 1 kb bucket. Plots were generated using the ggplot2 package in R (Wickham, 2009; R Development Core Team, 2014) and fitted with a generalized additive model curve. Further r^2 means were calculated for other genomic distances (Table 1).

Table 1. Analysis of linkage disequilibrium for the full set of 84 *Brachypodium distachyon* accessions and subsets corresponding to Populations 1 and 2.

	Full set		Population 1		Population 2	
Range	Measures	r^2	Measures	r^2	Measures	r^2
kb						
<1	101682	0.141	93890	0.126	90430	0.175
1–10	40004	0.327	24077	0.409	18151	0.408
10–20	38667	0.293	23915	0.360	16979	0.350
20–30	37868	0.273	23152	0.326	16366	0.331
30–50	72706	0.264	43151	0.301	31195	0.305
50–70	71140	0.248	42269	0.267	30496	0.284
70–100	106461	0.233	62861	0.236	45629	0.270
100–200	348034	0.213	202529	0.194	146346	0.250
200–300	343430	0.191	195570	0.151	142376	0.227
300–400	334274	0.178	186478	0.128	138778	0.217
400–500	335245	0.168	186152	0.113	138968	0.203
500–600	332148	0.164	183407	0.104	136280	0.197
500–700	330413	0.159	182499	0.097	136301	0.188
700–1000	958789	0.155	523570	0.089	402062	0.184
1000–5000	2164688	0.168	1924432	0.078	1923235	0.188

Flowering-Time Measurements

In Supplemental Data S10, the 63 accessions for which flowering time was quantified are listed. Seed availability permitting, three or four seeds per accession per treatment were surface sterilized as described above and placed at 4°C for 2 or 6 wk. Although some seeds—particularly those cold-treated for 6 wk—germinated in the cold, seedling growth was minimal at 4°C. Cold treatments were staggered so that seeds were all sown in the greenhouse on the same day. Pots were arranged in a randomized block design, and flats of pots were rotated once a week. Flowering time was measured, from August to January, as the number of days from sowing to initial emergence of the inflorescence, that is, when the awn of the top-most flower was first visible. This stage corresponds to the start of heading, stage 51 on the Biologische Bundesanstalt Bundessortenamt and Chemische Industrie scale (Hong et al., 2011), and is consistent with prior measurements of flowering time at stage 50 on the Zadoks scale (Ream et al., 2014). Natural light intensity varied over the course of our flowering-time experiment at our greenhouse location in Amherst, MA (approximately 42°23′30.9″ N, 72°31′55.1″ W). Supplemental lighting was used to maintain irradiance at or above 450 W m^{-2} for 16 h d^{-1}. For purposes of calculation, plants that had not flowered by the end of the experiment at 149 d were assigned a flowering time of 150 d.

Genome-Wide Association Studies Analysis of Flowering Time in *Brachypodium distachyon* Accessions

Genome-wide association studies analysis was performed using TASSEL 3.0 (Bradbury et al., 2007) for the 54,392 SNPS in the *B. distachyon*-only set, including individuals from Populations 1, 2, and 5 together with the flowering-time data from plants that had received 6 wk of vernalization as described above. We performed both general linear model (GLM; Bradbury et al., 2007) and mixed linear model (MLM; Zhang et al., 2010) analyses via TASSEL. We used a 5% minor allele frequency cutoff and verified that results were similar with a 10% cutoff. A kinship matrix, generated within TASSEL, was added as part of the MLM analysis to factor in relatedness between accessions (Yu et al., 2006), and a $Q + K$ model was used. For the GLM, 1000 permutations were performed. The MLM with the preset optimum level of compression was used; variance components were estimated after each marker. Quantile-quantile plots were compared to determine if MLM or GLM displayed a tighter adherence to expected *p*-values.

To reduce Type I error rates that result from testing with multiple SNPs, an adjusted significance threshold was calculated. Because of our small population size, we used a highly conservative method of dividing a traditional 0.05 *p*-value threshold by the total number of SNPs tested. This adjusted significance threshold was used as the genome-wide cutoff. The SNPs that crossed

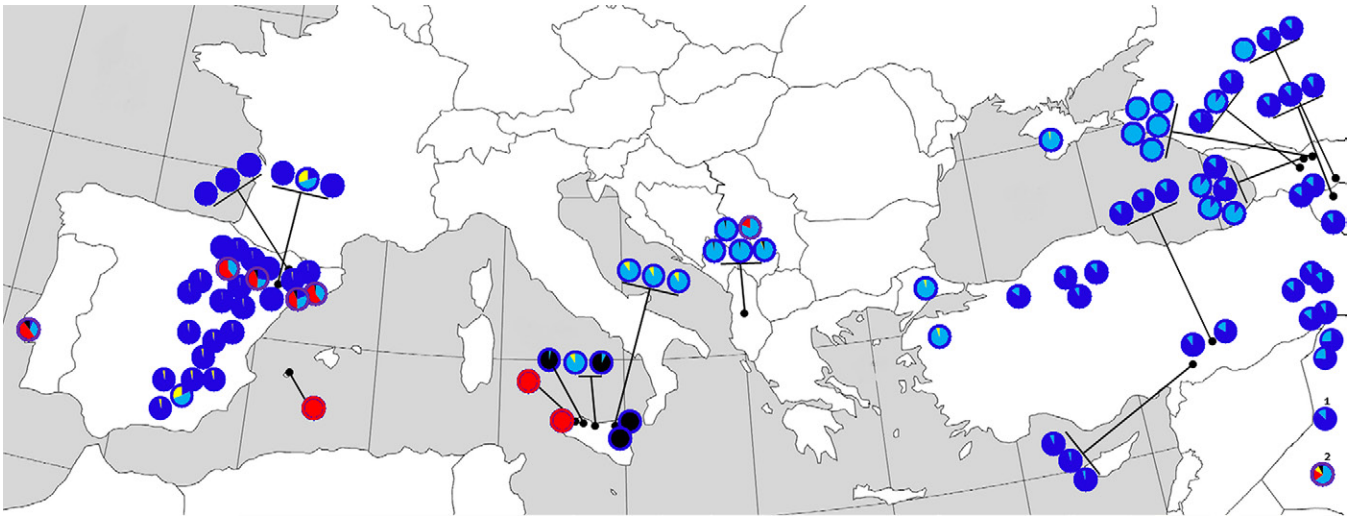

Fig. 1. Geographic distribution of *Brachypodium* accessions used in this study. Each dot represents an individual accession, with the color of the outer ring representing inferred species (*B. distachyon* in dark blue, *B. stacei* in red, and *B. hybridum* in purple). The inner pie chart represents the inferred ancestries of each accession as determined by STRUCTURE analysis (Population 1 in dark blue, Population 2 in cyan, Population 3 in red, Population 4 in yellow, and Population 5 in black). 1, the exact location of this collection site in Iraq is unknown; 2, the collection site is Kalafabad, Iran (not shown). Detailed collection site information is listed in Supplemental Table S1. Species assignments and STRUCTURE analysis are discussed in the main text.

this threshold were considered to represent significant marker–trait associations (MTAs) (Supplemental Table S4). To determine if candidate flowering-time regulators were within or near the significant MTAs, we identified genes 250 kb upstream of the left border and 250 kb downstream of the right border of each significant MTA within the *B. distachyon* v1.0 reference genome. These boundary genes were subsequently used to anchor MTA intervals to the *B. distachyon* v3.1 gene annotation. We explored the available annotation for all genes within these regions including for the v3.1 annotation; resulting candidates for regulating flowering time in *B. distachyon* are reported in Supplemental Table S4.

Results

Diverse Germplasm Was Characterized via Genotyping-by-Sequencing

Our study encompasses 94 *Brachypodium* accessions from a broad geographic range and includes reference lines for *B. hybridum* and *B. stacei*, as well as *B. distachyon* (Fig. 1; Supplemental Table S1). Many of the samples were newly collected and developed for this study and include countries for which *Brachypodium* germplasm has not yet been characterized such as Italy, Albania, Armenia, and Georgia. While some collection sites are represented by a single accession, at other sites, seeds were harvested from two to five different individuals. To maximize diversity, both newly harvested germplasm and established inbred lines were included. Names of accessions and collection sites are listed in Supplemental Table S1.

Genotyping-by-sequencing (Elshire et al., 2011) of these accessions yielded 99,872 high-quality SNPs for the full set of 94 *Brachypodium* lines characterized (Supplemental Data S1). These SNPs were fairly evenly distributed across the five nuclear chromosomes of *B. distachyon*, with one SNP approximately every 2.6 to 3.0 kb. However, as expected, the frequency of SNPs was reduced near centromeres; *Ape*KI, the restriction enzyme used to digest our genomic DNA samples, is partially methylation-sensitive and therefore unlikely to cut within regions of repetitive DNA.

As a control, Bd21, the source of the fully sequenced *B. distachyon* reference genome, was subjected to GBS with the other 93 accessions. Only 111 SNPs were identified between our Bd21 sample and the reference genome, suggesting a low likelihood of false-positive SNP calls. The few SNPs that were identified in our GBS analysis of Bd21 probably are due to a combination of sequencing errors either in the reference or our data and remaining trace heterozygosity in Bd21, as has been reported (Gordon et al., 2014). For the other accessions, GBS detected from 3380 to 42,345 clear differences from the reference (Supplemental Data S2).

The three known *B. stacei* lines had by far the highest number of differences compared with Bd21, that is, more than 40,000 SNPs per line. All other accessions had fewer than 18,500 SNPs compared to the Bd21 reference (Supplemental Data S2). Pairwise comparisons of SNP calls likewise emphasized the distinct nature of the *B. stacei* lines ABR114, Ita-Sic-QVL1, and Ita-Sic-VMO1. Although these lines were distinguished from each other by no more than 1200 SNPs, upward of 35,000 SNPs often separated *B. stacei* from other accessions. Exceptions to this general trend included the reference *B. hybridum* lines ABR100 and ABR113, which differed from ABR114 by 7015 and 7161 SNPs respectively (Supplemental Data S3).

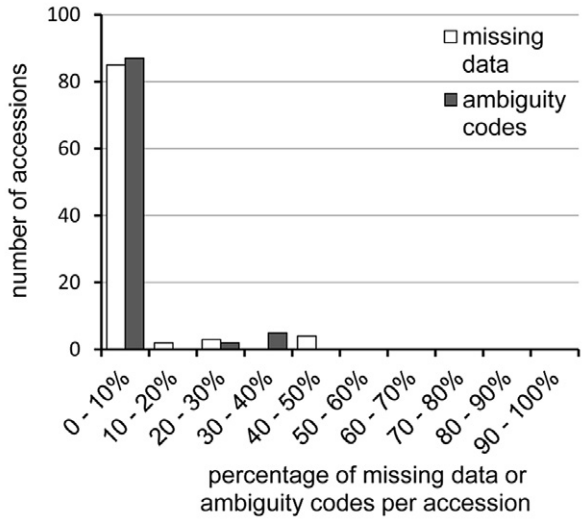

Fig. 2. Levels of missing data and apparent heterozygosity in genotyping-by-sequencing results for *Brachypodium* accessions. The number of accessions with a given percentage of single-nucleotide polymorphisms represented by missing data (alleles called as N's, white bars) or nucleotide ambiguity codes (indicating two alleles, gray bars) is shown.

We also determined the proportion of missing SNP information for each accession. For 85 out of 94 accessions, <10% of the SNPs corresponded to missing data, testifying to the overall high quality of GBS (Fig. 2). Five lines exhibited between 10 and 30% missing data, and four lines had 40 to 50% missing data (Fig. 2). These last four lines consisted of Bd3-1, an established *B. distachyon* reference line from Iraq, and three *B. stacei* lines: ABR114, Ita-Sic-QVL1, and Ita-Sic-VMO1. The appearance of Bd3-1 as an outlier may indicate a technical issue with this sample. The high proportion of missing data for the *B. stacei* lines, however, probably is due to the fact that SNPs were determined by mapping sequence reads to the *B. distachyon* Bd21 genome. It is likely that many *B. stacei* sequences were too divergent to meet the threshold for mapping to *B. distachyon* loci and were therefore considered to be missing; alternatively, some *B. distachyon* loci may be deleted in the *B. stacei* genome. In this context, the thousands of SNPs differentiating *B. stacei* lines from other accessions are even more striking (Supplemental Data S2, S3); SNPs with missing nucleotide information were excluded from these totals, suggesting that the GBS SNP calls represent a dramatic underestimate of the actual, genome-wide sequence differences.

