Crop Science

A global resource for exploring and exploiting genetic variation in sorghum crop wild relatives

Emma S. Mace[1,2,*] | Alan W. Cruickshank[1,*] | Yongfu Tao[2] | Colleen H. Hunt[1,2] | David R. Jordan[1,2]

[1] Agri-Science Queensland, Department of Agriculture and Fisheries (DAF), Hermitage Research Facility, Warwick, QLD 4370, Australia

[2] Queensland Alliance for Agriculture and Food Innovation (QAAFI), The University of Queensland, Hermitage Research Facility, Warwick, QLD 4370, Australia

Correspondence
David Jordan, Agri-Science Queensland, Department of Agriculture and Fisheries (DAF), Hermitage Research Facility, Warwick, QLD4370, Australia.
Email: david.jordan@uq.edu.au

[*] These authors contributed equally to this work

Assigned to Associate Editor Benjamin Kilian.

Abstract

One response to mitigate the impact of climate change on agricultural systems is to develop new varieties that are tolerant to the new range of biotic and abiotic challenges this change causes. This requires access to novel variants of genes for complex adaptive traits. Crop wild relatives (CWRs) are a potentially valuable source of these genes; however, these materials are often difficult to work with and identifying valuable alleles is difficult without substantial investment in prebreeding. In this study, we describe the development of a nested association mapping population for sorghum [*Sorghum bicolor* (L.) Moench] using two cultivated grain sorghum reference parents and nine wild and exotic sorghum accessions as donors. The donor parents come from the *S. bicolor* (L.) Moench subsp. *verticilliflorum* (Steud.) de Wet ex Wiersema & J. Dahlb., *S. bicolor* (L.) Moench subsp. *drummondii* (Steud.) de Wet ex Davidse, and *S. bicolor* (L.) Moench subsp. *bicolor margaritiferum* taxa and were sampled from a range of environments across Africa. In total, the resource consists of 13 populations and a total of 1,224 lines. The population has been genotyped with diversity array technology (DArT) markers that produced 42,372 unique single nucleotide polymorphism (SNP) markers covering the genome. We determine the utility of the resource for high resolution mapping of complex traits by demonstrating that the exotics contain unique alleles for some example adaptive trait loci and by using the population for genome-wide association study (GWAS). The resource should provide useful material for plant breeders attempting to deal with the challenges generated by climate change.

1 | INTRODUCTION

Sorghum is an important source of food and fodder in water-limited agricultural production systems and, as such, is a staple for over 500 million people worldwide. The crop is grown predominantly in water-limited environments in the tropics and subtropics. Recent modelling suggests that climate change is expected to increase incidence of abiotic stress,

destabilizing sorghum production (Burke, Lobell, & Guarino, 2009, Lobell et al., 2015). To enhance the productivity and sustainability of the sorghum grain crop, breeders will need access to novel genetic variation in adaptive traits.

The publication and annotation of a reference genome sequence for sorghum, based on the elite grain type BTx623 (Paterson et al., 2009), has fast-tracked gene and novel sequence variants discovery and has enabled resequencing studies (Mace et al., 2013, Zheng et al., 2011) to identify millions of sequence variants and signatures of domestication. Like all cultivated crops, sorghum experienced a genetic bottleneck associated with domestication and the resequencing studies identified a dramatic restriction in the diversity of modern, cultivated lines. As a result, cultivated germplasm lines sample only a small proportion of the genetic diversity of the species constrained by the set of genes selected by the first farmers. These include genes that were valuable in early agriculture, alleles that were selected by chance, and new mutations that have arisen subsequently. Subsequent changes in climate and agricultural systems mean that this initial set of genes selected by the first farmers will not contain all of the variation required for our current challenges. With this in mind, strong consideration should be given to exploring and exploiting the variation in CWRs in which sorghum is particularly rich.

All grain sorghum belongs to *S. bicolor*, a native of Africa that has three subspecies: *bicolor*, *verticilliflorum*, and *drummondii*. Subspecies *verticilliflorum* is the wild progenitor species with seed <1 mm thick that shatter. Subspecies *bicolor* incorporates all the cultivated types characterized by nonshattering habit, larger seed, and usually less tillering than the wild types. Subspecies *drummondii* incorporates all material derived from natural crossing between *bicolor* and *verticilliflorum* types (Wiersema & Dahlberg, 2007) and, as such, can display almost any combination of traits from parents. There have been at least two, possibly more, domestications of sorghum. The original domestication is thought to have occurred in East Africa—somewhere in Ethiopia or possibly Sudan. The group distinguished by the name *margaritiferum*, previously regarded as a subrace, has been identified as the progeny of a second, more recent, domestication in West Africa (de Alencar Figueiredo et al., 2008, Deu et al., 1995, Mace et al., 2013, Sagnard et al., 2011). These plants have many features of the wild species and are grown separately from other sorghum, named as a different crop, and prepared in different foods.

Crop wild relatives have been used in many species as sources of novel resistance to biotic stresses but much less so for adaptation to abiotic stresses (Hajjar & Hodgkin, 2007, Rao, Reddy, & Bramel, 2003). One of the reasons for this is the difficulty in screening wild species germplasm for these traits in a way that is consistent with identifying variation that will be impactful in modern breeding as a result of the confounding effects of plant architecture and phenol-

Core Ideas

- A novel nested association mapping population based on sorghum wild relatives is described.
- The novel resource developed is suitable for GWAS trait dissection activities.
- The resource provides useful alleles for breeding germplasm adapting to climate change.

ogy (Jordan, Mace, Cruickshank, Hunt, & Henzell, 2011). As a result, even if useful variation for complex traits exist, the cost of identifying and using it is beyond the resources of most breeding programs. In sorghum, a major step forward in the use of exotic germplasm was made by the sorghum conversion program (Stephens, Miller, & Rosenow, 1967), which introgressed genes for photoperiod insensitivity and short stature into unadapted tropical germplasm to make it useful for sorghum breeders working in temperate regions. This material has been widely used, but one of the implications of this program was that genetic diversity was dramatically reduced on sorghum chromosomes 6 and 10 that carried major genes for height and maturity (Morris et al., 2013, Thurber, Ma, Higgins, & Brown, 2013). More recently nested association mapping (NAM) populations have been proposed as a powerful tool for elucidating the genetic architecture of traits (Yu, Holland, McMullen, & Buckler, 2008). When NAM populations are generated with a single backcross to the adapted parent (BCNAM), their utility for introgression of novel traits into an elite genetic background is even greater for crop improvement programs (Jordan et al., 2011). Both NAM and BCNAM studies have been used in maize (*Zea mays* L.), sorghum, rice (*Oryza sativa* L.), wheat (*Triticum aestivum* L.) and other crops to discover the genetic architecture of important traits (Bouchet et al., 2017, Buckler et al., 2009, Fragoso et al., 2017, Tao et al., 2020). With the availability of high-density SNP data, these resources have the potential to identify candidate genes for target traits. In the Australian grain sorghum prebreeding program, several wild accessions were used as exotic parents in the previously developed BCNAM populations (Jordan et al., 2011). These populations contained at least the expected 25% exotic parent DNA but conformed to the nonshattering, 3-dwarf height and photoperiod-insensitive phenotype desired for grain sorghum breeding in Australia and proved to have considerable value in identify novel alleles for quantitative traits (Tao et al., 2020)

