Although natural genetic variation provides the foundation for crop improvement, the amount of diversity currently used in most breeding programs is limited. This overall decline in crop species’ diversity has both historical and modern origins. In the past, strong selection for domestication-related traits created a severe genetic bottleneck and reduced diversity in domesticates compared with wild relatives (Hyten et al., 2006). Today, most high-yielding cultivars are derived from crosses among genetically related varieties instead of less adapted but more variable genotypes (Tanksley and McCouch, 1997). Consequently, modern breeding practices have further constrained the amount of extant diversity in crop species and limited genetic gains in breeding programs. Because narrow or limited diversity can also have disastrous consequences in the field, the 1970 southern corn leaf blight (Bipolaris maydis) epidemics for example (Tatum, 1971), broadening the genetic base of germplasm available to breeders is of paramount importance.

Sorghum bicolor (L.) Moench, a tropical grass probably domesticated in East Africa 3000 to 6000 yr ago (Kimber, 2000), is a staple...
cereal food for millions of people in the developing world. In the United States, sorghum is grown primarily for animal feed, but recently grain, forage, and sweet sorghum types have received increased attention as potential energy crops (Rooney et al., 2006). In the early 1960s, sorghum breeders recognized that elite U.S. cultivars had experienced strong genetic bottlenecks as a result of breeding practices. To provide a long-term solution for the development of a sustainable sorghum production system, the USDA in cooperation with Texas A&M University initiated the Sorghum Conversion Program (SCP), a strategy to introduce novel genetic variation from exotic, tropical germplasm into modern U.S. cultivars (Stephens et al., 1967). To expedite their use in temperate-zone breeding programs, tropical lines were converted to photoperiod insensitive, early maturing, and short stature phenotypes. This was accomplished by crossing each tropical line to a temperate elite and selecting the progeny for day-neutral flowering and reduced height. Progeny were then backcrossed repeatedly to the tropical parent until the resultant lines were fixed for temperate alleles at major loci controlling maturity and height while retaining ~90% of the tropical genome (Lin et al., 1995). To date, ~850 converted tropical lines have been released by the SCP and this germplasm has allowed breeders to exploit novel variation for insect and disease resistance, drought tolerance, heterosis, and grain quality. As a result, most of the U.S. sorghum hybrids grown today have some tropical germplasm in their pedigrees (Gabriel, 2005). Because the SCP lines contain much of the genetic diversity present in tropical sorghums (Stephens et al., 1967), the use of this material offers a unique opportunity for dissecting the genetic and molecular bases of agriculturally important traits through association mapping.

Association mapping is a powerful tool for high-resolution mapping of loci underlying quantitative traits and is dependent on the structure of linkage disequilibrium or the non-random association of alleles or polymorphisms at different loci (Flint-Garcia et al., 2003). Significant associations between genotypes and phenotypes can be caused (i) by marker loci harboring causal polymorphisms, (ii) by marker loci being physically linked to a polymorphism that influences a particular phenotype, and, of greater concern, (iii) from the effects of population structure or familial relationship (kinship) between individuals comprising the test population. Individuals belonging to the same subpopulations or that are related by descent (kin), are more likely to both resemble each other phenotypically and share common alleles, independently of these alleles being linked or not to the causal polymorphism (leading to spurious associations). Therefore, knowledge of population structure and kinship in association mapping populations is critical. In fact, Yu et al. (2006) have recently shown that controlling for such demographic factors can lead to a significant reduction in the number of spurious associations in maize (Zea mays L.).

Our goal in this research was to assemble and characterize the genetic and phenotypic diversity of a panel of sorghum germplasm suitable for association mapping. We assessed the levels of population structure and familial relatedness with simple sequence repeat (SSR) loci and evaluated the performance of several models in controlling for spurious associations (i.e., Type I error).

**MATERIALS AND METHODS**

**Sorghum Association Panel**

The sorghum panel evaluated in this study contained 377 accessions comprising both converted tropical (n = 228) and breeding (n = 149) lines (Supplementary Table 1; all supplementary information can be viewed and downloaded from www.sorghumdiversity.org/sorghum/supplementary.html). Converted tropical lines, products of the SCP, were selected mainly on the basis of (i) geographic and genotypic diversity, attempting to ensure that material from most African regions, cultivated races (Harlan and de Wet, 1972), and working groups (within-race morphological subdivisions) (Dahlberg, 2000) were represented, and (ii) historical importance, so lines that have significantly contributed to breeding as sources of resistance (or tolerance) to biotic and abiotic stresses were included. The breeding lines comprised photoperiod-insensitive elite inbreds, improved cultivars, and landraces that either have been or are being extensively used in U.S. sorghum breeding programs. The breeding panel composition was based less on race and working group designation per se and more on historical importance in sorghum improvement. There was an effort to represent all important types of sorghum (e.g., grain, forage, and sweet) and also to include examples of lines that have been significant in places other than the United States (Central America, Africa, and India). Some modern breeding lines from Texas and Kansas were included, but much more emphasis was placed on including inbreds that have been successful in the seed industry.