In addition, we examined the level of heterozygosity in each accession (Fig. 2). The majority of accessions (87 out of 94, or 93%) had two alleles at <6% of the SNP loci; in fact, 89% of the accessions had <2% heterozygosity. This high level of homozygosity is consistent with previous characterizations of Turkish and Spanish germplasm using SSR markers as well as an extremely high frequency of inbreeding reported for *B. distachyon* (Vogel et al., 2009; Mur et al., 2011; Giraldo et al., 2012). In the seven remaining accessions in our study, heterozygosity was markedly higher, ranging from 21 to 38% (Fig. 2). Both of the *B. hybridum* reference lines, ABR100 and ABR114, had high levels of nucleotide ambiguity codes, leading us to hypothesize that the other five lines in this group (Alb-AL_1D and Spa-Nor-S1A, -S2C, -S7C, and -S10C) might also be *B. hybridum*

The presence of both *B. stacei*-like and *B. distachyon*-like sequences for homeologous loci in the allotretraploid could result in the appearance of heterozygosity for those loci. To confirm this possibility, we focused on SNPs for which all three *B. stacei* lines lacked sequence information. In this subset of 40,959 out of 99,872 high-quality SNPs, ambiguity codes were found for no more than 1.7% of the positions in any of the seven putative *B. hybridum*. This observation is consistent with *B. hybridum* merely appearing to be highly heterozygous because its genome is derived from divergent *B. stacei*-like and *B. distachyon*-like parents. Together, the results on levels of missing SNP data and heterozygosity (Fig. 2) were the first indication that the GBS data clearly distinguished between the three species (*B. stacei*, *B. hybridum*, and *B. distachyon*) at a genome-wide level.

Brachypodium hybridum Is Intermediate to *Brachypodium stacei* and *Brachypodium distachyon* in Population Structure Analyses

To elucidate the population structure in our sample, we investigated the complete set of 99,872 SNPs using the distance-based method of PCA (Patterson et al., 2006; Price et al., 2006). The 94 *Brachypodium* accessions formed three clusters well separated from each other along the first principal component, which explains the largest percentage of the variation in the data (Fig. 3). Because reference lines for *B. stacei*, *B. hybridum*, and *B. distachyon* separated into PCA Clusters 1, 2, and 3, respectively, we hypothesized that species identity explains the greatest amount of variation between the accessions. Interestingly, the putative *B. hybridum* cluster (Cluster 2) was centrally located between the *B. stacei* and *B. distachyon* clusters along the first principal component (Fig. 3). This result hints at the mixed parentage of the allotetraploid. In addition, the 84 accessions in Cluster 3 formed three subgroups separating along the second principal component, which explains the next largest percentage of the variation (Fig. 3). While Cluster 3A consisted of four Italian accessions, Clusters 3B and 3C were of mixed geographic origin; Albania, Georgia, Italy, Spain, Turkey, and Ukraine were represented in Cluster 3B and Armenia, Georgia, Iraq, Spain, and Turkey were represented in Cluster 3C. The PCA thus suggests that the accessions in Cluster 3 are all *B. distachyon* but that membership in the subclusters is determined by a factor other than geography. Accessions comprising each cluster, with the corresponding values for the first 10 principal components, are listed in Supplemental Data S4.

To further clarify the population structure, we analyzed the SNP data with STRUCTURE, a model-based

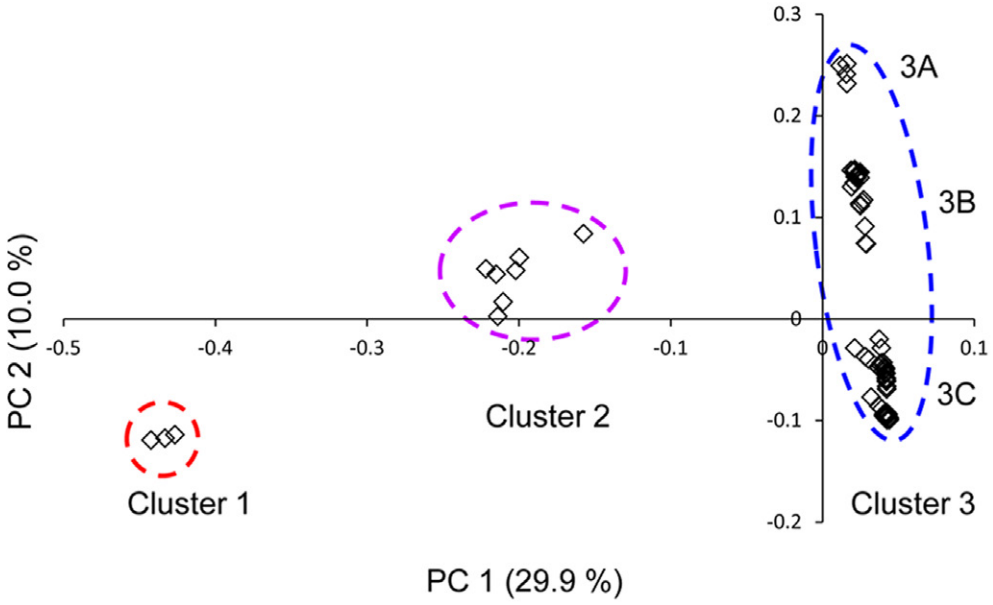

Fig. 3. Principal component analysis clusters the three species separately, with *Brachypodium hybridum* intermediate to *B. stacei* and *B. distachyon*. Principal component analysis was performed for 99,872 single-nucleotide polymorphisms using EIGENSOFT (Patterson et al., 2006; Price et al., 2006). The values of principal component 1 and 2 (PC 1 and PC 2, explaining 29.9 and 10.0% of the variation in the data set and represented by the *x*- and *y*-axes, respectively) were plotted for each accession. Each open diamond represents an individual accession. Because of overlap, not all diamonds are clearly visible. Cluster 1 contains known *B. stacei* lines exclusively. Cluster 2 includes the reference *B. hybridum* lines ABR113 and ABR100. Cluster 3 contains reference *B. distachyon* lines in two of the three subgroups: Bd1-1 in subgroup 3B, and Bd21, Bd21-3, and Bd3-1 in 3C. Subgroup 3A contains four Italian accessions.

clustering method (Pritchard et al., 2000; Falush et al., 2003; Evanno et al., 2005; Falush et al., 2007; Hubisz et al., 2009), and found a clear peak in ΔK for $K = 5$ populations (Supplemental Data S5). A striking feature of the inferred ancestry is that the putative *B. stacei* accessions appeared to comprise a single population (Population 3, color-coded red in Fig. 4; Supplemental Data S6). Moreover, this population was almost undetectable in the inferred ancestries of *B. distachyon* individuals. In *B. distachyon*, alleles from three populations— Population 1 (dark blue), Population 2 (cyan), or Population 5 (black)—tended to predominate (Fig. 4). Population 4 (yellow) contributed a relatively small proportion of alleles to some accessions. The seven lines known or hypothesized to be *B. hybridum* on the basis of the PCA all exhibited mixed ancestries derived to varying degrees from the *B. stacei*-like Population 3 plus *B. distachyon*-like populations (Fig. 4). Population 2, in particular, was prominently represented in the inferred ancestries of all the *B. hybridum* lines tested. This mosaic ancestry demonstrates, with the resolution achieved by thousands of SNPs distributed across the chromosomes, that the *B. hybridum* genome is a mixture of *B. stacei*-like and *B. distachyon*-like sequences.

Genotyping-by-Sequencing Yields Unambiguous Species Assignments Confirmed by Sequencing of the DNA Barcoding Locus *trnLF*

To confirm the distinctions between species observed with our GBS data, we employed the DNA barcoding method established by López-Alvarez and colleagues for the plastid *trnLF* locus (López-Alvarez et al., 2012). The noncoding *trnLF* region, located between the chloroplast *trnL* and *trnF* genes, is maternally inherited and distinct between *B. stacei* and *B. distachyon* (López-Alvarez et al., 2012). In most of the *B. hybridum* examined, the maternal parent appears to have been *B. stacei*-like (López-Alvarez et al., 2012). Therefore, the presence of a *B. stacei trnLF* sequence in a *Brachypodium* plant indicates that the species is not *B. distachyon*, but could be either *B. stacei* or *B. hybridum*. For a small percentage of *B. hybridum* individuals examined to date, the maternal parent seems to have been *B. distachyon*-like (López-Alvarez et al., 2012). Thus, identification of a *B. distachyon*-like *trnLF* sequence could be consistent with a species assignment of either *B. distachyon* or *B. hybridum*.

Figure 5 shows a neighbor-joining tree of *trnLF* sequences for accessions analyzed by GBS including reference *B. stacei*, *B. hybridum*, and *B. distachyon*. The nucleotide sequences clearly formed two separate clades each with 100% bootstrap support. The smaller clade consisted of *B. stacei*-like sequences and included the type *B. stacei* line ABR114 as well as the established *B. hybridum* lines ABR100 and ABR113. Also in this small clade were two Italian accessions (Ita-Sic-QVL1 and

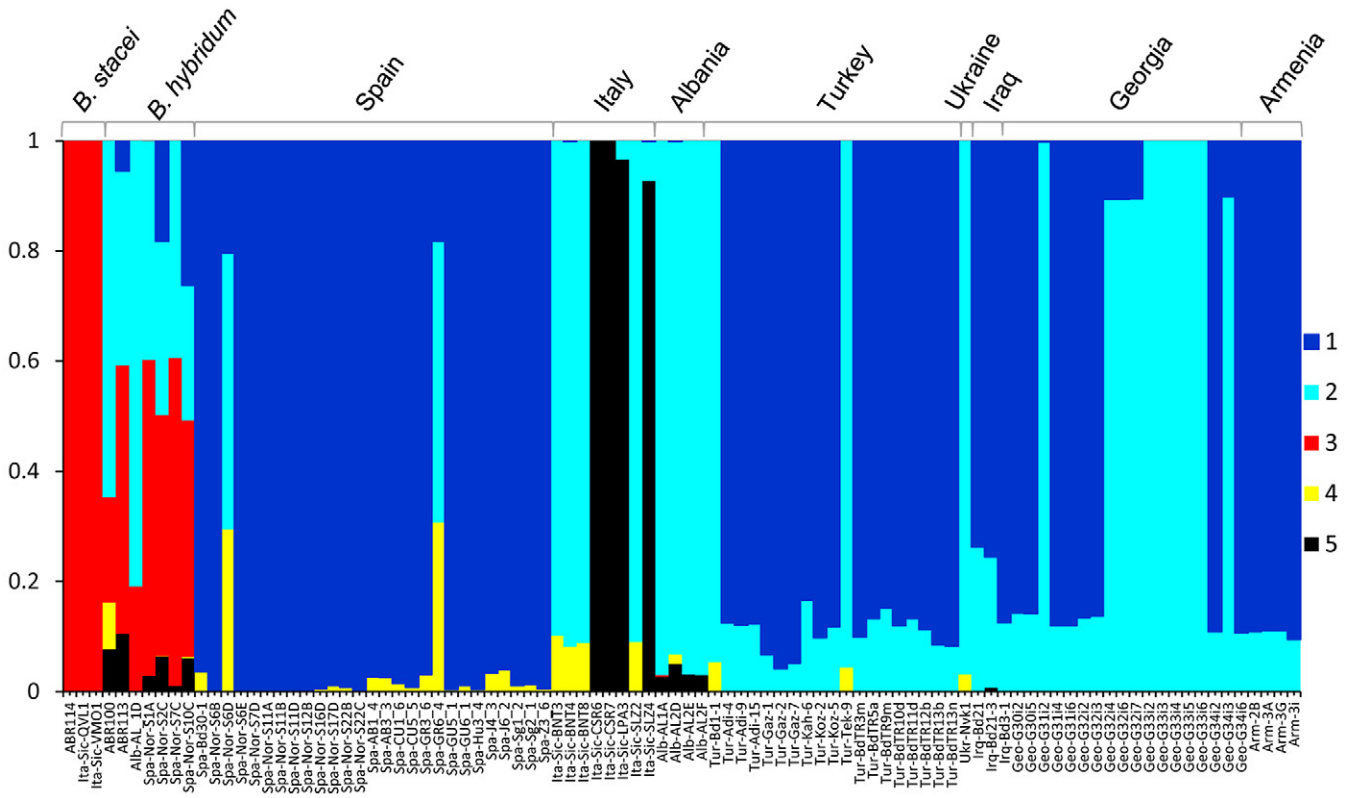

Fig. 4. STRUCTURE analysis reveals distinct genomic signatures for *Brachypodium stacei*, *B. hybridum*, and *B. distachyon*. The results of a representative replicate for $K = 5$ populations are shown. Each of the 94 accessions is represented by a vertical bar. Color coding along the *y*-axis indicates the proportion of the individual's ancestry inferred to be derived from Population 1 (dark blue), 2 (cyan), 3 (red), 4 (yellow), or 5 (black). Accessions are ordered from left to right with *B. stacei* first, *B. hybridum* second, and then *B. distachyon* accessions grouped by country of origin.