The project 'Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives' is an investment in greater collection of and prebreeding with CWRs to improve adaptation to climate change. Knowing the value of BCNAM populations and the genetic diversity present in sorghum CWRs, we have produced a BCNAM

TABLE 1 Identifiers and origins of parents used in the development of a sorghum crop wild relative introgression panel

Donor parents					
Primary identifier	Current taxonomy	Other identifiers	Country	Latitude, Longitude °	Elevation m
AusTRCF317030	*Sorghum* sp. (previously *S. versicolor*)	–	South Africa	–	–
AusTRCF317961[a,b]	*S. bicolor* subsp. *verticilliflorum*	–	Australia	–	<20
IS3121[b]	*S. bicolor* subsp. *verticilliflorum*	K.50188, PI226096	Kenya	–	–
PI300119[a]	*S. bicolor* subsp. *verticilliflorum*	AJO-0464	South Africa	−25.52, 31.55	396
PI330272[a]	*S. bicolor* subsp. *drummondii*	R-244, BL1067	Ethiopia	13.04, 38.84	1500
PI525695[a,b]	*S. bicolor* subrace *margaritiferum*	IS 25806, SG4767	Mali	11.9, −8.1	–
PI532566[b]	*S. bicolor* subsp. *verticilliflorum*	TCD070	Chad	13.50, 16.50	320
PI535995	*S. bicolor* subsp. *verticilliflorum*	C120	Cameroon	10.57, 13.58	970
PI536008[b]	*Sorghum purpureosericeum*	C048	Cameroon	11.25, 14.67	300
Adapted parents					
Primary identifier	Traits		Pedigree	Country	Breeding program of origin
Macia[a]	2-dwarf plant type, midge resistant, inbred cultivar		SAS3220/M91057	Zimbabwe	ICRISAT
QL36	3-dwarf plant type, midge resistant, inbred restorer		QL27/RTx2767	Australia	Queensland Government

[a]Resequencing data generated previously (Mace et al., 2013).
[b]De novo genome assemblies developed.

panel of 1,224 lines using exotic parents from the *verticilliflorum*, *drummondii*, and *margaritiferum* taxa. Using simple phenotypic traits as examples, we show the utility of this panel for identifying and mapping novel variation.

2 | MATERIALS AND METHODS

2.1 | Population development

In winter 2015, 18 biparental crosses were made in the glasshouse. Nine CWR accessions (Table 1) were successfully crossed with the two recurrent parents QL36 and Macia. To generate backcross populations, the F_1 hybrids were sown in a photoperiod manipulation facility to simulate short days and induce flowering in the 2015–2016 summer. Once floral induction commenced, hybrids were transplanted to an outdoor crossing block (at Warwick in South Queensland) where the recurrent parents were planted earlier. A combination of mechanical and subsequent hand sowing was used to increase

the spread of flowering in the recurrent parents. For the QL36 backcrosses, we were able to use an isoline segregating for the *ms3* gene. For Macia backcrosses, the plastic bag emasculation technique (Schertz & Clark, 1967) was used; however, this was not successful for all crosses.

In the subsequent inbreeding generations selection was minimized but photoperiod sensitivity and necessary selection against shattering had an impact both on number of progenies advanced and which individual plants successfully reproduced within progenies. Bulk BC_1F_1 and F_2 populations were grown at Ayr in North Queensland through winter 2016. In total, 2,198 $BC_1F_{1:2}$ and $F_{2:3}$ progenies were advanced. This included more progenies from QL36 than Macia crosses, but all nine CWR exotics had progeny advanced. Approximately 12% of selections came from shattering heads but were advanced one more generation to allow the possibility of more recombination near shattering genes.

Single-plant progeny rows were grown in at Warwick in the 2016–2017 summer. This allowed further selection against shattering and photoperiod sensitivity. Single heads were

TABLE 2 Cross codes and pedigrees of all crosses with >20 lines in the crop wild relative backcross nested association mapping panel

Duplicate code	Cross code	Pedigree	No. of lines panel
R16011	R16011	Macia//PI300119/Macia	54
R16016	R16016	Macia//PI536008/Macia	54
R16008	R16008	F2_ms3*8_QL36//AusTRCF317030/QL36	243
R16001	R16001	F2_ms3*8_QL36//AusTRCF317961/QL36	42
R16002D{	R16002	F2_ms3*8_QL36//PI300119/QL36	34
	R16019	PI300119/QL36//QL36	31
R16005D{	R16005	F2_ms3*8_QL36//PI330272/QL36	50
	R16022	QL36//PI330272/QL36	17
R16003	R16003	F2_ms3*8_QL36//PI532566/QL36	161
R16004	R16004	F2_ms3*8_QL36//PI535995/QL36	185
R16007	R16007	F2_ms3*8_QL36//PI536008/QL36	114
R15183	R15183	PI330272/Macia	21
R15184	R15184	IS3121/QL36	40
R15190	R15190	PI525695/QL36	63
R15180	R15180	PI535995/QL36	46

harvested from 1,779 progenies and advanced to a nursery at Ayr in winter 2017, and BC_1F_4 and F_5 progenies were harvested in late 2017.

2.2 | The crop wild relative panel

The final panel genotyped and lodged with the Australian Grains Genebank has 1,224 individual lines. These come from 14 backcross populations and eight F_2 populations involving nine exotic parents and two adapted elite lines, with the backcross derived populations representing >80% of the panel. Five of the exotic parents have been used successfully in crosses with both adapted lines. For comparisons of populations, a subset of 1,155 lines was used; from the total of 22 populations, seven were excluded because they had <20 progeny and two pairs of populations with the same parentage but different crossing method were aggregated (Table 2). Each exotic has at least one representative population in the subset of the 13 larger populations.

2.3 | Characterization experiments

In the 2017–2018 summer, the panel was grown for seed increase at Warwick and two partially replicated preliminary characterization experiments were grown at Gatton (humid subtropical, −27°33′, 152°20′) and Emerald (semiarid tropical, −23°32′, 148°11′).

Trials were laid out in a row−column design with 1,231 genotypes at Emerald with 820 of these planted at Gatton. Each site had a partially replicated design (Cullis, Smith, &

Coombes, 2006) generated as a two-site multienvironment trial design using the R optimal design package OD (Butler, 2019). At each site, 30% of the genotypes had two plots with the remaining genotypes having a single plot, where different genotypes were replicated at each site. At each site, there was a check genotype that had 33 replicates at Emerald and 18 replicates at Gatton. The two recurrent parents had 4 replicates at each site. In Australian regional trials, partially replicated experiments of this nature predict height and days to flower with heritabilities usually over 70% and always over 50%.