**Phenotyping**

The sorghum association panel was grown in a randomized complete block design in Weslaco (two replicates), College Station (two replicates), and Lubbock (four replicates), TX, in the summer of 2006 in 6-m rows spaced at either 0.50 or 0.75 m. Eight traits were evaluated on a per-row basis: flag leaf length and width (measured only at Weslaco and College Station), plant height, terminal branch length, flowering time (measured as time to mid-anthesis), panicle length, and flag leaf height and exsertion. A linear model was used to account for the effects of location and replication and derive a single phenotypic value for each genotype.

**Genotyping and Statistical Analyses**

**DNA Preparation and Polymerase Chain Reactions**

Six to 10 seeds from each accession were germinated in small pots in the dark at 27°C. Total genomic DNA was isolated from pooled etiolated seedlings (~7 d old) following a standard CTAB extraction protocol (Doyle and Doyle, 1987). Amplifications were performed in 10-μL volumes using one of two temperature cycling protocols.
(Supplementary Table 2) depending on how fluorescent amplifiers were labeled (either with locus-specific or universal primers). The following reaction components were common between the two protocols: 20 ng of total genomic DNA, 1X polymerase chain reaction (PCR) buffer, 2.5 mmol L⁻¹ MgCl₂, 0.2 mmol L⁻¹ dNTPs, 5% DMSO, and 0.5U of Taq DNA polymerase. For PCRs with end-labeled primers, 2 pmol each of 5’-labeled forward and unlabeled reverse primers were used. Cycling protocol consisted of 95°C for 3 min; followed by 27 cycles of 95°C for 30 s, 55°C for 20 s, and 72°C for 30 s; and incubation at 72°C for 45 min. Polymerase chain reactions with universal primers were performed using unlabeled locus-specific primers, one of which contained a 5’ binding site for the universal primer (i.e., “pigtail”), and a 5’-labeled universal primer (see Supplementary Table 2). Primer concentrations and cycling conditions were as previously described (Schuelke, 2000) and fluorescent SSR detection and allele scoring was performed according to Casa et al. (2005).

**Simple Sequence Repeat Loci and Diversity Estimates**

A total of 49 SSR loci (Supplementary Table 2) were evaluated. Loci were selected based primarily on their genomic location (i.e., to achieve fairly uniform genome coverage) and secondarily on information content (see Casa et al., 2005). Approximately half of the SSRs assayed were also used in a previous study to characterize 3000 accessions from the world sorghum collection (Billot and Hash, 2006). Summary statistics, including number of alleles, allele frequencies, and polymorphism information content for each locus, were calculated with PowerMarker version 3.0 (Liu and Muse, 2005).

**Population Structure and Kinship**

The program Structure, version 2.1 (Pritchard et al., 2000), was used to determine the presence of population structure and assign sorghum lines to subpopulations. This program implements a model-based clustering method for inferring population structure using genotypic data from unlinked markers. We used an ancestry model that allowed population admixture, and allele frequencies among populations were assumed to be correlated (i.e., allele frequencies were likely to be similar due to shared ancestry or migration). We also tested for the optimal number of subpopulations, k (distinct from K, the kinship matrix; see below). We allowed k to vary from 1 to 12, with three independent runs for each value. The optimal k value was determined based on the estimated logarithmic likelihood of the data and its performance in the unified-mixed model for association analysis (Yu et al., 2006; and see below). Initially, $5 \times 10^5$ burn-in lengths (Pritchard et al., 2000) and sampling periods of iterations were used for each k value, while $5 \times 10^5$ burn-in and sampling periods of iterations were used for the optimal k value. Runs that did not meet the convergence criterion were not analyzed. A graphical display of subpopulation composition was generated with DISTRUCT (Rosenberg, 2004).