-VMO1) previously described as *B. stacei* (López-Alvarez et al., 2012) and four Spanish lines (Fig. 5). These four Spanish accessions—like the *B. hybridum* reference lines—were in PCA Cluster 2; their membership in the smaller *trnLF* clade is consistent with a *B. stacei*-like maternal parent for the hybrid species (Fig. 5). The larger clade in the tree consisted of *B. distachyon*-like sequences and included the reference *B. distachyon* lines Bd21, Bd1-1, and Bd21-3 in addition to the remaining accessions for which clear *trnLF* sequence was available (Fig. 5). Overall, the *trnLF* tree (Fig. 5; Supplemental Table S1) supported the idea that accessions separated along the first principal component (Fig. 3) based on species and that population structure captures species designations (Fig. 4). When other researchers had made previous species assignments, our combined PCA, *trnLF*, and STRUCTURE results agreed with the earlier assignments (Filiz et al., 2009; Vogel et al., 2009; López-Alvarez et al., 2012). Thus, our analysis of thousands of SNPs not only reliably distinguished between species but also provided a deeper understanding of the genomic composition of individuals within these species.

Geographic Origin Does Not Fully Explain the Population Structure Observed within *Brachypodium distachyon*

Within the 84 *B. distachyon* accessions, the three subgroupings revealed by PCA and STRUCTURE did not correlate broadly with geographic origin (Supplemental Data S4, S6). The divergent genomic signatures are particularly apparent in Fig. 4 (see also Fig. 1), in which the *B. distachyon* accessions are ordered by country of origin roughly from west to east. Notably, representatives of STRUCTURE Populations 1 and 2 are present in both the western and eastern edges of the geographic distribution (Fig. 4), although Population 1 does predominate in Spain. Additionally, in several instances, accessions harvested from the same location had dramatically different inferred ancestries. For example, Ita-Sic-SLZ2 and-SLZ4 originated from different individuals at the same site in Schilizzi near Palermo, Sicily (Supplemental Table S1); however, STRUCTURE Populations 2 and 5 represented 91 and 92% of the inferred ancestry in Ita-Sic-SLZ2 and Ita-Sic-SLZ4, respectively. Similarly, Geo-G31i2, -G31i4, and -G31i6 all originated from a single site in southeastern Georgia but exhibited alleles primarily from STRUCTURE Population 2 in Geo-G31i2 and Population 1 in Geo-G31i4 and -G31i6. Other accessions with the same geographic origin but contrasting genomic profiles

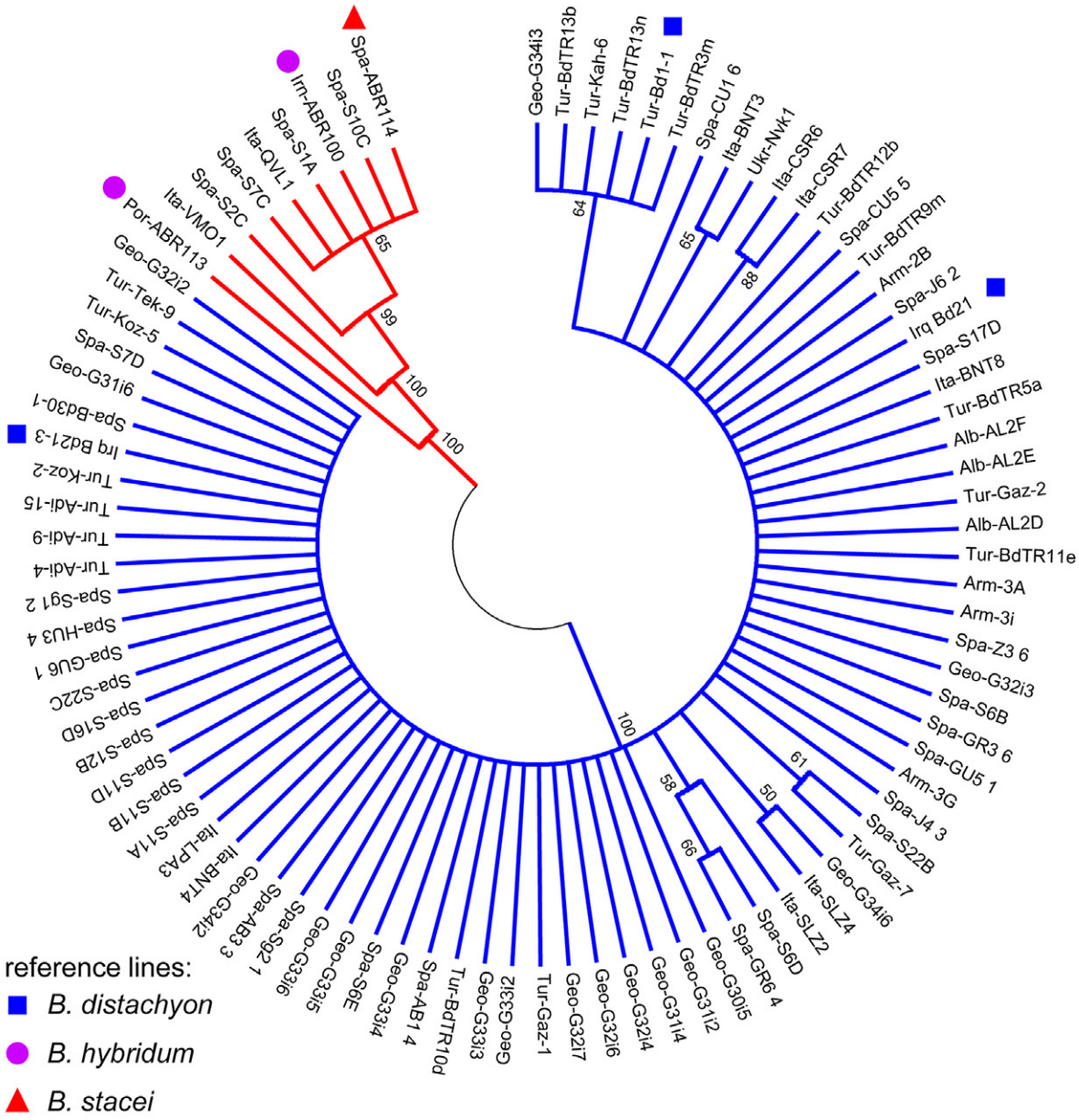

Fig. 5. Sequences for the plastid *trnLF* locus distinguish *Brachypodium distachyon* from *B. stacei* and *B. hybridum*. This neighbor-joining tree—generated in MEGA5 (Tamura et al., 2011)—is based on nucleotide sequences from the region between the chloroplast *trnL* and *trnF* genes for accessions analyzed by genotyping-by-sequencing. The tree is a consensus from 500 bootstrap replicates. Numbers indicate bootstrap support, and only branches with >50% support are shown. Blue lines indicate *B. distachyon*-like sequences, while red lines indicate *B. stacei*-like sequences. Blue squares mark the reference *B. distachyon* lines Bd21, Bd1-1, and Bd21-3; purple circles mark the reference *B. hybridum* lines ABR113 and ABR100; and a red triangle marks the reference *B. stacei* line ABR114.

included Geo-G32i2, -G32i4, -G32i6, and -G32i7; Geo-G34i2, -G34i3, and -G34i6; and Spa-Nor-S6B, -S6D, and -S6E (Fig. 4; Supplemental Data S6).

A neighbor-joining phylogenetic tree constructed for *B. stacei* and *B. distachyon* with the full SNP set revealed local geographic structure within major *B. distachyon* groups, but no geographic structure between major groups (Supplemental Fig. S1). The two major clades were monophyletic with high support and were consistent with the division of most *B. distachyon* accessions into PCA Clusters 3B and 3C or STRUCTURE Populations 2 and 1. Tree analysis further supported the distinction of PCA Cluster 3A or STRUCTURE Population 5 as a separate clade of Italian accessions more closely related to STRUCTURE Population 2 (Supplemental Fig. S1). Major clades tended to represent multiple geographic origins, though some of the subclades represented single regions such as Albania, Italy, northern Spain, and Turkey, indicating local relatedness among many *B. distachyon* samples (Supplemental Fig. S1).

To investigate whether species identity was obscuring the true population structure within *B. distachyon*, we excluded *B. stacei* and *B. hybridum* accessions and repeated the PCA using the 54,392 SNPs polymorphic

for *B. distachyon*. This PCA separated the *B. distachyon* accessions into three well-defined groups for the first two principal components (Supplemental Data S7), consistent with the all-*Brachypodium* analysis: the four Italian accessions Ita-Sic-CSR6, -CSR7, -LPA3, and -SLZ4 formed a distinct cluster (3A); 23 accessions from six countries comprised a second cluster (3B); and 57 *B. distachyon* accessions from five countries comprised the third cluster (3C; Supplemental Data S7). Only along the third vs. fifth principal components did accessions begin to form clusters consistent with geographic origin. Turkish accessions separated from Spanish ones along the third principal component, while Albanian and Georgian accessions separated along the fifth principal component (Supplemental Data S7). Thus, although evidence of relatedness as a result of similar geographic origin was apparent, geography was not the primary, defining feature of the population structure.

We also analyzed 9907 SNPs selected from the *B. distachyon*-only set with STRUCTURE. In this case, ΔK exhibited a peak for two populations, and PCA-defined subgroups 3A and 3B were combined, consistent with their sister-clade designation in the phylogenetic trees. However, subgroups 3A and 3B remained clearly separated from 3C and included accessions from multiple countries as above (Supplemental Data S8, S9).

Flowering Time Differences Are Largely Consistent with the *Brachypodium distachyon* Population Structure

Observing *B. distachyon* clusters of mixed geographic origin prompted us to look for another factor that might more fully explain population structure. Although limited phenotypic information was available for most of the accessions, the reference Bd lines were known to have contrasting flowering times. Notably, Bd21, Bd21-3, and Bd3-1—all of which derived most of their ancestry from the all-*Brachypodium* STRUCTURE Population 1 (dark blue)—are consistently early flowering, while Bd1-1, assigned to Population 2 (cyan), is late-flowering (Vogel et al., 2009; Schwartz et al., 2010; Ream et al., 2014). In addition, whereas many Turkish accessions, including Adi-4 and BdTR9m, flower relatively early, a few, such as Tek-9, are extremely late flowering (Filiz et al., 2009; Vogel et al., 2009). Bd1-1 and Tek-9 were the only Turkish lines in our collection assigned primarily to Population 2 (cyan) (Fig. 4). Interestingly, our line Ukr-Nvk1, with alleles predominantly from Population 2, originated from the same seed stock as Bd29-1, which has been characterized as delayed flowering similar to Bd1-1 (Ream et al., 2014). Moreover, as plants were growing for tissue and seed harvesting, we noticed that *B. distachyon* accessions with either Population 2 (cyan) or Population 5 (black) comprising most of the inferred ancestry tended to flower late. To determine whether this trend held, we quantified the flowering time for three-fourths of the *B. distachyon* accessions.