Plant height of main stem and a rust rating were recorded at both the Gatton and Emerald sites. Rust in both environments was the result of natural infection. Sorghum rust in Eastern Australia is known to have multiple pathotypes (White, Ryley, George, & Kong, 2015) but it was beyond the resources of these experiments to identify how many or which pathotypes were present in the two environments. Because of constraints, days to flower were recorded at Emerald only. Each trait was analyzed separately using a linear mixed model for the data from both sites combined for plant height and rust and a single site for days to flower. All sites and traits were assessed for possible spatial effects resulting from extraneous field effects and neighbor effects, and these were included in the model as necessary. Best linear unbiased predictors (BLUPs) were produced for each genotype for each trait and averaged across sites. All statistical analyses were conducted using R (www.R-project.org). Linear mixed models were fitted using the ASRemL-R package (Butler, Cullis, Gilmour, & Gogel, 2009). Generalized heritabilities were calculated for each trait and site using the average standard errors and the genetic variance as discussed in Cullis et al. (2006).

2.4 | Genotyping

The BCNAM panel was genotyped using medium to high-density genome-wide SNPs provided by Diversity Arrays Technology Pty Ltd (www.diversityarrays.com). DNA was extracted from bulked young leaves of five plants from a plot of each genotype in the panel from the BC_1F_4 and F_5 generations using a previously described CTAB method (Doyle & Doyle, 1987). The DNA samples were then genotyped following an integrated DArT and genotyping-by-sequencing methodology, which involves complexity reduction of the genomic DNA to remove repetitive sequences using methylation sensitive restriction enzymes prior to sequencing on Next Generation sequencing platforms. The sequence data generated were then aligned to version v3.1.1 of the sorghum reference genome sequence (McCormick et al., 2018) to identify SNPs.

Parental allelic contributions were determined by identifying homozygous polymorphic SNPs between the two parents for each population separately and calculating the number of progeny within each populations that carried the exotic allele at each SNP locus. Moving means were then calculated based on averaging the exotic allele proportions of all the SNPs within a 2-cM window using the consensus map genetic linkage distances (Mace et al., 2009) for each population. The production of color-coded exotic genome introgression heatmaps was performed in R (www.R-project.org).

2.5 | Genome wide association study

Genome-wide association study of plant height was conducted using FarmCPU (Liu, Huang, Fan, Buckler, & Zhang, 2016) with 42,372 imputed SNP filtered on minor allele frequency >0.01 with all other settings as standard. The across-site BLUPs were used as phenotypic data. The pedigree of this population was used as a covariate in the analysis to control population structure. Threshold for defining significant associations was set using the Bonferroni correction $(0.05/42372 = 1.18 \times 10^{-6})$. The SNPs identified as being significantly associated with plant height were further compared with reported plant height QTL extracted from the sorghum QTL atlas (Mace et al., 2019).

3 | RESULTS

3.1 | Novel genetic variation in crop wild relatives

Whole-genome resequencing data (Mace et al., 2013) has been generated previously for six of the 11 parents (five exotic

and one of the adapted lines) (Table 1). In addition to the exotic parental lines, the resequencing dataset included an additional three wild and weedy genotypes, 18 landraces, and 18 improved inbred lines, encompassing all racial groups within cultivated sorghum and identified ~10 M sequence polymorphisms. Principal components analysis (Figure 1a) demonstrated that the cultivated types clustered very closely together, with *S. bicolor* subsp. *verticilliflorum* separating on principal component 1 followed by *S. margaritiferums* on principal component 2, explaining 55% of the variation. The increased level of diversity among the wild and weedy genotypes was significantly higher than the cultivated types ($P < 2.2 \times 10^{-16}$ by paired *t*-test). This pattern is also clearly indicated through the total number of SNPs identified in each group (Figure 1b), where the wild- and weedy-specific alleles (34%) were more abundant than improved inbred specific alleles (8%) and landrace specific alleles (18%). The study also demonstrated that the average total number of SNPs and insertions and deletions per genotype was highest in the wild and weedy genotypes (one SNP per 763 bp) in comparison to the landraces (one SNP per 1,282 bp) and the inbred lines (one SNP per 1,543 bp). Of particular note, the *margaritiferum* genotypes (including CWR exotic parent PI525695) had approximately twice the SNP density of the landrace group (one SNP per 691 bp vs. one SNP per 1,282 bp, respectively). The wild and weedy genotypes, including the *margaritiferums*, were also shown to contain more functional diversity with, on average, fivefold more large-effect SNPs within genes vs. members of the improved inbred and landrace groups. Figure 2 details the distinct allelic variants identified from the CWR exotic parents at three major adaptation loci including the aluminum tolerance loci Alt_{SB} (Magalhaes et al., 2007), the deeper rooting loci *DRO1*(Arai-Sanoh et al., 2014) and a drought adaptation loci, *PIN5*, the candidate gene for the stay-green QTL (*stg2*) (A. K. Borrell, personal communication, 2018). In addition, further resequencing data and de novo genome assemblies are being generated for six of the exotic CWRs (Table 1), which will enable gene content variation to be explored in more detail.

3.2 | Recovery of donor genotype in the backcross nested association mapping panel

In total, 42,372 genome-wide SNPs were identified in the panel (data available via https://ics.hutton.ac.uk/cwr/sorghum/), corresponding to an average of one SNP per 15 kb, and were used to investigate parental allele segregation in the 13 populations with >20 lines (Table 2). Figure 3 shows the distribution of the average proportion of exotic genome alleles present in the 13 populations. The overall mean percentage exotic parent genome was 26% across all backcross

FIGURE 1 (a) Principal component analysis of sorghum accessions from Mace et al. (2013) indicating novel diversity available in wild and weedy accessions. (b) Distribution of novel single nucleotide polymorphism alleles across germplasm groups included in Mace et al. (2013)