Simple sequence repeat based relative kinship estimates, defined as $F_{ij} = (Q_{ij} - Q_m)/(1 - Q_m) \cong \Theta_{ij}$, where $\Theta_{ij}$ is the pairwise kinship coefficient ($F_{ij}$ is an estimator of the coefficient), $Q_{ij}$ is the probability of identity by state for random genes from i and j, and $Q_m$ is the average probability of identity by state for genes coming from random individuals in the population from which i and j were drawn, were obtained as previously described (Loiselle et al., 1995; Ritland, 1996; Lynch and Ritland, 1999; Rousset, 2002) using SPAGeDi 1.2 (Hardy and Vekemans, 2002). Confidence intervals (95%) for kinship coefficients were calculated as follows:

$$
\bar{F}_{ij} \pm 1.96\sigma_{\bar{F}_{ij}}
$$

where

$$
\sigma_{\bar{F}_{ij}} = \frac{\sigma_{F_{ij}}}{\sqrt{n}}
$$

$\sigma_{\bar{F}_{ij}}$ is the standard deviation for $F_{ij}$; $F_{ij}$ is the average kinship between any given individual in (a) subpopulation and all other individuals in the same (sub)population, and n is the population size. For each subpopulation identified we also estimated the expected heterozygosity, $H_e$ (also referred to as unbiased gene diversity, D), and confidence intervals based on 1000 bootstrap replicates using PowerMarker version 3.0.

**Association Model Testing**

In all, we tested the performance of six different association models in controlling for false positives or spurious associations (Type I error). These included (i) a model that did not control for population structure or relatedness (naive), expressed as $y = A\alpha + e$; (ii) a model that accounted for population structure (Q), $y = A\alpha + Qv + e$; (iii) a model that controlled for familial relatedness or kinship (K), $y = A\alpha + Zu + e$; (iv) an alternative kinship-based model (K'), $y = A\alpha + Zu' + e$; (v) a mixed model that accounted for both population structure and kinship (QK), $y = X\beta + A\alpha + Qv + Zu + e$, and (vi) a mixed model with the alternative kinship estimates (QK'), $y = X\beta + A\alpha + Qv + Zu' + e$. Here, $y$ is a vector of phenotypic observation; $\alpha$ is a vector of allelic effects; $e$ is a vector of residual effects; $v$ is a vector of population effects; $\beta$ is a vector of fixed effects other than allelic or population group effects; $u$ is a vector of polygenic background effects; Q is the population membership assignment matrix (based on SSR genotypic data and calculated using Structure) relating y to $v$; and X, A, and Z are incidence matrices of 1s and 0s relating y to $\beta$, $\alpha$, and $u$, respectively. The variances of the random effects are expressed as $\text{Var}(u) = K'V'$ and $\text{Var}(e) = R'V'R$, where $K$ is the kinship matrix (based on SSR genotypic data and calculated using SPAGeDi), R is a matrix with the off-diagonal elements being zero and the diagonal elements being the reciprocal of the number of observations for which each phenotypic data point was observed, $V$ is the genetic variance, and $V_K$ is the residual variance. Best linear unbiased estimates of $\beta$, $\alpha$, and $v$ (fixed effects), and best linear unbiased predictions of $u$ (random effects) were obtained by solving the mixed-model equations, Eq. [5] and [6], above. We should note that the naive, Q, K, and QK models have been described previously (Yu et al., 2006). To our knowledge, this is the first time that the K’ and QK’ models have been tested.

In this study, we tested models that used two different methods for estimating kinship (the K and K’ matrices). In the first round of simulations (K matrix), the negative kinship values were simply set to zero as suggested by Yu et al. (2006). In a second round of simulations, $F_{ij} = \min(F_{ij})$ was used instead.
to compute a new kinship matrix (K′ matrix), where F′, = (F, − F,)/{(1 − F,)} ≥ Θ,, F, is the raw pairwise kinship coefficients from the SPAGeDi output, and min F, is the average of the minimum (negative) F, values from each row (or column) of the untransformed kinship matrix. In the case of K′, negative F, values quantify divergence between individuals belonging to different populations under drift and not under selection or adaptation (which is better accounted for by Q).

The Type I error rate was simulated based on the method described by Yu et al. (2006). Because of the low marker density relative to the extent of linkage disequilibrium in sorghum (Hamblin et al., 2005), few, if any, randomly distributed SSRs should associate with particular phenotype(s). Consequently, the random SSRs provide an empirical null distribution with which the models (above) can be tested for their ability to control Type I error. Only alleles with a frequency >10% in our sample were used for the simulation. Because both fixed and random effects were involved, only likelihood-based methods could be used for model comparison. In this study we used two methods, the −2 residual log likelihood (for comparing nested models) and the Bayesian information criterion (BIC) (Schwarz, 1978) (for non-nested models).