At the time of testing, sufficient seed was available for 48 of 57 (84%) of the Population 1 accessions (dark blue), 12 of 23 (52%) of the Population 2 accessions (cyan), and three of four (75%) of the Population 3 accessions (black). Three or four seeds for each accession were cold treated for 2 or 6 wk and then sown in soil and grown with a photoperiod of 16 h. Figure 6 shows the resulting flowering times. Supplemental Data S10 includes the full data set.

As has been reported previously (Vogel et al., 2009; Schwartz et al., 2010; Ream et al., 2014; Tyler et al., 2014), extended cold treatment tended to trigger earlier flowering in most accessions (Fig. 6A, 6C; Supplemental Data S10). For example, Bd21-3 cold treated for 6 wk flowered 31.2 d earlier than with the 2-wk cold treatment. Dramatic variation in flowering times between accessions was also observed (Fig. 6B, 6C), with Bd3-1 flowering at 18 d and several late-flowering accessions, such as Ita-Sic-SLZ2 and Tek-9, remaining vegetative over the course of the experiment even after 6 wk of cold treatment (Fig. 6; Supplemental Data S10). Because a day length of 20 h can further promote flowering (Ream et al., 2014), the 16-h photoperiod used for this study likely accentuated differences in vernalization responsiveness.

To determine whether flowering time was correlated with the observed population structure, we averaged the flowering times of accessions in PCA Cluster 3C–STRUCTURE Population 1 and compared them with flowering times of accessions in Cluster 3B–Population 2 for the 6-wk cold treatment (Fig. 6D). Cluster 3A–Population 5 was omitted because of its small sample size. Strikingly, although the variability within each cluster was high, Population 1 accessions had an average flowering time of 58.2 d compared with 122.2 d for Population 2 accessions (Fig. 6B). A single-factor analysis of variance (ANOVA) indicated that this difference between populations was statistically significant at $P = 7.63 \times 10^{-6}$. However, outliers occurred; for example, the Armenian accessions Arm-3A, -3G, and -3i as well as the Georgian accession Geo-G34i6, all of which were assigned to Population 1 (blue), required 130 d or more to flower (Supplemental Data S10). The earliest-flowering Population 2 accession was Spa-Nor-S6D. With an average flowering time of 27.7 d (±2.5 d SD), Spa-Nor-S6D flowered even earlier than many Population 1 accessions. Interestingly, two other accessions, Spa-Nor-S6B and -S6E, which originated from the same site as Spa-Nor-S6D, were relatively early flowering and had Population 1-type genomic signatures. In summary, in spite of the intriguing exceptions, flowering time differences are a major distinction between the main *B. distachyon* populations.

Many Single-Nucleotide Polymorphism Are Highly Differentiated or Fixed between *Brachypodium distachyon* Populations 1 and 2

We examined the degree of divergence between the two principal *B. distachyon* populations with *F*-statistic (F_{ST}) analyses using the programs LOSITAN and BayeScan

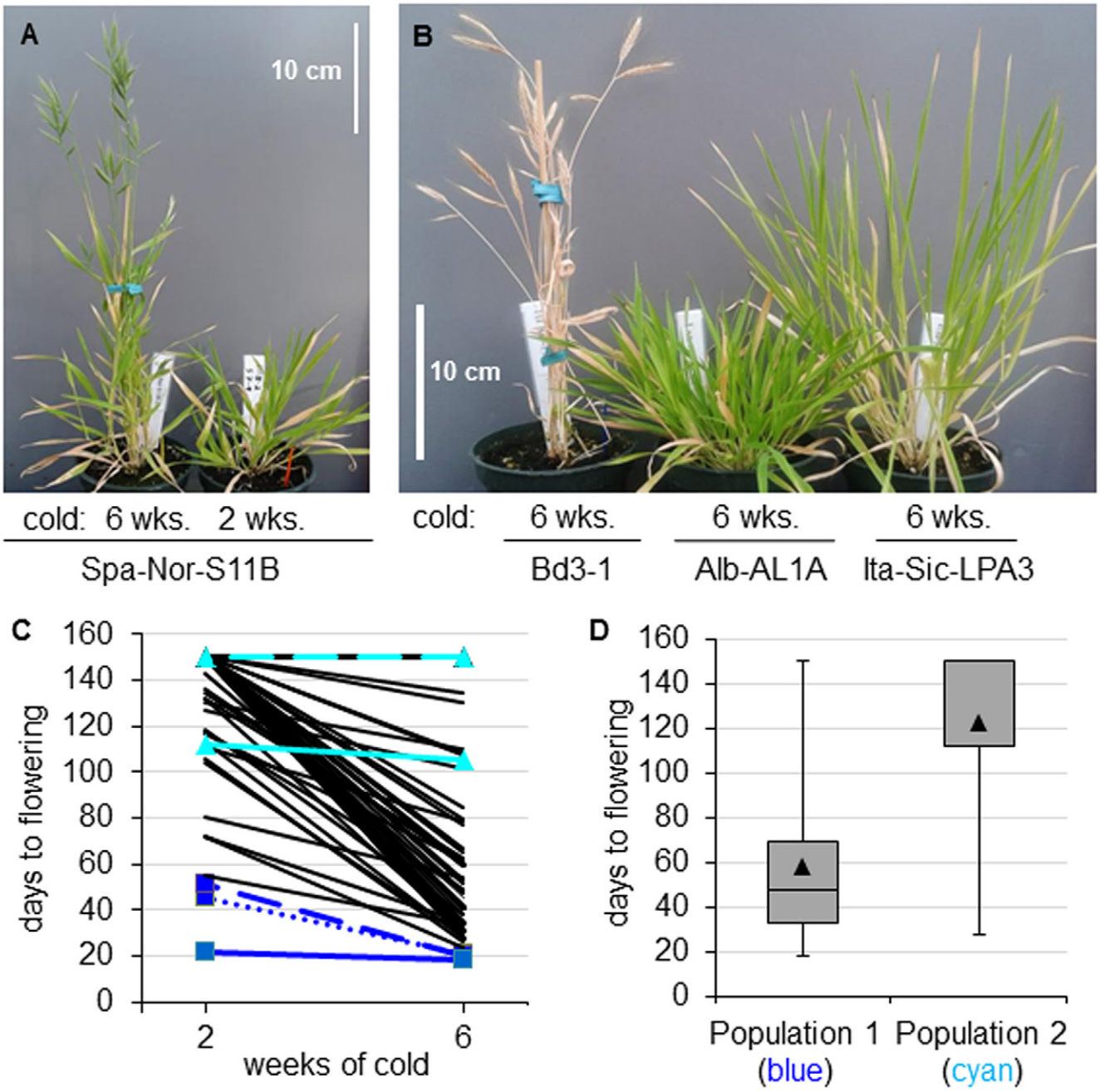

Fig. 6. Vernalization responsiveness and flowering time vary between accessions. Flowering time was measured after seeds had been cold treated for 2 or 6 wk. (a) Plants of the Population 1 accession Spa-Nor-S11B were photographed after 8 wk of growth; the plants on the left and right experienced 6 and 2 wk of cold treatment, respectively, before sowing. (b) Plants that had been cold treated for 6 wk as seeds were photographed after 9 wk of growth; Bd3-1 is from Population 1 (blue), Alb-AL1A from Population 2 (cyan), and Ita-Sic-LPA3 from Population 5 (black). The scale bars in (a) and (b) represent 10 cm. (c) Flowering time is shown as the average number of days required for flowering for two to four plants per accession per treatment. Plants that had not flowered by the end of the experiment were assigned a flowering time of 150 d. Most accessions are represented by black lines; select accessions are highlighted as follows: Irq-Bd3–1, solid, dark blue line; Irq-Bd21, dotted, blue line; Irq-Bd21-3, dashed, blue line; Tur-Bd1-1, solid, cyan line; and Tur-Tek-9, dashed, cyan line. (d) On average, after 6 wk of cold treatment, Population 1 accessions flowered earlier than Population 2 accessions. The black triangles indicate the average flowering time in days. The shaded boxes represent the second and third quartiles, while the whiskers indicate the maximum and minimum flowering times for accessions in the two populations. The median is represented by a horizontal black line across the shaded box for Population 1 and is at 150 d for Population 2. Sample sizes are 48 and 10 accessions for Populations 1 and 2, respectively.

(Beaumont and Nichols, 1996; Antao et al., 2008; Foll and Gaggiotti, 2008). The estimated F_{ST} for the dataset was high for both programs (0.51 and 0.62, respectively), indicating extensive differentiation across the genome between populations of *B. distachyon*. In total, LOSITAN identified 632 loci as fixed (F_{ST} = 1) and 2772 loci as high F_{ST} outliers; BayeScan identified 686 loci, including fixed loci, as outliers (Supplemental Data S11, S12). We focused on the 218 fixed loci between Populations 1 and 2, which occur on all five *B. distachyon* chromosomes. In addition, 14 SNPs on chromosomes 1, 2, and 3 were identified by both programs as highly differentiated (Supplemental Data S13). Besides indicating the types of loci that differentiate the two populations, this set of SNPs could be useful for genotyping new germplasm to determine likely membership in Population 1 vs. 2.

We determined the identity of the *B. distachyon* gene nearest each of these highly divergent or fixed SNPs and then identified the closest rice homologous gene (Supplemental Table S2). Gene ontology term analysis of the rice orthologs revealed 20 biological process terms significantly overrepresented in our dataset (Table 2; Supplemental Table S3; Supplemental Fig. S2). Remarkably, the majority of these terms are related to reproduction and developmental processes including flower development (Table 2; Supplemental Fig. S2). Inspection of the individual annotations for the rice and closest *A. thaliana* orthologs likewise revealed various genes associated with flowering, organ development, and meristem maintenance including a DELLA-domain GRAS transcription factor gene, a gibberellin-receptor-like *GID1L2* gene, *AINTEGUMENTA*, *APETALA1*, *FERTILIZATION-INDEPENDENT ENDOSPERM*, *AUXIN RESPONSE FACTOR10*, *BRASSINOSTEROID INSENTITIVE1*, *HECATE*, *LATERAL SUPPRESOR*, *PERIANTHIA*, *SEPALLATA1*, *REPLUMLESS*, *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE9*, *LOST MERISTEMS3*, *FILAMENTOUS FLOWER*, and *TARGET OF MONOPTEROS6* (Ikeda et al., 2001; Greb et al., 2003; Malcomber and Kellogg, 2005; Nole-Wilson and Krizek, 2006; Gremski et al., 2007; Schwarz et al., 2008; Mosquna et al., 2009; Ding and Friml, 2010; Li et al., 2010; Schlereth et al., 2010; Schulze et al., 2010; Maier et al., 2011; Jiang et al., 2014; Pabón-Mora et al., 2014) (Supplemental Table S2). These results suggest that genes involved in flowering and developmental programming are among the most highly differentiated between the two *B. distachyon* populations. It is possible that differences in at least some of these genes may play a role in the flowering time differences observed between populations under our common garden conditions.

Given the high levels of differentiation between *B. distachyon* populations, we attempted to determine when they had diverged from each other using DIYABC (Cornuet et al., 2008, 2014). The median of the posterior distribution for the estimated time of divergence was 5920 generations (95% confidence interval: 1290–9620) (Supplemental Fig. S3). Because *B. distachyon* is annual, with likely one generation per year, this result indicates that divergence between *B. distachyon* populations occurred <10,000 yr ago. In contrast, the *B. stacei* and *B. distachyon* lineages split an estimated 9.9 and 4.3 million yr ago from the common ancestor, respectively (López-Alvarez et al., 2012). The relatively recent divergence observed here for *B. distachyon* populations, after the end of Pleistocene glaciations, suggests that differences in flowering phenology evolved rapidly among *B. distachyon* populations.