FIGURE 2 Gene trees highlighting novel variation identified in the crop wild relatives used in the panel (highlighted in red) extracted from a whole genome resequencing data set (Mace et al., 2013) for (a) Alt_{SB}, (b) $DRO1$, and (c) $PIN5$

derived populations and 48% across the F_2 derived populations; however, this varied considerably between populations. In the F_2 populations, the highest recovery of exotic parent genotype occurred in R15184 (61%), with the lowest being in R15190 (41%) (Figure 3). Six of the backcross populations had average donor contribution below the theoretical 25%, with the minimum being 17% (R16016), but all populations had individual lines with higher proportions of the donor genome. Three backcross populations had average donor contribution above 25%, with R16005D, the progeny of PI330272 and QL36, having an average exotic contribution of 42%. Figure 4 presents heat maps of the average introgression in chromosomes SBI-01 and SBI-06 for the larger populations. The figure highlights the location of genes known to influence shattering status, photoperiod sensitivity, and plant height. These are all known to be important in adaptation to cultivation in subtropical environments and would be expected to be under selection particularly in the summer nurseries employed in breeding this material. Significant contrasts in

degree of introgression are apparent both between populations and between chromosomes. The F_2 populations and the R16005D backcross population have a much higher proportion of donor genotype on SBI-01 than other populations. All populations show an increased proportion of the adapted genotype around the *Sh1* shattering gene (Lin et al., 2012). There is more variation in degree of introgression around the location of the maturity locus *Ma3*, with R16016, R160007, and R16001 showing the highest proportion of adapted genotype in this genomic region. On SBI-06, there is a disparity between high concentration of the adapted genotype on the short chromosome arm carrying the major maturity and height loci—*Ma1*, *Ma6*, and *Dw2*—and much more introgression of the donor genotype on the long arm. Both SBI-01 and SBI-06 reflect the much greater level of donor genotype in population R15184, already noted on a whole-genome basis. The distribution of the average proportion of exotic genome alleles across all chromosomes is presented in Supplemental Figure S1.

FIGURE 3 Variation of the genome-wide exotic allele introgression across 13 populations
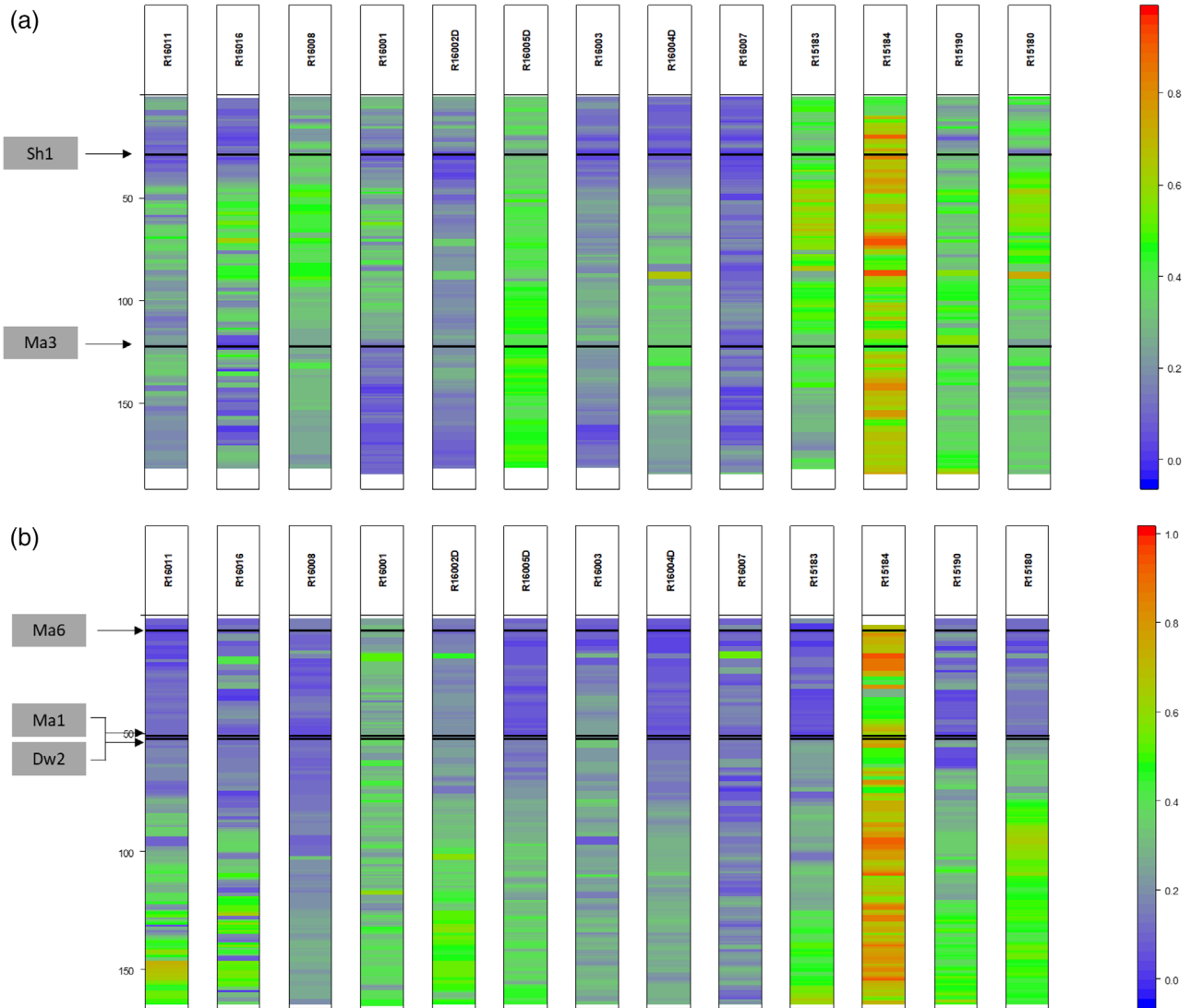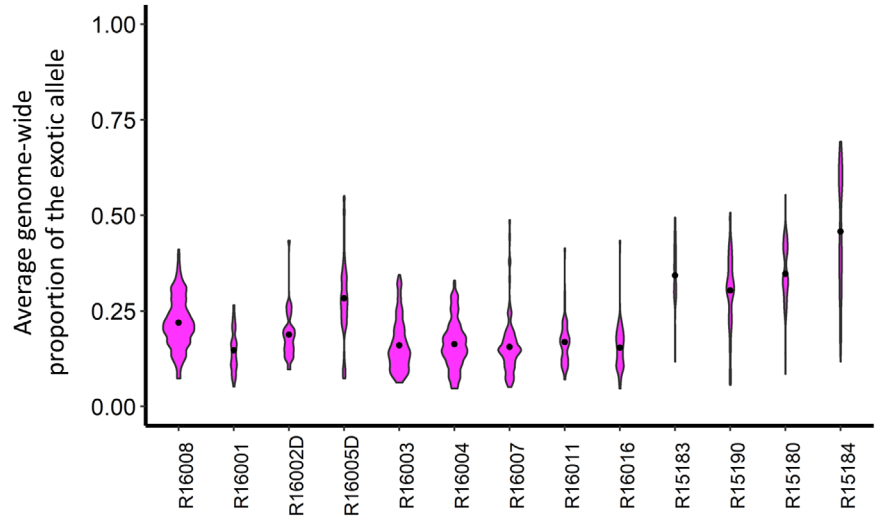
FIGURE 4 Heat map indicating the average exotic genome introgression rate on two chromosomes: (a) SBI-01 and (b) SBI-06. Major effect genes of relevance are highlighted

3.3 | Characterization experiments

The phenotypic characterization of the panel found significant genetic variation and heritability in a range of traits and showed that a wide range of plant types were conserved: from fine-stemmed 3-dwarf types to robust 1- and 2-dwarf types with thick stems. Plant height had heritabilities >90% in both environments and a correlation between sites of 0.88; the rust rating had heritabilities of 66% at Gatton and 51% at Emerald where the disease pressure was lower. Days to flower had a heritability of 80% at Emerald (Supplemental Table S1A and S1B; Supplemental Figure S2).