RESULTS AND DISCUSSION

In the study presented here, we characterized a sorghum panel comprising 377 lines that were assembled for association mapping studies. Such panels have already been compiled for a number of other important crop species including maize (Flint-Garcia et al., 2005), durum wheat (Triticum durum Desf.; Maccafferri et al., 2006), bread wheat (Triticum aestivum L.; Breseghello and Sorrells, 2006), and barley (Hordeum vulgare L.; Rostoks et al., 2006), and association-mapping-based approaches have allowed the identification of genes underlying the variation in key maize traits including flowering time (Thornsberry et al., 2001) and kernel composition (Whitt et al., 2002; Palaisa et al., 2003; Wilson et al., 2004).

Diversity Levels and Partitioning in the Association Panel

Levels of genetic diversity in converted tropical (n = 228) and breeding (n = 149) panels were assessed at 49 SSR loci. Two loci (Xtxp065 and Xtxp287) exhibited >20% missing data and were discarded from further analysis. Summary statistics for the remaining 47 loci are presented in Supplementary Table 2. A total of 553 alleles were detected in the entire panel, with the number of alleles per locus ranging from 2 (Xcup06, 19, 23, and 53) to 50 (Xtxp343). Polymorphism information content values ranged from 0.01 (Xcup19 and Xcup53) to 0.93 (Xtxp067 and Xtxp343), with an average value of 0.56. Although the converted tropical sorghums exhibited both a higher average number of alleles per locus (10.9) and greater diversity (0.56) than the breeding lines (7.5 and 0.51, respectively), analysis of molecular variance indicated that only 2% of the variation was due to differences between panels. As a whole, therefore, the SCP and breeding lines were not significantly differentiated from each other. This result was not surprising, considering that the breeding panel was designed to contain as much diversity as possible and many of the breeding lines had tropical progenitors.

Population Structure and Kinship

To assess population structure, we used a model-based method for determining the number of subpopulations, k, in our panel. An accession was assigned to the subpopulation or group to which it showed the highest probability of membership. When k varied from 1 to 12, the posterior probability of the data improved steadily for k ≥ 8 and reached a plateau for k ≥ 9 (Fig. 1). Based on this result and results from further tests using our phenotypic data (see below), splitting the panel into either nine (k = 9) or 10 (k = 10) subpopulations best described population structure for testing the association mapping models.

The primary racial compositions along with kinship and heterozygosity statistics for subpopulations in Table 1. While kafir, durra, West African, and zerazera accessions were consistently assigned for both k = 9 and k = 10 and showed highly correlated probability of assignment across simulations (i.e., across replicated runs for a given k, as well as for runs where 7 ≤ k ≤ 10), the probability of assignment of individuals to specific groupings was not always consistent. For example, broomcorn, sudanense, and other loose-headed S. bicolor accessions were assigned to different populations in different runs for the

![Figure 1. Posterior probability, ln P(D), of the data as a function of the number of subpopulations (k), where k was allowed to range from 1 to 12. Diamonds represent independent runs for each value of k.](image-url)
same value of \( k \). For \( k = 9 \), Subpopulation I was comprised mostly of nigricans (caudatum type) and guinea accessions from East Africa and India, whereas Subpopulation VIII contained mostly caudatum and guinea types from West Africa (Table 1). Because racial classification relies on a limited number of morphological traits, it was not surprising that some caudatum and guinea accessions clustered by geographic location rather than by race. Caudatum accessions were also prevalent in two other subpopulations: IV (mostly zerazera working group) and IX (a mix of accessions from different caudatum working groups including “hybrid/intermediate” caudatum types). Subpopulation III was comprised primarily of kafir types, whereas durra sorghums were split between two groups: V (accessions from India and Ethiopia) and VI (milo durras). Accessions classified as bicolor were prevalent in Subpopulations II (containing sudanense and broomcorn types) and VII (comprising dochna-bicolors and margaritiferum, a guinea-type sorghum grown primarily in West Africa). The primary differences between groupings in \( k = 9 \) and \( k = 10 \) were the division of Subpopulation IX into two groups (\( k = 10 \), Subpopulations IX and X) and alternative clustering of accessions assigned to poorly defined groups such as \( k = 9 \) Subpopulations I, II, and VII (Table 1).