Linkage Disequilibrium Declines More Quickly in Population 1 than Population 2

Having detected broad flowering time differences and highly differentiated SNPs between Populations 1 and 2, we next examined LD to assess the resolution of association mapping in these populations of *B. distachyon*. Linkage disequilibrium describes the degree of association between alleles at different loci; rapid decay of LD within a short genomic distance is favorable for GWAS, as it decreases the number of markers associated with a phenotype of interest and allows much finer resolution when mapping. We examined the decay of LD with distance in Population 1, Population 2, and the full set of *B. distachyon* accessions. Figure 7 shows how LD measures change with distance; Table 1 summarizes the results of this analysis in bins ranging from 10 to 5000 kb. Linkage disequilibrium decay is not particularly rapid in *B. distachyon*; in the full set, average r^2 levels of 0.1 are not reached by 5 Mb. This is in contrast to maize, where r^2 levels of 0.1 are reached on average by 2 to 5 kb (Yan et al., 2009), or *A. thaliana*, where levels of 0.2 are reached by about 10 kb (Kim et al., 2007). The highly selfing nature of *B. distachyon* (Vogel et al., 2009) can lead to high levels of LD; however, LD levels in *B. distachyon* can also be affected by the strong population structure within the species. When Populations 1 and 2 are examined separately, LD decay occurs more rapidly. In particular, LD decay is most rapid in Population 1, where r^2 levels close to 0.1 are reached by ~500 kb, and half-maximum values are reached by 100 to 200 kb, similar to what is observed in some varieties of cultivated rice (Mather et al., 2007; Xu et al., 2012). In contrast, LD decay is quite slow in Population 2, with levels of r^2 of 0.1 again not reached by 5 Mb and half-maximum values reached by 400 to 500 kb. Extensive LD within this population may be due to the moderate sample size (n = 23). Together, these results suggest that GWAS will have greater resolution if mapping is performed within Population 1. Nevertheless, increased mapping resolution could be achieved if the LD rate of decay were increased, perhaps through sampling of more accessions, even in this population.

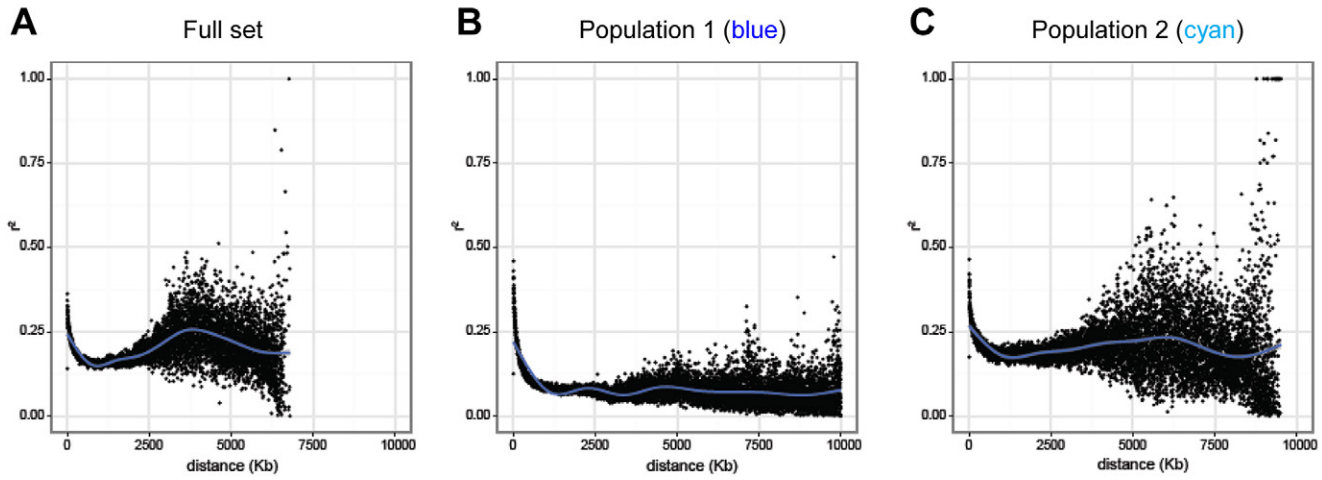

Fig. 7. Linkage disequilibrium decays more quickly in *Brachypodium distachyon* Population 1 than Population 2. Linkage disequilibrium (LD) plots for the complete set of 84 *B. distachyon* accessions, 57 Population 1 accessions, and 23 Population 2 accessions are shown in (a), (b), and (c), respectively. Linkage disequilibrium was not analyzed separately for the four Population 5 accessions because of the small sample size. Pairwise LD values (r^2) are plotted against physical distance. Trend lines represent plot data fitted with a generalized additive model curve.

Table 2. Significantly overrepresented biological process gene ontology (GO) terms among genes near highly divergent SNP markers.

GO term	Description	Input count	Genome total count	*p*-value	False discovery rate
GO:0006807	Nitrogen compound metabolic process	43	3898	0.00046	0.0077
GO:0022414	Reproductive process	15	732	0.00013	0.0077
GO:0000003	Reproduction	22	1449	0.00028	0.0077
GO:0006139	Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	43	3898	0.00046	0.0077
GO:0048856	Anatomical structure development	19	1250	0.00074	0.0099
GO:0009058	Biosynthetic process	53	5311	0.00093	0.01
GO:0009908	Flower development	10	469	0.0013	0.011
GO:0048608	Reproductive structure development	10	474	0.0015	0.011
GO:0003006	Reproductive developmental process	10	474	0.0015	0.011
GO:0030154	Cell differentiation	10	534	0.0034	0.021
GO:0048869	Cellular developmental process	10	534	0.0034	0.021
GO:0032501	Multicellular organismal process	29	2636	0.0048	0.027
GO:0009653	Anatomical structure morphogenesis	13	886	0.0068	0.034
GO:0007275	Multicellular organismal development	28	2591	0.0071	0.034
GO:0009791	Post-embryonic development	18	1441	0.0079	0.035
GO:0016049	Cell growth	8	465	0.013	0.045
GO:0090066	Regulation of anatomical structure size	8	465	0.013	0.045
GO:0032502	Developmental process	29	2806	0.011	0.045
GO:0008361	Regulation of cell size	8	465	0.013	0.045

Genome-Wide Association Mapping Reveals Significant Associations for Flowering Time after Six Weeks of Vernalization

For many of the accessions tested, vernalization of 2 wk was insufficient to trigger flowering before the experiment ended (Supplemental Data S10). Therefore, to test the performance of GWAS in *B. distachyon*, we searched for significant MTAs between our high-density SNP data and time of flowering for 60 accessions after 6 wk of vernalization. We initially attempted GWAS using Population 1 only because of the relatively low LD within this group and to limit the effects of population structure. After correcting for multiple tests, no significant MTAs were identified within this small sample of accessions. Accordingly, we used the complete set of phenotyped accessions despite the risk of false negatives as a result of an association of flowering time with population structure (Brachi et al., 2011). Comparison of quantile–quantile plots demonstrated that the MLM displayed a tighter adherence to expected *P*-values than the GLM, though even with the MLM, *P*-values were still somewhat inflated (Supplemental Fig. S4). Focusing on results from the MLM analysis, we used a conservative adjusted significance threshold by dividing the 0.5 significance threshold by the number of SNPs tested, resulting in a genome-wide significance cutoff of 1.62×10^{-6}. We identified nine distinct regions of MTAs comprised of one or more significantly associated SNPs (Fig. 8). Four of the nine MTAs consisted of a single SNP, while the remaining MTAs contained multiple SNPs (Supplemental Table S4). Clear phenotypic differences were observed between individuals carrying alternative alleles at each MTA (Supplemental Fig. S5). Thus, these nine MTAs represent genomic loci that may significantly influence flowering time following 6 wk of vernalization.

The genetics of flowering time has been extensively studied in numerous species, including grasses, and 163 predicted flowering-time genes have been described in *B. distachyon* (Higgins et al., 2010; Ream et al., 2014). We therefore compared locations of these *B. distachyon* genes with MTAs (Fig. 8; Supplemental Table S4). As a conservative cutoff for distance, based on the decay of LD in Population 1, we chose 250 kb. In five cases, candidate genes lay within 250 kb of a significant MTA or

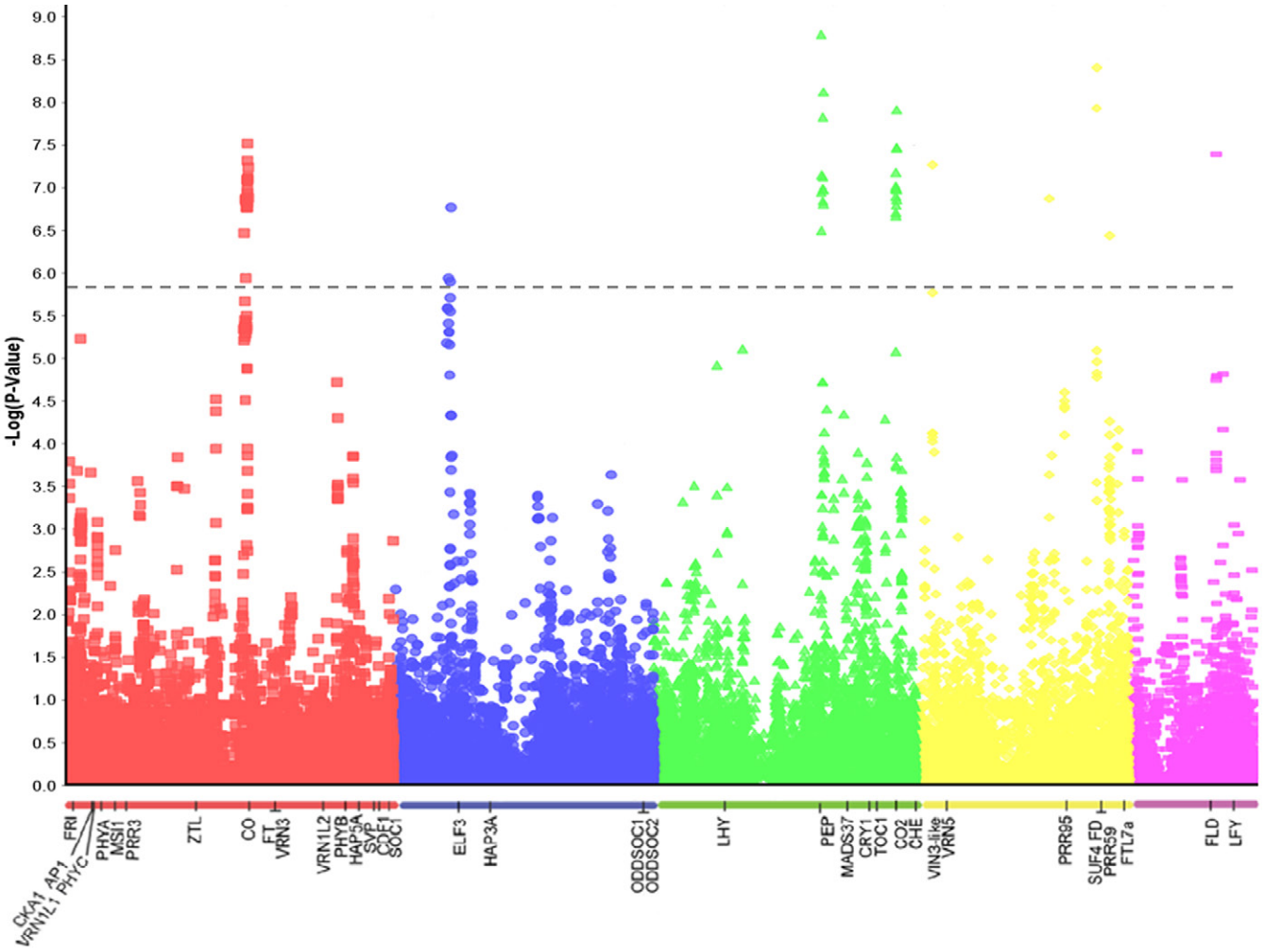

Fig. 8. Manhattan plot showing single-nucleotide polymorphism marker associations with flowering time following 6 wk of vernalization across the five chromosomes of *Brachypodium distachyon*. The chromosomal locations of *B. distachyon* homologs of previously characterized flowering time, circadian rhythm, and photoperiod genes are indicated. A horizontal gray dashed line represents the adjusted genome-wide significance threshold.