Figure 5 presents the characterization variation in each of the 13 larger populations. For plant height and rust, the BLUPs were averaged across sites. The majority of lines in the panel, regardless of their adapted parent, are in the plant height range between 1 and 2 m, but each pedigree also has some lines shorter and taller than this range including some lines in the 2.5–3 m range. The rust reaction of the panel is more constricted with the rust rating BLUPs for most lines between 4 and 5 on this discrete scale. Only a small number of lines had a predicted rating below 4 and an even smaller number had a score over 6.

Days to flower has a different distribution with most of the panel flowering in a range 5–10 d earlier than either of the adapted parents. Also, populations differed from one another more markedly for this trait. There was no general tendency for backcross populations to be closer to the adapted parents than F_2 populations. Nor was there any general tendency for exotic parents to produce populations consistently earlier or later, although the latest F_2 and the latest backcross population were both the product of crosses with PI330272.

3.4 | Genome-wide association study of plant height

To test the utility of this panel for dissecting the genetic architecture of agronomic traits, we conducted GWAS for plant height using 42372 imputed SNP with minor allele frequency >0.01. The GWAS identified strong signal for the trait, reflecting the significant genetic variance and high heritability of plant height in this resource, and identified 17 significantly associated SNPs (Figure 6; Supplemental Table S2). The top three most significant SNPs corresponded with previously reported plant height genes in sorghum, including *Dw1* on chromosome 9, *Dw2* on chromosome 6, and *Dw3* on chromosome 7, supporting the reliability of using this population to dissect agronomic traits. The GWAS also detected seven other previously identified height QTL reported in the sorghum QTL atlas (Mace et al., 2019). Novel loci, not yet reported in previous studies, were also identified

(Supplemental Table S2), indicating the exotic resource used in this population may provide genetic variation not previously recorded in studies with cultivated sorghum.

4 | DISCUSSION

Climate change is placing increasing pressure on agriculture, causing changes in the frequency of abiotic and biotic stresses potentially destabilizing critical cropping systems. An important strategy for dealing with this challenge is to develop new varieties of crops that are more tolerant to these stresses. A key limitation to plant breeders' capacity to perform this task is their ability to access useful genetic variation for some of these traits in easily accessible breeding material. Domestication of modern crops resulted in a dramatic reduction in genetic diversity, and for many years CWRs have been seen as a source of novel genes for crop improvement but their use has been limited to major genes because of the restrictions imposed by difficulties in phenotyping and linkage drag. Recent advances in sequencing and gene mapping have created opportunities to detect useful allelic variants from wild germplasm, while techniques such as marker-assisted selection, genetic modification, and gene editing are providing more rapid ways of deploying these alleles in elite germplasm. Here we describe a genetic and genomic resource we have developed for exploiting genetic variation from CWRs in sorghum.

4.1 | Sorghum wild relatives are a source of novel genetic variation for crop improvement