A graphical display of estimated population structure for \( k = 9 \) is presented in Supplementary Fig. 1. Despite the amount of structure and the prevalence of certain sorghum racial types within particular groups, the sorghum panel also exhibited substantial genetic admixture (depicted in Supplementary Fig. 1 as more than one color within a single vertical bar). Kafi and zerazera/caudatum types, for example, occurred in the genetic backgrounds of about half of the accessions comprising the milo/feterita group (Subpopulation VI). In addition, the highest membership assignment to any one population was <0.5 for ~15% of the accessions.

We also calculated the mean kinship, \( \bar{F}_{ij} \), and the expected heterozygosity, \( H_e \), for each population identified for \( k = 9 \) and \( k = 10 \) (Table 1). Among the well-defined subpopulations (i.e., those that were consistently defined between analyses and for which individuals showed highly correlated probability of assignment across simulations, shaded areas in Table 1), relatively high \( \bar{F}_{ij} \) and low \( H_e \) were observed for Subpopulations III and VIII. Here, strong kinship coupled with low diversity suggests that the kafir and West African guinea/caudatum groups have experienced a more severe genetic bottleneck than the other well-defined subpopulations. This observation is consistent with the historical geographic isolation of both groups and temperate adaptation of kafir types (see Casa et al., 2005). As would be expected, a tendency toward weaker kinship and higher gene diversity was observed among the loosely defined subpopulations (i.e., I, VII, and IX/X).

### Phenotypic Diversity

This panel contains much of the phenotypic diversity present in tropical sorghums (e.g., variation for plant morphology and panicle architecture), and represents diverse geographic and climatic regions (representatives from the Americas, Asia, and the entire African continent, from high and low elevations, rainy and dry environments). Therefore, the collection presents an excellent source of variability for dissecting the genetic bases of agriculturally important traits and adaptation.

Phenotypic variation for eight traits, organized by subpopulation (\( k = 9 \)), is shown in Table 2. Except for the sudanense/broomcorn subpopulation, which was considerably taller than the other groups, the amount of variation for
both plant height and flowering time (i.e., maturity–related traits) across all subpopulations was similar. Because variation in maturity can confound the phenotypic evaluation of other agriculturally important traits, use of this germplasm panel in association studies should simplify dissection of these traits in sorghum. Perhaps more importantly, this panel should be well suited for association studies in higher latitude environments, such as the United States. Besides plant height and flowering time, all other measured traits exhibited substantial variation both within and among subpopulations (Table 2), particularly for inflorescence–related traits such as panicle and terminal branch lengths.

### Association Models

Population structure and kinship among individuals not only affect the amount and patterns of diversity in a collection but can also lead to spurious associations between genotypes and phenotypes (i.e., Type I error).
In this study, we tested the performance of six models in minimizing Type I error: one model that did not control for population structure or relatedness (naive), one that accounted for population structure only (Q), two that controlled only for familial relatedness, each using a different kinship matrix (K and K’), and two that considered both population structure and kinship (QK and QK’) (see above). The population membership assignment (Q) and kinship (K) matrices for our association panel are provided as Supplementary Tables 3 to 6. We should also note that the effective number of markers used in this analysis was 45 (two of the 47 SSR loci, Xcup19 and 55, were not included because we detected only two alleles and the minor allele frequency was <0.10).

Optimizing the Number of Subpopulations for Model Testing

Results from our analysis of population structure (see above) indicated that the probability of groupings based on the genotypic data improved steadily for \( k \geq 8 \) but reached a plateau for \( k \geq 9 \) (Fig. 1). Therefore, we refined estimates of the optimal number of subpopulations in our panel by testing the likelihood of each of these \( k \) values against the phenotypic data for all measured traits. Figure 2 shows the performance of the QK model for some phenotypic traits, measured by the BIC, as a function of \( k \). While results for only four traits are shown, the lowest BIC values, and therefore the best likelihood for all traits, were obtained for \( k = 9 \) and 10 using mixed models QK and QK’. We, therefore, used membership assignment matrices for both \( k = 9 \) and \( k = 10 \) (Q_9 and Q_10, respectively; Supplementary Tables 3 and 4) for model testing and further analyses.