associated interval. For example, *BdCONSTANS1* (*CO1*), a homolog of *A. thaliana CONSTANS,* is 27 kb outside the MTA interval on chromosome 1. In *A. thaliana*, CO expression in the dark activates *FLOWERING LOCUS T* (*FT*), which encodes the mobile florigen signal that initiates the floral transition (Kardailsky et al., 1999; Samach et al., 2000; Jaeger and Wigge, 2007), and a rice gene closely related to *CO*, *Hd1/photoperiod sensitive1,* influences flowering time differentially under short and long days (Yano et al., 2000). A *B. distachyon* homolog of the *A. thaliana* circadian clock gene *EARLY FLOWERING3* (*ELF3*) lies within the boundaries of the MTA interval on chromosome 2. This gene was originally characterized in a screen for early flowering *A. thaliana* mutants (Zagotta et al., 1992). In barley, mutations in *Praematurum-a*, the barley homolog of *ELF3*, result in similar early flowering phenotypes (Zakhrabekova et al., 2012). On chromosome 3, one of the two MTAs is 150 kb from a gene orthologous to *A. thaliana GROWTH FACTOR REGULATOR 7* (*GFR7*). Although *GFR* genes have not been shown to directly impact flowering time, they have been implicated in floral organ differentiation and meristem patterning (Pajoro et al., 2014). Two of the four MTAs on chromosome 4 are located <150 kb from *BdPSEUDO-RESPONSE REGULATOR59* (*BdPRR59*) and *BdPRR95*, homologs of the *A. thaliana* circadian clock genes *PRR5* and *PRR9* (Nakamichi et al., 2005). While the role of *PRR*s in grasses is less well understood than in *A. thaliana*, in wheat, a 2 kb deletion upstream of the *PRR* family member *Ppd-D1* resulted in early flowering in both short and long days (Beales et al., 2007).

The remaining four MTAs—one on chromosome 3, two on chromosome 4, and one on chromosome 5—are more than 250 kb from obvious a priori candidates. These MTAs may correspond to novel, previously undetected, flowering-time-regulatory genes, or, given the slow LD decay in our *B. distachyon* sample, may be associated with more distant candidate genes.

Because the *B. distachyon* genome is compact, many potential flowering-time genes have been described, and we used a relatively large (250 kb) cutoff for loci associated with an MTA, we considered whether obtaining five

MTAs close to flowering-time candidate genes could have occurred by chance. In a random sample of 10,000 of our GBS markers, 1738 markers lie 149 kb or less from a candidate flowering-time gene listed in Higgins et al., 2010 and Ream et al., 2014. Also, 149 kb represents the greatest distance we observed for the five MTAs we considered to be close to flowering-time candidates. This suggests there is a 17.4% chance of any MTA falling close to a flowering-time gene whether or not there is a true association. In a sample of nine MTAs, we thus would have expected 1.56 to be within 149 kb of a flowering-time gene by chance. With five MTAs at distances of $\leq$149 kb from a flowering-time candidate, our sample does seem enriched for described flowering-time loci. Though more work is needed to confirm involvement of these genes in flowering-time variation in *B. distachyon*, our GWAS analysis has provided some fruitful candidates for exploration.

Discussion

By identifying SNPs adjacent to restriction enzyme sites, GBS captured a fraction of the genetic diversity in this germplasm collection. High-coverage, whole-genome resequencing would undoubtedly uncover an even greater richness of genetic variation including insertions, deletions, and rearrangements. For instance, resequencing of six core *B. distachyon* lines revealed upward of 1.6 million SNPs between the late-flowering line Bd1-1 and the early flowering line Bd21 (Gordon et al., 2014). In comparison, our GBS analyses detected approximately 12,000 SNPs between Bd1-1 and Bd21. The cost-effectiveness of reduced-representation sequencing, however, allowed us to survey genetic diversity across a greater number of accessions. The SNP information yielded by this broad approach made it possible to distinguish the genomic signatures of three *Brachypodium* species, characterize population structure within *B. distachyon* across its native range, and identify candidate loci potentially controlling flowering time.

In multiple analyses (PCA, STRUCTURE results, and phylogenetic trees), *B. stacei*, *B. hybridum*, and *B. distachyon* were clearly distinct (Fig. 3, 4; Supplemental Fig. S1). Because SNP calling involved mapping sequences to the *B. distachyon* reference genome, this distinction was also evident on examining levels of missing SNP data and heterozygosity (Fig. 2). A practical application of these findings is the identification of numerous candidate SNPs for use as markers in PCR-based species assignments; examination of those SNP loci for which sequence information was present for all three *B. stacei* accessions showed that, in many cases, the *B. stacei* and *B. distachyon* accessions carried contrasting nucleotides, while the seven *B. hybridum* accessions uniformly showed evidence of both *B. stacei*-like and *B. distachyon*-like alleles (Supplemental Data S1, S14). Although additional *B. stacei* and *B. hybridum* accessions must be tested to confirm the broad utility of each marker, subsets drawn from each chromosome could be used as cleaved, amplified polymorphic sequence markers (Konieczny and Ausubel, 1993) to rapidly assign species membership.

This study elucidates the differences between species across the nuclear genome at an average resolution of <5 kb. Our work thus reinforces and refines the three-species concept first revealed by fluorescence in situ hybridization and examination of organellar and nuclear loci (Hasterok et al., 2004; Idziak et al., 2011; Catalán et al., 2012; Giraldo et al., 2012; López-Alvarez et al., 2012). The PCA, summarizing information from nearly 100,000 SNPs, placed *B. hybridum* intermediate to *B. stacei* and *B. distachyon* along the first principal component (Fig. 3); thus, species identity is, genome-wide, the primary distinguishing factor for the 94 accessions in our collection. Additionally, STRUCTURE analysis showed the extent to which *B. hybridum* genomes are a mixture of alleles derived from *B. stacei*- and *B. distachyon*-like populations (Fig. 4). The diversity of individual ancestries inferred for different *B. hybridum* accessions (Fig. 4) is consistent with multiple origins of this allotetraploid species, as proposed by the research groups of P. Catalán and colleagues (López-Alvarez et al., 2012). The origination of these *B. hybridum* lines from Portugal, Spain, Albania, and Iran further supports the idea of multiple hybridization events. Strikingly, the *B. hybridum* accessions surveyed here all carried ancestry from *B. distachyon* Population 2, perhaps indicating a greater proclivity or opportunity for members of this population to cross with *B. stacei*. It will be interesting to determine whether a contribution from Population 2 is consistently present when larger numbers of *B. hybridum* accessions are examined.

The presence of only three *B. stacei* accessions in our study reflects both the relative scarcity of established *B. stacei* in *Brachypodium* collections, as well as our intentional focus on *B. distachyon* as a model. The five *B. hybridum* accessions newly identified in our study were initially presumed to be *B. distachyon* based on their morphology but clearly cluster with the reference *B. hybridum* lines in principal component and STRUCTURE analyses (Fig. 3, 4). These results emphasize the difficulty of making accurate species assignments for individual *Brachypodium* accessions based solely on morphological traits. While each species might, on average, be phenotypically distinct for a particular trait, the intraspecific variation is large (Tyler et al., 2014), and the phenotypic and geographic ranges can overlap (Mur et al., 2011; Catalán et al., 2012).

With regard to the 84 accessions confirmed to be *B. distachyon*, the data collected provide additional information about key reference lines and, importantly, allow the characterization of new germplasm. Our success in expanding the known genetic diversity for *B. distachyon* is illustrated by our identification of over 54,000 SNPs for accessions spanning the native range of this model grass. This total is more than three times the number of SNPs detected in a GBS analysis of Turkish accessions (Dell'Acqua et al., 2014). Interestingly, the inferred ancestries of Bd21 and Bd21-3 contained a relatively high

proportion of Population 2 in comparison to Bd3-1 and most other early-flowering lines (Fig. 4; Supplemental Data S6). This observation reinforces the utility of Bd3-1 as a mapping partner for crosses with Bd21 or Bd21-3. The Bd21 and Bd3-1 lines have already been used to generate a recombinant inbred population, which has proven to be a powerful tool in investigating the genetic basis of traits such as pathogen resistance (Cui et al., 2012). Our results could indicate additional promising candidates for high-coverage genome resequencing, generating recombinants, and other resource-intensive studies.

While Turkish inbred lines genotyped with 43 SSR markers had separated into distinct phylogenetic clades largely consistent with flowering time differences (Vogel et al., 2009), our GBS data showed that this distinction held across the native range of *B. distachyon* from Spain in the west to Georgia in the east. In terms of evolutionary history, the structuring of *B. distachyon* into at least two highly diverged genetic populations is intriguing, especially because the two main populations occupy similar geographic ranges and because the divergence seems to be relatively recent. In some of our samples, individuals belonging to different populations were found in the same locality or close by. This result indicates there is little gene flow between populations, consistent with high levels of self-fertilization in *B. distachyon* (Vogel et al., 2009). Although geographic subdivisions are evident within our main populations, the broad population structure is more strongly associated with what appears as flowering time in our common garden experiment. However, it is not intuitive why accessions could have different flowering times and yet thrive in the same locality subject to the same environmental conditions. We note, though, that the differences in flowering time observed under our greenhouse conditions do not necessarily represent true flowering times in the wild. Accessions that flower at different times in our experiments may flower concurrently in natural populations, despite divergent genetic backgrounds, as a response to local environmental conditions that act as cues for flowering. For example, the length of vernalization in many localities may exceed the lengths of time tested in our greenhouse. In fact, because most of our collections were made at a single time point, it is possible that all individuals at a single locality flower at similar times in the wild, regardless of genetic background. Nevertheless, our F_{ST} analyses strongly support the existence of some reproductive-phase morphological or developmental differences between the two main *B. distachyon* populations. Intriguingly, many genes close to SNPs highly differentiated between populations seem to be involved in the autonomous flowering pathway, suggesting that observed flowering time differences might be related to developmental phenotypes.

None of the fixed or high F_{ST} loci (Supplemental Table S2) between populations overlapped with MTAs detected by GWAS (Supplemental Table S4). This is expected and suggests, first, that our tests for association controlled well for confounding population structure (Brachi et al., 2011) and, second, that phenotypic diversity between populations and within populations may involve different flowering pathways. Polymorphic loci underlying flowering differences between populations are likely to display high F_{ST} and are best captured through such population analyses. That our GWAS yielded nine significant associations indicates that flowering time variation within populations is still sufficient to permit mapping of loci responsible for these differences. Additional investigation will be required to determine if any casual polymorphisms exist within the five a priori candidate genes or potential novel regulators.

Whereas GWAS has recently been used in *B. distachyon* to identify genetic loci associated with environmental variables (Dell'Acqua et al., 2014), to our knowledge, our findings are the first to illustrate the feasibility of linking genotype to a developmental phenotype in *B. distachyon*. Examination of our MTAs for known flowering-time, circadian-rhythm, and photoperiod genes revealed candidate genes underlying five of the nine MTAs. Our identification of four MTAs far from any previously described candidate genes suggests that further study is needed to better understand regulation of flowering time in the grasses. Our estimates of LD decay in *B. distachyon* indicate that improvements to GWAS in this species can be achieved through expansion of characterized germplasm, in particular from Population 1, which may decrease the size of genomic intervals at which LD decays significantly.

Conclusions

As demonstrated for flowering time in *B. distachyon*, GWAS can detect large effects even in small populations. On the other hand, detecting significant MTAs of small-effect or rare variants requires large populations (Atwell et al., 2010). Further expanding *B. distachyon* germplasm collections is therefore crucial for increased power in future GWAS. Because flowering time is a major factor separating populations, harvesting multiple times throughout the year at a single location could enhance the diversity of the germplasm collected and increase the likelihood that accessions most useful for GWAS—early-flowering accessions from STRUCTURE-defined Population 1—would be recovered. As well as capturing additional diversity, the availability of new *B. distachyon* accessions would increase the resolution of GWAS and thus greatly facilitate the rapid identification of candidate genes for functional testing in this model system.