Cultivated sorghum is the result of at least two domestication events. Given the high degree of genetic diversity contained in wild sorghum (Mace et al., 2013) and that it is found over a wide range of environments in Africa and beyond, it is reasonable to assume that the CWRs hold novel genetic variation that confers adaptation to a wide range of abiotic and biotic challenges (Tao, Zhao, Mace, Henry, & Jordan, 2019). Genotyping, sequencing, and genetic mapping all suggest that sorghum CWRs contain both novel alleles of genes known in cultivated sorghum and genes at loci unknown in cultivated sorghum (Campbell et al., 2016, Tao et al., 2017, Tao et al., 2018). In the former case, the novel alleles from CWRs often segregate at very low frequencies in most panels and populations developed previously, making them very hard to detect. In the latter case there may be novel genes in structural variations not present in cultivated sorghum and, hence, these are also very difficult to detect. We have demonstrated that this panel is a resource for finding both these forms of novel variation.
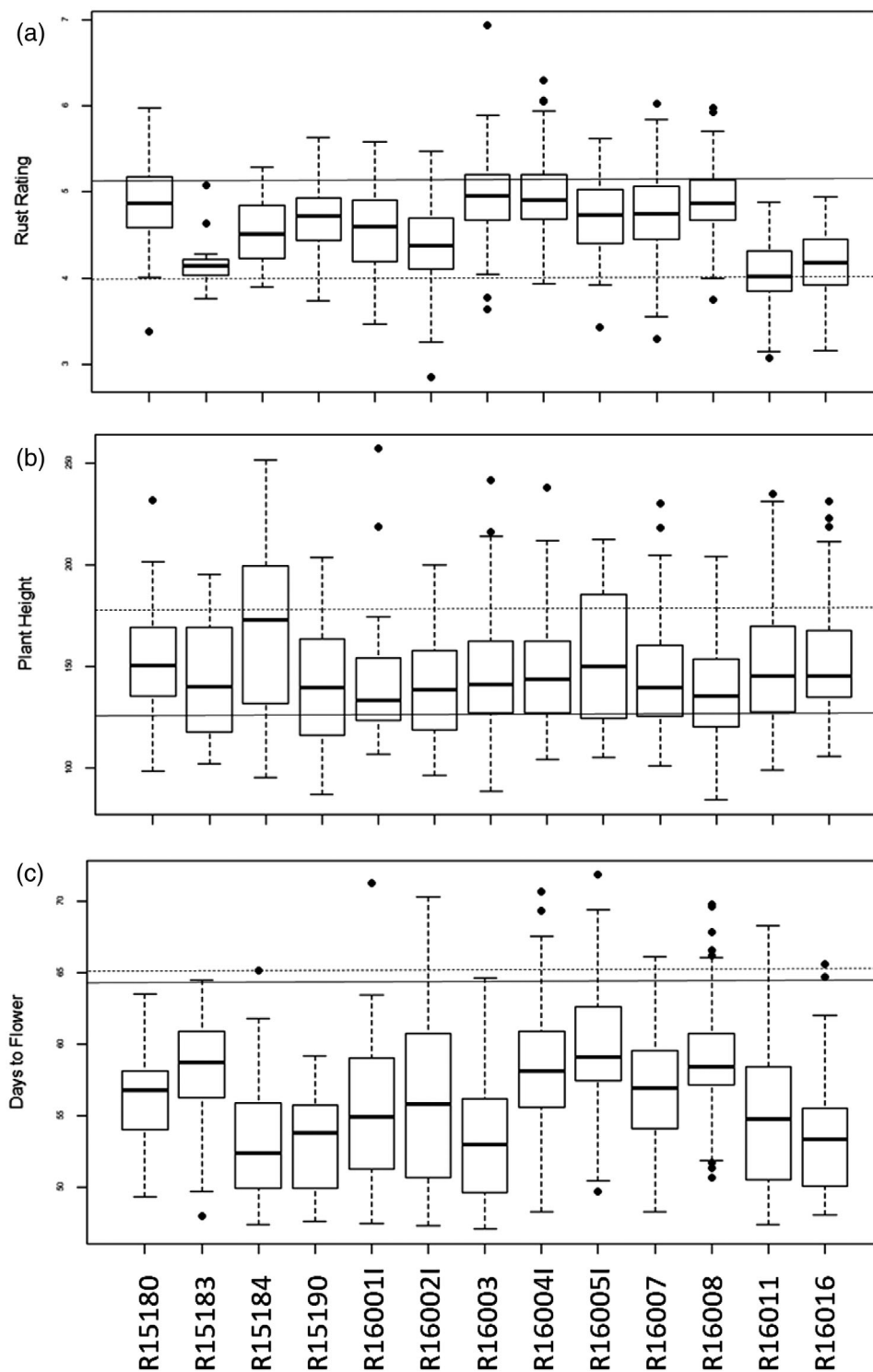
FIGURE 5 Box plots of best linear unbiased projections (BLUPs) of the sorghum populations for (a) rust ratings data based from Emerald and Gatton sites. Rating: 1 = highly resistant and 9 = susceptible. (b) plant height data based from Emerald and Gatton sites. (c) Days to 50% anthesis based on data from Emerald Qld. Horizontal lines represent the BLUPs of the recurrent parents with the solid line being QL36 and the dotted line being Macia
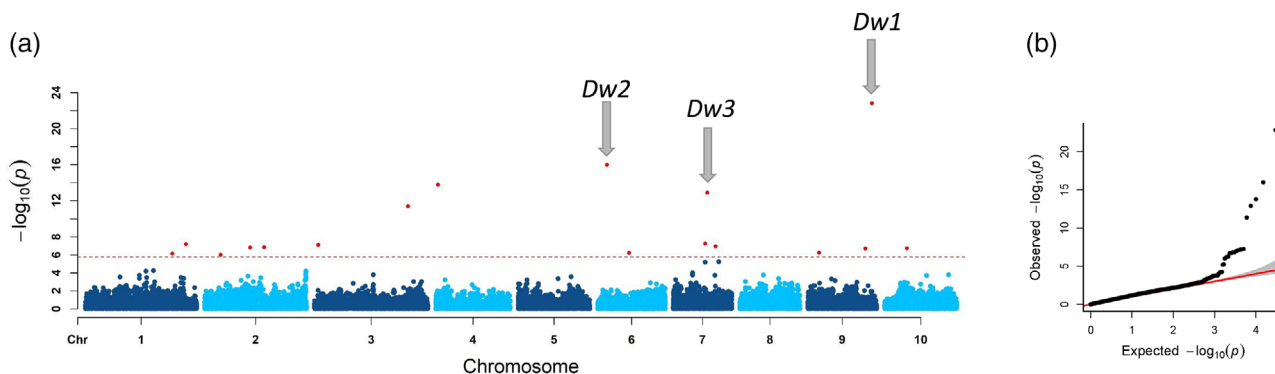
FIGURE 6 (a) Manhattan plot displaying results of plant height genome-wide association study (GWAS). The red dashed line indicates the significance threshold (p-value $< 1.18 \times 10^{-6}$) based on the Bonferroni correction. Known plant height genes colocating with significant SNPs are highlighted. (b) The quantile–quantile plot of plant height GWAS

Selection for adaptation to subtropical and temperate zones has led to documented reduction in diversity in particular genomic regions of cultivated sorghum (Morris et al., 2013, Thurber et al., 2013). This is due to selection against photoperiod sensitivity genes and height genes, with SBI-06 shown to be the most affected. In our resource, there were only a very limited number of regions genome-wide where the adapted allele was selected at much higher frequencies than expected (>90%) across multiple populations, with the short-arm of SBI-06 showing this extreme segregation distortion in 10 out of the 13 populations. However, even in this genomic region, there were individuals within eight of the 13 populations that contained the exotic allele at multiple SNPs (>400 in total) on the short-arm of SBI-06, indicating that substantial novel variation has been retained on SBI-06 and elsewhere. Selection for the adapted allele was also in evidence around the shattering gene, *Sh1*, on SBI-01 (Figure 4). However, higher frequencies than expected of the exotic allele were observed in some populations (e.g. R15184), possibly an indication of multigenic influence on the shattering trait. Only one of our CWR parents is presently classified as subspecies *drummondii,* which by definition is a wild sorghum with introgression from cultivated sorghum but it is likely that other wild parents have had historic introgression from cultivated sorghum. As such, these weedy relatives with a history of admixture may represent a valuable resource for breeding as a result of the additional cycles of recombination between cultivated and wild parents that they contain.

The genome-wide introgression of the exotic alleles was broadly in line with theoretical expectations based on the population development strategy, that is, 25% for the backcross derived population and 50% for the F_2 populations. However, within each population, there was considerable variation both within individuals and in specific genomic regions. Within the backcross derived populations, the exotic allele was segregating at a higher frequency than the adapted allele at ~1% of all SNPs genome-wide (480 SNPs); this increased 10-fold in the

F_2 derived populations with the exotic allele occurring at a higher frequency in 11% of the SNPs genome wide. The high frequency of the exotic allele was nonrandomly distributed, with particular genomic regions having high proportions of the exotic allele across multiple populations. For example, the long arm of chromosome SBI-03, which contains known genes and QTL for adaptive traits, including drought tolerance (e.g. *DRO1* and *PIN5*) and aluminum tolerance (Alt_{SB}), was found to contain a high frequency of exotic alleles in eight of the 13 populations including the *margaritiferum*-derived population (R15190) (Supplemental Figure S1). This is in line with findings from Mace et al. (2013) who observed that the western African origins of the *margaritiferums*, where poor soils and unreliable rainfall are noteworthy, suggest a potentially valuable genetic resource for low soil pH and toxic aluminum tolerance. The centromeric region on chromosome SBI-05, which contains the stay-green QTL (*stg4*), was also found to contain a high frequency of exotic alleles in nine of the 13 populations (Supplemental Figure S1). This highlights the value of this resource for novel, exotic alleles associated with adaptive traits, providing breeders and researchers with a key resource to navigate and exploit this diversity in sorghum and other species.