It is not surprising that the optimal number of sub-populations is trait dependent (i.e., different estimates of the number of subpopulations need to be considered for model comparison and population analyses). Although individuals may share little or no identity by descent (kinship coefficient is near zero), they can still belong to the same population and share many quantitative trait locus alleles at loci underlying adaptation since they are under the same selection pressure. In this case, populations will be more differentiated for loci underlying adaptation-related traits than for random loci. Therefore, correcting for population structure that correlates with adaptation (which can be trait specific) is also important for preventing false positives that associate with population differentiation.

![Figure 2. Influence of the number of subpopulations (k) on the performance of the QK model, a mixed model that accounts for both population structure and kinship, measured by the Bayesian information criterion (BIC), for selected phenotypic traits. Lower BIC values indicate best likelihood.](image-url)
Performance of Various Association Models

Simulations of Type I error for all models and all quantitative traits combined are presented in Fig. 3. As expected, the naive model showed the highest (25% of the P values are under the 5% threshold) inflation of $P$ values (i.e., $P$ values were not uniformly distributed), and consequently the highest Type I error. Controlling for population structure ($Q$ model) yielded a slight improvement over the naive model but a considerable inflation of $P$ values (15% of the $P$ values are under the 5% threshold) can still be seen (Fig. 3). On the other hand, all other models ($K$, $K'$, $QK$, and $QK'$) showed a good approximation to a uniform distribution of $P$ values, with $QK$ and $QK'$ performing slightly better (4.8% of the $P$ values are under the 5% threshold) than the $K$ or $K'$ models (5.1% of the $P$ values are under the 5% threshold).

The relative performance of each model was evaluated with a likelihood-ratio test. For each trait measured, 12 model comparisons were performed. Results presented in Table 3 reveal two important aspects of the data. First, models that account for both population structure and kinship tended to perform better than those that controlled only for $Q$ or $K$. Second, the magnitude of improvement achieved by accounting for both $Q$ and $K$ was trait dependent. For example, flag leaf height and length were less affected by population structure and kinship (Table 3, higher $P$ values than other traits for most pairwise comparisons). Inflorescence-related characters (e.g., terminal branch length and panicle length), on the other hand, tended to be more sensitive to $Q$ and $K$ (as shown by their lower $P$ values). Because sorghum races have traditionally been defined by panicle architecture and seed morphology characters, it was not surprising to see that these traits were more correlated to population structure.

Bayesian information criteria for each model and trait are shown in Table 4. As noted above, models that accounted for both population structure and kinship tended to perform better than simple models. Moreover, mixed models that used $K$ (either with membership assignment matrices $Q_9$ or $Q_{10}$) tended to perform better than those using $K'$ and, in general, the $QK$ model should be favored. We should note, however, that although the $K$ matrix tended to capture a larger proportion of the phenotypic variance and exhibit improved BIC (Table 4) than $K'$, the $QK'$ model may be best for some traits since $K'$ showed better convergence properties than $K$ alone (as shown in Table 4, $K$ did not converge for panicle length and flag leaf width).

Community Resources

During the past 6 yr, work by several research groups has provided a wealth of genetic and genomic information for *S. bicolor*. Molecular genetic resources in the public domain include detailed genetic (Menz et al., 2002; Bow- ers et al., 2003), physical (Arizona Genomics Institute, 2007), and comparative maps (Paterson et al., 2004), more than 200,000 expressed sequence tag sequences (National Center for Biotechnology Information, 2007), and 1X coverage of the sorghum gene space (from methylation-filtered genomic clones) (Bedell et al., 2005). Genomics efforts have been strengthened significantly in 2007 with the preliminary release of the *S. bicolor* genome sequence (www.phytozome.net/sorghum [verified 6 Oct. 2007]). In addition to extensive mapping and DNA sequencing, characteristics of the sorghum
Table 3. Association model comparison based on the likelihood-ratio test.†