Supplemental Material

Supplemental Tables S1–S4, Supplemental Data files S1–S15 (including SNP genotypes in Supplemental Data S1 and S15), and Supplemental Fig. S1–S5 are available online:

Supplemental Table S1. Summary information for accessions used in this study.

Supplemental Table S2. Annotation of genes near highly divergent SNPs.

Supplemental Table S3. Enrichment of GO terms for genes near highly divergent SNPs.

Supplemental Table S4. Significant marker-trait association intervals as determined by mixed linear model analysis for flowering time after 6 wk of vernalization.

Supplemental Data S1. High-quality SNPs for 94 *Brachypodium* accessions.

Supplemental Data S2. Number of SNPs compared to the Bd21 reference.

Supplemental Data S3. Number of SNPs in pairwise comparisons.

Supplemental Data S4. Principal component values for 94 *Brachypodium* accessions.

Supplemental Data S5. Delta *K* results for 94 *Brachypodium* accessions.

Supplemental Data S6. Inferred individual ancestry for 94 *Brachypodium* accessions.

Supplemental Data S7. Principal component values for 84 *B. distachyon* accessions.

Supplemental Data S8. Delta *K* results for 84 *B. distachyon* accessions.

Supplemental Data S9. Inferred individual ancestry for 84 *B. distachyon* accessions.

Supplemental Data S10. *B. distachyon* flowering times after cold treatment.

Supplemental Data S11. LOSITAN results for highly differentiated SNPs.

Supplemental Data S12. BayeScan results for highly differentiated SNPs.

Supplemental Data S13. SNPs highly differentiated between *B. distachyon* populations.

Supplemental Data S14. SNPs highly differentiated between *Brachypodium* species.

Supplemental Data S15. High-quality SNPs for 84 *B. distachyon* accessions.

Supplemental Fig. S1. Neighbor-joining tree for *B. distachyon*.

Supplemental Fig. S2. AgriGo graphic of GO terms for genes near highly divergent SNPs.

Supplemental Fig. S3. Timing of the *B. distachyon* population split.

Supplemental Fig. S4. Quantile-Quantile plots for general linear and mixed linear models for flowering time of *B. distachyon* accessions after 6 wk of vernalization.

Supplemental Fig. S5. Box plots of flowering time following 6 wk of vernalization for each significant MTA based on allele.

Acknowledgments

The authors thank the University of Oslo, Norway, for the Bioportal service used for STRUCTURE analysis. We are also grateful to Sherin Perera and Zhongyun Huang (University of Massachusetts–Amherst) for technical assistance and help with a bioinformatic analysis, respectively. The authors state they have no conflicts of interest to declare.

This material is based on work supported by the National Institute of Food and Agriculture (NIFA), USDA, the Massachusetts Agricultural Experiment Station, and the Biology Department of the University of Massachusetts–Amherst under project number MAS00419 to A.L.C. and S.P.H., and the Office of Science (BER) Department of Energy Grant DE-SC0006621 to S.P.H. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the USDA, NIFA, or other funding agencies.

References

Alves, S.C., B. Worland, V. Thole, J.W. Snape, M.W. Bevan, and P. Vain. 2009. A protocol for Agrobacterium-mediated transformation of *Brachypodium distachyon* community standard line Bd21. Nat. Protoc. 4:638–649. doi:10.1038/nprot.2009.30

Antao, T., A. Lopes, R.J. Lopes, A. Beja-Pereira, and G. Luikart. 2008. LOSITAN: A workbench to detect molecular adaptation based on a F_{st}-outlier method. BMC Bioinformatics 9:323. doi:10.1186/1471-2105-9-323

Atwell, S., Y.S. Huang, B.J. Vilhjálmsson, G. Willems, M. Horton, Y. Li, et al. 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature 465:627–631. doi:10.1038/nature08800

Beales, J., A. Turner, S. Griffiths, J.W. Snape, and D.A. Laurie. 2007. A pseudo-response regulator is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 115:721–733. doi:10.1007/s00122-007-0603-4

Beaumont, M.A., and R.A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. Lond. B Biol. Sci. 263:1619–1626. doi:10.1098/rspb.1996.0237

Beaumont, M.A., W. Zhang, and D.J. Balding. 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025–2035.

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.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:10.1093/bioinformatics/btm308

Bragg, J.N., L. Tyler, and J.P. Vogel. 2012. Brachypodium. In: C. Kole, C.P. Joshi, and D.R. Shonnard, editors, Handbook of bioenergy crop plants. Taylor and Francis, Boca Raton, FL. p. 593–618.

Brkljacic, J., E. Grotewold, R. Scholl, T. Mockler, D.F. Garvin, P. Vain, et al. 2011. Brachypodium as a model for the grasses: Today and the future. Plant Physiol. 157:3–13. doi:10.1104/pp.111.179531

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T.L. Madden. 2009. BLAST+: Architecture and applications. BMC Bioinformatics 10:421. doi:10.1186/1471-2105-10-421

Catalán, P., J. Müller, R. Hasterok, G. Jenkins, L.A.J. Mur, T. Langdon, A. Betekhtin, D. Siwinska, M. Pimentel, and D. López-Alvarez. 2012. Evolution and taxonomic split of the model grass *Brachypodium distachyon*. Ann. Bot. (Lond.) 109:385–405. doi:10.1093/aob/mcr294

Cornuet, J.M., P. Pudlo, J. Veyssier, A. Dehne-Garcia, M. Gautier, R. Leblois, J.M. Marin, and A. Estoup. 2014. DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. Bioinformatics 30:1187–1189. doi:10.1093/bioinformatics/btt763

Cornuet, J.M., F. Santos, M.A. Beaumont, C.P. Robert, J.M. Marin, D.J. Balding, T. Guillemaud, and A. Estoup. 2008. Inferring population history with DIY ABC: A user-friendly approach to approximate Bayesian computation. Bioinformatics 24:2713–2719. doi:10.1093/bioinformatics/btn514

Cui, Y., M.Y. Lee, N. Huo, J. Bragg, L. Yan, C. Yuan, C. Li, S.J. Holditch, J. Xie, M.C. Luo, et al. 2012. Fine mapping of the *Bsr1* barley stripe mosaic virus resistance gene in the model grass *Brachypodium distachyon*. PLoS ONE 7:E38333. doi:10.1371/journal.pone.0038333

Dell'Acqua, M., A. Zuccolo, M. Tuna, L. Gianfranceschi, and M.E. Pè. 2014. Targeting environmental adaptation in the monocot model *Brachypodium distachyon*: A multi-faceted approach. BMC Genomics 15:801. doi:10.1186/1471-2164-15-801

Ding, Z., and J. Friml. 2010. Auxin regulates distal stem cell differentiation in Arabidopsis roots. Proc. Natl. Acad. Sci. USA 107:12046–12051. doi:10.1073/pnas.1000672107

Draper, J., L.A. Mur, G. Jenkins, G.C. Ghosh-Biswas, P. Bablak, R. Hasterok, and A.P. Routledge. 2001. *Brachypodium distachyon*: A new model system for functional genomics in grasses. Plant Physiol. 127:1539–1555. doi:10.1104/pp.010196

Du, Z., X. Zhou, Y. Ling, Z. Zhang, and Z. Su. 2010. agriGO: A GO analysis toolkit for the agricultural community. Nucleic Acids Res. 38:W64–W70. doi:10.1093/nar/gkq310

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:E19379. doi:10.1371/journal.pone.0019379

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: A simulation study. Mol. Ecol. 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Falush, D., M. Stephens, and J.K. Pritchard. 2007. Inference of population structure using multilocus genotype data: Dominant markers and null alleles. Mol. Ecol. Notes 7:574–578. doi:10.1111/j.1471-8286.2007.01758.x

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791. doi:10.2307/2408678

Filiz, E., B.S. Ozdemir, F. Budak, J.P. Vogel, M. Tuna, and H. Budak. 2009. Molecular, morphological, and cytological analysis of diverse *Brachypodium distachyon* inbred lines. Genome 52:876–890. doi:10.1139/g09-062

Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian Perspective. Genetics 180:977–993. doi:10.1534/genetics.108.092221

Garvin, D.F., Y.Q. Gu, R. Hasterok, S.P. Hazen, G. Jenkins, T.C. Mockler, L.A.J. Mur, and J.P. Vogel. 2008. Development of genetic and genomic research resources for *Brachypodium distachyon*, a new model system for grass crop research RID B-3176-2009. Crop Sci. 48:S69–S84. doi:10.2135/cropsci2007.06.0332tpg

Giraldo, P., M. Rodríguez-Quijano, J.F. Vázquez, J.M. Carrillo, and E. Benavente. 2012. Validation of microsatellite markers for cytotype discrimination in the model grass *Brachypodium distachyon*. Genome 55:523–527. doi:10.1139/g2012-039

Girin, T., L.C. David, C. Chardin, R. Sibout, A. Krapp, S. Ferrario-Méry, and F. Daniel-Vedele. 2014. Brachypodium: A promising hub between model species and cereals. J. Exp. Bot. 65:5683–5696. doi:10.1093/jxb/eru376

Goff, S.A., M. Vaughn, S. McKay, E. Lyons, A.E. Stapleton, D. Gessler, N. Matasci, L. Wang, M. Hanlon, A. Lenards, et al. 2011. The iPlant collaborative: Cyberinfrastructure for plant biology. Front. Plant Sci. 2:34.

Gomez, L.D., J.K. Bristow, E.R. Statham, and S.J. McQueen-Mason. 2008. Analysis of saccharification in *Brachypodium distachyon* stems under mild conditions of hydrolysis. Biotechnol. Biofuels 1:15. doi:10.1186/1754-6834-1-15

Gordon, S.P., H. Priest, D.L. Des Marais, W. Schackwitz, M. Figueroa, J. Martin, et al. 2014. Genome diversity in *Brachypodium distachyon*: Deep sequencing of highly diverse inbred lines. Plant J. 79:361–374. doi:10.1111/tpj.12569

Greb, T., O. Clarenz, E. Schäfer, D. Müller, R. Herrero, G. Schmitz, and K. Theres. 2003. Molecular analysis of the *LATERAL SUPPRESSOR* gene in Arabidopsis reveals a conserved control mechanism for axillary meristem formation. Genes Dev. 17:1175–1187. doi:10.1101/gad.260703

Gremski, K., G. Ditta, and M.F. Yanofsky. 2007. The *HECATE* genes regulate female reproductive tract development in *Arabidopsis thaliana*. Development 134:3593–3601. doi:10.1242/dev.011510

Hasterok, R., J. Draper, and G. Jenkins. 2004. Laying the cytotaxonomic foundations of a new model grass, *Brachypodium distachyon* (L.) Beauv. Chromosome Res. 12:397–403. doi:10.1023/B:CHRO.0000034130.35983.99

Higgins, J.A., P.C. Bailey, and D.A. Laurie. 2010. Comparative genomics of flowering time pathways using *Brachypodium distachyon* as a model for the temperate grasses. PLoS ONE 5:E10065. doi:10.1371/journal.pone.0010065

Hong, S.Y., J.H. Park, S.H. Cho, M.S. Yang, and C.M. Park. 2011. Phenological growth stages of *Brachypodium distachyon*: Codification and description. Weed Res. 51:612–620. doi:10.1111/j.1365-3180.2011.00877.x

Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, C. Zhu, T. Lu, Z. Zhang, et al. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. 42:961–967. doi:10.1038/ng.695

Hubisz, M.J., D. Falush, M. Stephens, and J.K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. Mol. Ecol. Resour. 9:1322–1332. doi:10.1111/j.1755-0998.2009.02591.x

Huson, D.H., D.C. Richter, C. Rausch, T. Dezulian, M. Franz, and R. Rupp. 2007. Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics 8:460. doi:10.1186/1471-2105-8-460

Idziak, D., A. Betekhtin, E. Wolny, K. Lesniewska, J. Wright, M. Febrer, M.W. Bevan, G. Jenkins, and R. Hasterok. 2011. Painting the chromosomes of Brachypodium-current status and future prospects. Chromosoma 120:469–479. doi:10.1007/s00412-011-0326-9