4.2 | Sorghum BCNAM population is a valuable tool for identifying genes contributing to key adaptive traits

The sorghum CWR NAM populations were developed following two different strategies: the first was a backcross strategy crossing nine different exotic parental lines to two adapted elite lines (Macia and QL36) and backcrossing to the adapted lines with selection applied against shattering and photoperiod sensitivity; and the second was a biparental crossing strategy, advanced to the F_2 generation and with selection again applied against shattering and photoperiod sensitivity. In both

cases, some selection was applied to constrain variation in height and maturity to ensure the populations were amenable to phenotyping. The more constrained diversity in phenology of the progeny indicates the success of the selection strategy and that extreme variation for flowering time has been minimized. This is a critical feature of the panel because the breeding value of any new allele for a quantitative trait is context dependent (Jordan et al., 2011). The application of targeted selection in this panel provides a relevant genetic context to evaluate the impact of exotic alleles, thus enabling effective comparisons for complex traits across multiple environments. The phenotypic diversity captured for height and rust resistance demonstrates that the selection imposed has not had negative consequences for capturing the exotic genome as also evidenced by the genotypic investigation of exotic allele introgression genome wide. The final population size of over 1,200 represents one of the largest structured mapping resources developed for sorghum.

To demonstrate the utility of this resource for association mapping, we conducted a GWAS for height—an important adaptive trait—with high heritability (>90% across the two environments included in our study). The GWAS identified 17 SNPs associated with plant height, of which, the three most significant SNPs corresponded with previously identified major effect genes for height (*Dw1*, *Dw2*, and *Dw3*). The most significant SNP was within 125 kb of *Dw1* (Hilley, Truong, Olson, Morishige, & Mullet, 2016), indicating that this resource provides the resolution and power required for complex trait dissection. The higher rate of linkage disequilibrium (LD) decay in wild and weedy genotypes (LD decaying to 50% by 7 kb) in comparison to cultivated lines (LD decaying to 50% by 20 kb) in sorghum, as observed in Mace et al. (2013), provides increased opportunities for gene-level resolution in association mapping when coupled with high-density genotyping. The availability of whole-genome sequence data, either through resequencing or de novo assembly, for all of the exotic parents used in this resource will enable high-density SNP imputation across the progeny therefore allowing the promised high mapping resolution to be achieved. The ongoing de novo assembly of exotic parents will further facilitate gene discovery using this resource given the prevalence of present–absent variation of genes within a species (Tao, Jordan, & Mace, 2019, Tao et al., 2019).

4.3 | Source of new alleles for breeding for future climate

The material developed and described in this paper incorporates novel alleles from multiple exotic parental lines into relevant genetic backgrounds for sorghum breeders. The resource has the potential to deliver outcomes not only for applied breeding but also for research purposes. The sorghum CWR

NAM panel is maintained at the Australian Grains Genebank and is available for use under standard material transfer agreements. The phenotypic characterization data and SNP data are managed in the sorghum Germinate database (Raubach et al., 2020). The results presented emphasize what can be gained from incorporating alleles from diverse genetic backgrounds, particularly abiotic and biotic stress related genes. The resolution of specific differences between alleles enables us to rapidly identify and select alleles and loci that may prove advantageous. These can be rapidly introgressed into new backgrounds using conventional phenotypic selection or combined with genome-editing technologies to target specific SNPs. The sorghum CWR panel is an unmatched resource for the improvement of sorghum, providing researchers with a key resource to navigate and exploit this diversity in sorghum.

CONFLICT OF INTEREST
The authors have no conflict of interest to report.

ORCID
Alan W. Cruickshank https://orcid.org/0000-0002-7982-1746

REFERENCES

Arai-Sanoh, Y., Takai, T., Yoshinaga, S., Nakano, H., Kojima, M., Sakakibara, H., … Uga, Y. (2014). Deep rooting conferred by *DEEPER ROOTING 1* enhances rice yield in paddy fields. *Scientific Reports*, *4*, 5563. https://doi.org/10.1038/srep05563

Bouchet, S., Olatoye, M. O., Marla, S. R., Perumal, R., Tesso, T., Yu, J., … Morris, G. P. (2017). Increased power to dissect adaptive traits in global sorghum diversity using a nested association mapping population. *Genetics*, *206*, 573–585. https://doi.org/10.1534/genetics.116.198499

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., … Glaubitz, J. C. (2009). The genetic architecture of maize flowering time. *Science*, *325*, 714–718. https://doi.org/10.1126/science.1174276

Burke, M. B., Lobell, D. B., & Guarino, L. (2009). Shifts in African crop climates by 2050, and the implications for crop improvement and genetic resources conservation. *Global Environmental Change*, *19*, 317–325. https://doi.org/10.1016/j.gloenvcha.2009.04.003

Butler, D., Cullis, B. R., Gilmour, A., & Gogel, B. (2009). *ASReml-R reference manual*. Brisbane, Australia: The State of Queensland, Department of Primary Industries and Fisheries.

Butler, D. (2019). od: Generate optimal experimental designs. R package version 2.0.0

Campbell, B. C., Gilding, E. K., Mace, E. S., Tai, S. S., Tao, Y. F., Prentis, P. J., … Godwin, I. D. (2016). Domestication and the storage starch biosynthesis pathway: Signatures of selection from a whole sorghum genome sequencing strategy. *Plant Biotechnology Journal*, *14*, 2240–2253. https://doi.org/10.1111/pbi.12578

Cullis, B. R., Smith, A. B., & Coombes, N. E. (2006). On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological, and Environmental Statistics*, *11*, 381. https://doi.org/10.1198/108571106X154443

de Alencar Figueiredo, L. F., Calatayud, C., Dupuits, C., Billot, C., Rami, J. F., Brunel, D., … Glaszmann, J. C. (2008). Phylogeographic evidence of crop neodiversity in sorghum. *Genetics*, *179*, 997–1008. https://doi.org/10.1534/genetics.108.087312

Deu, M., Hamon, P., Chantereau, J., Dufour, P., D'Hont, A., & Lanaud, C. (1995). Mitochondrial DNA diversity in wild and cultivated sorghum. *Genome*, *38*, 635–645. https://doi.org/10.1139/g95-081

Doyle, J., & Doyle, J. (1987). Genomic plant DNA preparation from fresh tissue. *Phytochemical Bulletin*, *19*, 11–15.