<table>
<thead>
<tr>
<th>Trait</th>
<th>Naive vs. Q9</th>
<th>Naive vs. Q10</th>
<th>Naive vs. K</th>
<th>Naive vs. K′</th>
<th>Q9 vs. Q9K</th>
<th>Q9 vs. Q9K′</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A(λ)</td>
<td>P value</td>
<td>A(λ)</td>
<td>P value</td>
<td>A(λ)</td>
<td>P value</td>
</tr>
<tr>
<td>Flowering time</td>
<td>63</td>
<td>3.53 × 10⁻¹⁰</td>
<td>82.8</td>
<td>1.42 × 10⁻¹³</td>
<td>69.6</td>
<td>7.26 × 10⁻¹⁷</td>
</tr>
<tr>
<td>Terminal branch length</td>
<td>131.6</td>
<td>5.56 × 10⁻²⁴</td>
<td>85.9</td>
<td>3.47 × 10⁻¹⁴</td>
<td>92.3</td>
<td>7.45 × 10⁻²²</td>
</tr>
<tr>
<td>Flag leaf width</td>
<td>155.6</td>
<td>6.09 × 10⁻²⁹</td>
<td>177.7</td>
<td>7.03 × 10⁻₃³</td>
<td>5.8</td>
<td>3.30 × 10⁻⁸</td>
</tr>
<tr>
<td>Plant height</td>
<td>139.5</td>
<td>1.31 × 10⁻²⁵</td>
<td>120.7</td>
<td>3.65 × 10⁻₂¹</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Panicle length</td>
<td>17.5</td>
<td>4.14 × 10⁻²</td>
<td>22.9</td>
<td>1.11 × 10⁻³</td>
<td>15.3</td>
<td>9.17 × 10⁻⁴</td>
</tr>
<tr>
<td>Flag leaf length</td>
<td>33</td>
<td>1.34 × 10⁻⁴</td>
<td>23.1</td>
<td>1.04 × 10⁻²</td>
<td>6.8</td>
<td>9.11 × 10⁻³</td>
</tr>
<tr>
<td>Exsertion</td>
<td>61</td>
<td>8.60 × 10⁻¹⁰</td>
<td>76.6</td>
<td>2.32 × 10⁻¹²</td>
<td>46.7</td>
<td>8.27 × 10⁻¹²</td>
</tr>
</tbody>
</table>

Table 4. Performance of all models for each trait as measured by the Bayesian information criteria (BIC). Italics indicate models exhibiting lower BIC values (lower means better) for that trait.‡

<table>
<thead>
<tr>
<th>Trait</th>
<th>Naive</th>
<th>Qs</th>
<th>QsK</th>
<th>K</th>
<th>QsK</th>
<th>QsK′</th>
<th>QsK′</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A(λ)</td>
<td>P value</td>
<td>A(λ)</td>
<td>P value</td>
<td>A(λ)</td>
<td>P value</td>
<td>A(λ)</td>
</tr>
<tr>
<td>Flowering time</td>
<td>923.9</td>
<td>861.0</td>
<td>841.2</td>
<td>860.2</td>
<td>873.6</td>
<td>850.4</td>
<td>844.3</td>
</tr>
<tr>
<td>Terminal branch length</td>
<td>931.1</td>
<td>799.5</td>
<td>845.1</td>
<td>844.5</td>
<td>881.6</td>
<td>748.7</td>
<td>768.6</td>
</tr>
<tr>
<td>Flag leaf width</td>
<td>1041.1</td>
<td>885.5</td>
<td>863.4</td>
<td>–</td>
<td>–</td>
<td>945.9</td>
<td>862.4</td>
</tr>
<tr>
<td>Plant height</td>
<td>1071.0</td>
<td>1013.2</td>
<td>1015.8</td>
<td>1027.1</td>
<td>1033.8</td>
<td>988.3</td>
<td>989.4</td>
</tr>
<tr>
<td>Panicle length</td>
<td>928.2</td>
<td>788.7</td>
<td>807.4</td>
<td>–</td>
<td>–</td>
<td>835.9</td>
<td>751.3</td>
</tr>
<tr>
<td>Flag leaf height</td>
<td>924.5</td>
<td>906.9</td>
<td>901.6</td>
<td>915</td>
<td>905.8</td>
<td>887.9</td>
<td>884.9</td>
</tr>
<tr>
<td>Flag leaf length</td>
<td>1046.8</td>
<td>1013.8</td>
<td>1023.6</td>
<td>1045.9</td>
<td>1037.2</td>
<td>1017.4</td>
<td>1016.9</td>
</tr>
<tr>
<td>Exsertion</td>
<td>1069.4</td>
<td>1008.5</td>
<td>992.8</td>
<td>1028.6</td>
<td>1023.1</td>
<td>992</td>
<td>999.2</td>
</tr>
</tbody>
</table>

† Naive, model that did not control for population structure or relatedness; Q, model that accounted for only population structure; K, model that controlled for familial relatedness or kinship; K′, an alternative kinship-based model; QK, a mixed model that accounted for both population structure and kinship; QK′, a mixed model with the alternative kinship estimates (K′). Membership assignment matrices assumed either nine (Q) or 10 (QK) subpopulations.