Ikeda, A., M. Ueguchi-Tanaka, Y. Sonoda, H. Kitano, M. Koshioka, Y. Futsuhara, M. Matsuoka, and J. Yamaguchi. 2001. *slender* rice, a constitutive gibberellin response mutant, is caused by a null mutation of the *SLR1* gene, an ortholog of the height-regulating gene *GAI/RGA/RHT/D8*. Plant Cell 13:999–1010. doi:10.1105/tpc.13.5.999

International Brachypodium Initiative. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768. doi:10.1038/nature08747

Jaeger, K.E., and P.A. Wigge. 2007. FT protein acts as a long-range signal in Arabidopsis. Curr. Biol. 17:1050–1054. doi:10.1016/j.cub.2007.05.008

Jiang, G., Y. Xiang, J. Zhao, D. Yin, X. Zhao, L. Zhu, and W. Zhai. 2014. Regulation of inflorescence branch development in rice through a novel pathway involving the pentatricopeptide repeat protein sped1-D. Genetics 197:1395–1407. doi:10.1534/genetics.114.163931

Kardailsky, I., V.K. Shukla, J.H. Ahn, N. Dagenais, S.K. Christensen, J.T. Nguyen, J. Chory, M.J. Harrison, and D. Weigel. 1999. Activation tagging of the floral inducer *FT*. Science 286:1962–1965. doi:10.1126/science.286.5446.1962

Kim, S., V. Plagnol, T.T. Hu, C. Toomajian, R.M. Clark, S. Ossowski, J.R. Ecker, D. Weigel, and M. Nordborg. 2007. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nat. Genet. 39:1151–1155. doi:10.1038/ng2115

Konieczny, A., and F.M. Ausubel. 1993. A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J. 4:403–410. doi:10.1046/j.1365-313X.1993.04020403.x

Koornneef, M., and D. Meinke. 2010. The development of *Arabidopsis* as a model plant. Plant J. 61:909–921. doi:10.1111/j.1365-313X.2009.04086.x

Lee, S.J., T.A. Warnick, S. Pattathil, J.G. Alvelo-Maurosa, M.J. Serapiglia, H. McCormick, et al. 2012. Biological conversion assay using *Clostridium phytofermentans* to estimate plant feedstock quality. Biotechnol. Biofuels 5:5. doi:10.1186/1754-6834-5-5

Li, J., Y. Li, S. Chen, and L. An. 2010. Involvement of brassinosteroid signals in the floral-induction network of *Arabidopsis*. J. Exp. Bot. doi:10.1093/jxb/erq241

López-Alvarez, D., M.L. López-Herranz, A. Betekhtin, and P. Catalán. 2012. A DNA barcoding method to discriminate between the model plant *Brachypodium distachyon* and its close relatives *B. stacei* and *B. hybridum* (Poaceae). PLoS ONE 7:E51058. doi:10.1371/journal.pone.0051058

López-Alvarez, D., A.J. Manzaneda, P.J. Rey, P. Giraldo, E. Benavente, J. Allainguillaume, et al. 2015. Environmental niche variation and evolutionary diversification of the *Brachypodium distachyon* grass complex species in their native circum-Mediterranean range. Am. J. Bot. 102:1073–1088. doi:10.3732/ajb.1500128

Maier, A.T., S. Stehling-Sun, S.L. Offenburger, and J.U. Lohmann. 2011. The bZIP transcription factor PERIANTHIA: A multifunctional hub for meristem control. Front. Plant Sci. 2:79. doi:10.3389/fpls.2011.00079

Malcomber, S.T., and E.A. Kellogg. 2005. *SEPALLATA* gene diversification: Brave new whorls. Trends Plant Sci. 10:427–435. doi:10.1016/j.tplants.2005.07.008

Marriott, P.E., R. Sibout, C. Lapierre, J.U. Fangel, W.G.T. Willats, H. Hofte, L.D. Gómez, and S.J. McQueen-Mason. 2014. Range of cell-wall alterations enhance saccharification in *Brachypodium distachyon* mutants. Proc. Natl. Acad. Sci. USA 111: 14601–14606. doi:10.1073/pnas.1414020111

Mather, K.A., A.L. Caicedo, N.R. Polato, K.M. Olsen, S. McCouch, and M.D. Purugganan. 2007. The extent of linkage disequilibrium in rice (*Oryza sativa* L.). Genetics 177:2223–2232. doi:10.1534/genetics.107.079616

Mochida, K., and K. Shinozaki. 2013. Unlocking Triticeae genomics to sustainably feed the future. Plant Cell Physiol. 54:1931–1950. doi:10.1093/pcp/pct163

Mosquna, A., A. Katz, E.L. Decker, S.A. Rensing, R. Reski, and N. Ohad. 2009. Regulation of stem cell maintenance by the Polycomb protein FIE has been conserved during land plant evolution. Development 136:2433–2444. doi:10.1242/dev.035048

Mur, L.A.J., J. Allainguillaume, P. Catalán, R. Hasterok, G. Jenkins, K. Lesniewska, I. Thomas, and J. Vogel. 2011. Exploiting the Brachypodium tool box in cereal and grass research. New Phytol. 191:334–347. doi:10.1111/j.1469-8137.2011.03748.x

Nakamichi, N., M. Kita, S. Ito, T. Yamashino, and T. Mizuno. 2005. PSEUDO-RESPONSE REGULATORS, PRR9, PRR7 and PRR5, together play essential roles close to the circadian clock of *Arabidopsis thaliana*. Plant Cell Physiol. 46:686–698. doi:10.1093/pcp/pci086

Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford Univ. Press, New York.

Nole-Wilson, S., and B.A. Krizek. 2006. *AINTEGUMENTA* contributes to organ polarity and regulates growth of lateral organs in combination with *YABBY* Genes. Plant Physiol. 141:977–987. doi:10.1104/pp.106.076604

Pabón-Mora, N., G.K.S. Wong, and B.A. Ambrose. 2014. Evolution of fruit development genes in flowering plants. Front. Plant Sci. 5:300.

Pajoro, A., P. Madrigal, J.M. Muiño, J.T. Matus, J. Jin, M.A. Mecchia, et al. 2014. Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. Genome Biol. 15:R41. doi:10.1186/gb-2014-15-3-r41

Pasam, R.K., R. Sharma, M. Malosetti, F.A. van Eeuwijk, G. Haseneyer, B. Kilian, and A. Graner. 2012. Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC Plant Biol. 12:16. doi:10.1186/1471-2229-12-16

Patterson, N., A.L. Price, and D. Reich. 2006. Population structure and eigenanalysis. PLoS Genet. 2:E190. doi:10.1371/journal.pgen.0020190

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

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

R Development Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Stat. Comput., Vienna, Austria.

Ream, T.S., D.P. Woods, C.J. Schwartz, C.P. Sanabria, J.A. Mahoy, E.M. Walters, H.F. Kaeppler, and R.M. Amasino. 2014. Interaction of photoperiod and vernalization determines flowering time of *Brachypodium distachyon*. Plant Physiol. 164:694–709. doi:10.1104/pp.113.232678

Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.

Samach, A., H. Onouchi, S.E. Gold, G.S. Ditta, Z. Schwarz-Sommer, M.F. Yanofsky, and G. Coupland. 2000. Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*. Science 288:1613–1616. doi:10.1126/science.288.5471.1613

Schlereth, A., B. Möller, W. Liu, M. Kientz, J. Flipse, E.H. Rademacher, M. Schmid, G. Jürgens, and D. Weijers. 2010. MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. Nature 464:913–916. doi:10.1038/nature08836

Schulze, S., B.N. Schäfer, E.A. Parizotto, O. Voinnet, and K. Theres. 2010. *LOST MERISTEMS* genes regulate cell differentiation of central zone descendants in *Arabidopsis* shoot meristems. Plant J. 64:668–678. doi:10.1111/j.1365-313X.2010.04359.x

Schwartz, C.J., M.R. Doyle, A.J. Manzaneda, P.J. Rey, T. Mitchell-Olds, and R.M. Amasino. 2010. Natural variation of flowering time and vernalization responsiveness in *Brachypodium distachyon*. BioEnergy Res. 3:38–46. doi:10.1007/s12155-009-9069-3

Schwarz, S., A.V. Grande, N. Bujdoso, H. Saedler, and P. Huijser. 2008. The microRNA regulated SBP-box genes *SPL9* and *SPL15* control shoot maturation in Arabidopsis. Plant Mol. Biol. 67:183–195. doi:10.1007/s11103-008-9310-z

Slavov, G.T., R. Nipper, P. Robson, K. Farrar, G.G. Allison, M. Bosch, J.C. Clifton-Brown, I.S. Donnison, and E. Jensen. 2014. Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass *Miscanthus sinensis*. New Phytol. 201:1227–1239. doi:10.1111/nph.12621

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. doi:10.1093/bioinformatics/btu033

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17:1105–1109. doi:10.1007/BF00037152

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731–2739. doi:10.1093/molbev/msr121

Tian, F., P.J. Bradbury, P.J. Brown, H. Hung, Q. Sun, S. Flint-Garcia, T.R. Rocheford, M.D. McMullen, J.B. Holland, and E.S. Buckler. 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat. Genet. 43:159–162. doi:10.1038/ng.746

Tyler, L., J.U. Fangel, A.D. Fagerström, M.A. Steinwand, T.K. Raab, W.G. Willats, and J.P. Vogel. 2014. Selection and phenotypic characterization of a core collection of *Brachypodium distachyon* inbred lines. BMC Plant Biol. 14:25. doi:10.1186/1471-2229-14-25

Vain, P., B. Worland, V. Thole, N. McKenzie, S.C. Alves, M. Opanowicz, L.J. Fish, M.W. Bevan, and J.W. Snape. 2008. *Agrobacterium*-mediated transformation of the temperate grass *Brachypodium distachyon* (genotype Bd21) for T-DNA insertional mutagenesis. Plant Biotechnol. J. 6:236–245. doi:10.1111/j.1467-7652.2007.00308.x

Vogel, J.P., D.F. Garvin, O.M. Leong, and D.M. Hayden. 2006. *Agrobacterium*-mediated transformation and inbred line development in the model grass *Brachypodium distachyon*. Plant Cell Tissue Organ Cult. 84:199–211. doi:10.1007/s11240-005-9023-9

Vogel, J., and T. Hill. 2008. High-efficiency *Agrobacterium*-mediated transformation of *Brachypodium distachyon* inbred line Bd21-3. Plant Cell Rep. 27:471–478. doi:10.1007/s00299-007-0472-y

Vogel, J.P., M. Tuna, H. Budak, N. Huo, Y.Q. Gu, and M.A. Steinwand. 2009. Development of SSR markers and analysis of diversity in Turkish populations of *Brachypodium distachyon*. BMC Plant Biol. 9:88. doi:10.1186/1471-2229-9-88

Wickham, H. (2009) ggplot2: Elegant graphics for data analysis. Springer Science & Business Media, New York.

Xu, X., X. Liu, S. Ge, J.D. Jensen, F. Hu, X. Li, et al. 2012. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat. Biotechnol. 30:105–111. doi:10.1038/nbt.2050

Yan, J., T. Shah, M.L. Warburton, E.S. Buckler, M.D. McMullen, and J. Crouch. 2009. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS ONE 4:E8451. doi:10.1371/journal.pone.0008451

Yano, M., Y. Katayose, M. Ashikari, U. Yamanouchi, L. Monna, T. Fuse, et al. 2000. *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. Plant Cell 12:2473–2483. doi:10.1105/tpc.12.12.2473

Yu, J., G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, J.F. Doebley, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203–208. doi:10.1038/ng1702

Zagotta, M., S. Shannon, C. Jacobs, and D. Meeks-Wagner. 1992. Early-flowering mutants of *Arabidopsis thaliana*. Funct. Plant Biol. 19:411–418.

Zakhrabekova, S., S.P. Gough, I. Braumann, A.H. Müller, J. Lundqvist, K. Ahmann, et al. 2012. Induced mutations in circadian clock regulator *Mat-a* facilitated short-season adaptation and range extension in cultivated barley. Proc. Natl. Acad. Sci. USA 109:4326–4331. doi:10.1073/pnas.1113009109

Zhang, Z., E. Ersoz, C.Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, et al. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42:355–360. doi:10.1038/ng.546