Fragoso, C. A., Moreno, M., Wang, Z., Heffelfinger, C., Arbelaez, L. J., Aguirre, J. A., … Zhao, H. (2017). Genetic architecture of a rice nested association mapping population. *G3: Genes, Genomes, Genetics*, *7*, 1913–1926. https://doi.org/10.1534/g3.117.041608

Hajjar, R., & Hodgkin, T. (2007). The use of wild relatives in crop improvement: A survey of developments over the last 20 years. *Euphytica*, *156*, 1–13. https://doi.org/10.1007/s10681-007-9363-0

Hilley, J., Truong, S., Olson, S., Morishige, D., & Mullet, J. (2016). Identification of *Dw1*, a regulator of sorghum stem internode length. *PLoS ONE*, *11*, e0151271. https://doi.org/10.1371/journal.pone.0151271

Jordan, D., Mace, E., Cruickshank, A., Hunt, C., & Henzell, R. (2011). Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding program. *Crop Science*, *51*, 1444–1457. https://doi.org/10.2135/cropsci2010.06.0326

Lin, Z. W., Li, X. R., Shannon, L. M., Yeh, C. T., Wang, M. L., Bai, G. H., … Yu, J. M. (2012). Parallel domestication of the *Shattering1* genes in cereals. *Nature Genetics*, *44*, 720–724. https://doi.org/10.1038/ng.2281

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genetics*, *12*, e1005767. https://doi.org/10.1371/journal.pgen.1005767

Lobell, D. B., Hammer, G. L., Chenu, K., Zheng, B., McLean, G., & Chapman, S. C. (2015). The shifting influence of drought and heat stress for crops in northeast Australia. *Global Change Biology*, *21*, 4115–4127. https://doi.org/10.1111/gcb.13022

Mace, E., Innes, D., Hunt, C., Wang, X., Tao, Y., Baxter, J., … Jordan, D. (2019). The Sorghum QTL Atlas: A powerful tool for trait dissection, comparative genomics and crop improvement. *Theoretical and Applied Genetics*, *132*, 751–766. https://doi.org/10.1007/s00122-018-3212-5

Mace, E. S., Tai, S., Gilding, E. K., Li, Y., Prentis, P. J., Bian, L., … Han, X. (2013). Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nature Communications*, *4*, 2320. https://doi.org/10.1038/ncomms3320

Mace, E. S., Rami, J. F., Bouchet, S., Klein, P. E., Klein, R. R., Kilian, A., … Jordan, D. R. (2009). A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. *BMC Plant Biology*, *9*, 13. https://doi.org/10.1186/1471-2229-9-13

Magalhaes, J. V., Liu, J., Guimaraes, C. T., Lana, U. G., Alves, V. M., Wang, Y. H., … Kochian, L. V. (2007). A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. *Nature Genetics*, *39*, 1156–1161. https://doi.org/10.1038/ng2074

McCormick, R. F., Truong, S. K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., … Mullet, J. E. (2018). The *Sorghum bicolor* reference genome: Improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant Journal*, *93*, 338–354. https://doi.org/10.1111/tpj.13781

Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., … Kresovich, S. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 453–458. https://doi.org/10.1073/pnas.1215985110

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., … Rokhsar, D. S. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, *457*, 551–556. https://doi.org/10.1038/nature07723

Rao, N. K., Reddy, L., & Bramel, P. (2003). Potential of wild species for genetic enhancement of some semi-arid food crops. *Genetic Resources Crop Evolution*, *50*, 707–721. https://doi.org/10.1023/A:1025055018954

Raubach, S., Kilian, B., Dreher, K., Amri, A., Bassi, F. M., Boukar, O. … Shaw, P. D. (2020). From bits to bites: Advancement of the Germinate platform to support prebreeding informatics for crop wild relatives. *Crop Science*. https://doi.org/10.1002/csc2.20248 (in press).

Sagnard, F., Deu, M., Dembélé, D., Leblois, R., Touré, L., Diakité, M., … Mallé, Y. (2011). Genetic diversity, structure, gene flow and evolutionary relationships within the *Sorghum bicolor* wild–weedy–crop complex in a western African region. *Theoretical and Applied Genetics*, *123*, 1231. https://doi.org/10.1007/s00122-011-1662-0

Schertz, K., & Clark, L. (1967). Controlling dehiscence with plastic bags for hand crosses in sorghum. *Crop Science*, *7*, 540–542. https://doi.org/10.2135/cropsci1967.0011183X000700050040x

Stephens, J., Miller, F., & Rosenow, D. (1967). Conversion of alien sorghums to early combine genotypes. *Crop Science*, *7*, 396–396. https://doi.org/10.2135/cropsci1967.0011183X000700040036x

Tao, Y., Zhao, X., Wang, X., Hathorn, A., Hunt, C., Cruickshank, A., … Jordan, D. (2020). Large-scale GWAS in sorghum reveals common genetic control of grain size among cereals. *Plant Biotechnology Journal*, *18*, 1093–1105. https://doi.org/10.1111/pbi.13284

Tao, Y., Jordan, D. R., & Mace, E. S. (2019). Crop genomics goes beyond a single reference genome. *Trends in Plant Science*, *24*, 1072–1074. https://doi.org/10.1016/j.tplants.2019.10.001

Tao, Y., Zhao, X., Mace, E., Henry, R., & Jordan, D. (2019). Exploring and exploiting pan-genomics for crop improvement. *Molecular Plants*, *12*, 156–169. https://doi.org/10.1016/j.molp.2018.12.016

Tao, Y. F., Mace, E., George-Jaeggli, B., Hunt, C., Cruickshank, A., Henzell, R., & Jordan, D. (2018). Novel grain weight loci revealed in a cross between cultivated and wild sorghum. *The Plant Genome*, *11*, 1–10. https://doi.org/10.3835/plantgenome2017.10.0089

Tao, Y., Mace, E. S., Tai, S., Cruickshank, A., Campbell, B. C., Zhao, X., … Jordan, D. R. (2017). Whole-genome analysis of candidate genes associated with seed size and weight in *Sorghum bicolor* reveals signatures of artificial selection and insights into parallel domesti-

cation in cereal crops. *Frontiers in Plant Science*, 8, 1237. https://doi.org/10.3389/fpls.2017.01237

Thurber, C. S., Ma, J. M., Higgins, R. H., & Brown, P. J. (2013). Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. *Genome Biology*, *14*, R68. https://doi.org/10.1186/gb-2013-14-6-r68

White, J. A., Ryley, M. J., George, D. L., & Kong, G. A. (2015). Identification of pathotypes of the sorghum rust pathogen, *Puccinia purpurea*, in Australia. *Australasian Plant Pathology*, *44*, 1–4. https://doi.org/10.1007/s13313-014-0318-3

Wiersema, J. H., & Dahlberg, J. (2007). The nomenclature of *Sorghum bicolor* (L.) Moench (*Gramineae*). *Taxon*, *56*, 941–946. https://doi.org/10.2307/25065876

Yu, J., Holland, J. B., McMullen, M. D., & Buckler, E. S. (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics*, *178*, 539–551. https://doi.org/10.1534/genetics.107.074245

Zheng, L. Y., Guo, X. S., He, B., Sun, L. J., Peng, Y., Dong, S. S., … Jing, H. C. (2011). Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). *Genome Biology*, *12*, R114. https://doi.org/10.1186/gb-2011-12-11-r114

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.