‡ Likelihood ratio.

§ Comparison not possible (K model did not converge).

framework for genetic analyses aimed at improving sorghum for agronomically important traits.

We have developed an additional resource for the sorghum research community, a panel consisting of 377 diverse lines suitable for association mapping. Because genotypic data for this panel along with appropriate statistical models (QK method) for correcting for population structure and kinship are being made available to the entire sorghum community, researchers interested in using this germplasm can collect phenotypic data for their favorite trait or markers and candidate genes without the need for further

Table 4. Performance of all models for each trait as measured by the Bayesian information criteria (BIC). Italics indicate models exhibiting lower BIC values (lower means better) for that trait.‡

<table>
<thead>
<tr>
<th>Trait</th>
<th>Naive</th>
<th>Qs</th>
<th>QsK</th>
<th>K</th>
<th>QsK</th>
<th>QsK′</th>
<th>QsK′</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A(λ)</td>
<td>P value</td>
<td>A(λ)</td>
<td>P value</td>
<td>A(λ)</td>
<td>P value</td>
<td>A(λ)</td>
</tr>
<tr>
<td>Flowering time</td>
<td>923.9</td>
<td>861.0</td>
<td>841.2</td>
<td>860.2</td>
<td>873.6</td>
<td>850.4</td>
<td>844.3</td>
</tr>
<tr>
<td>Terminal branch length</td>
<td>931.1</td>
<td>799.5</td>
<td>845.1</td>
<td>844.5</td>
<td>881.6</td>
<td>748.7</td>
<td>768.6</td>
</tr>
<tr>
<td>Flag leaf width</td>
<td>1041.1</td>
<td>885.5</td>
<td>863.4</td>
<td>–</td>
<td>–</td>
<td>945.9</td>
<td>862.4</td>
</tr>
<tr>
<td>Plant height</td>
<td>1071.0</td>
<td>1013.2</td>
<td>1015.8</td>
<td>1027.1</td>
<td>1033.8</td>
<td>988.3</td>
<td>989.4</td>
</tr>
<tr>
<td>Panicle length</td>
<td>928.2</td>
<td>788.7</td>
<td>807.4</td>
<td>–</td>
<td>–</td>
<td>835.9</td>
<td>751.3</td>
</tr>
<tr>
<td>Flag leaf height</td>
<td>924.5</td>
<td>906.9</td>
<td>901.6</td>
<td>915</td>
<td>905.8</td>
<td>887.9</td>
<td>884.9</td>
</tr>
<tr>
<td>Flag leaf length</td>
<td>1046.8</td>
<td>1013.8</td>
<td>1023.6</td>
<td>1045.9</td>
<td>1037.2</td>
<td>1017.4</td>
<td>1016.9</td>
</tr>
<tr>
<td>Exsertion</td>
<td>1069.4</td>
<td>1008.5</td>
<td>992.8</td>
<td>1028.6</td>
<td>1023.1</td>
<td>992</td>
<td>999.2</td>
</tr>
</tbody>
</table>

Naive, model that did not control for population structure or relatedness; Q, model that accounted for only population structure; K, model that controlled for familial relatedness or kinship; K′, an alternative kinship-based model; QK, a mixed model that accounted for both population structure and kinship; QK′, a mixed model with the alternative kinship estimates (K′). Membership assignment matrices assumed either nine (Q) or 10 (QK) subpopulations.

K model did not converge.
SSR genotyping. For the short term, requests for a limited number of seeds of the sorghum association panel should be sent to Cleve Franks at the USDA-ARS Plant Stress and Germplasm Development Unit, Cropping Systems Research Laboratory, Lubbock, TX. In the near future, these lines will be maintained and distributed by the U.S. National Plant Germplasm System (www.ars-grin.gov/npgs/). Furthermore, 20 of the diverse lines characterized in this study are now being used to develop recombinant inbred populations in our labs for use in nested association mapping strategies (Yu et al., 2008). With the community resources presently available, *S. bicolor* is achieving sets of genetic data and genomics tools comparable to those of *Arabidopsis thaliana* and *C. sativa*.

**Acknowledgments**

Thanks to Charlotte Acharya for assistance with data collection and analysis and to Claire Billot (CIRAD) and Genoplante for making primer sequences available before publication. We also want to express our gratitude to Martha Hamblin for her comments and suggestions. Special thanks to Dr. Darrel Rosenow for assistance in classifying the accessions used in this study.

**References**


