Population genomic and genome-wide association analysis of lignin content in a global collection of 206 forage sorghum accessions

Hao Niu · Junai Ping · Yubin Wang · Xin Lv ·
Huiming Li · Fuyao Zhang · Jianqiang Chu ·
Yuanhuai Han

Abstract Forage sorghum (*Sorghum bicolor* (L.) Moench) is a wildly cultivated C4 cereal crop in many geographical regions and differs among germplasms in a number of important physiological traits. Lignin is a complex heteropolymer found in plant cell walls that adversely affects economic and environmental benefits of the crop. To understand the genetic basis, we resequenced the genomes of 206 sorghum accessions collected around the globe and identified 14,570,430 SNPs and 1,967,033 indels. Based on the SNP markers, we characterized the population structure and identified loci underlying lignin content by genome-wide association studies (GWAS). Analysis of the genetic relationships among the accessions revealed a more diverse spread of sorghum accessions and breeding lines from Asia, America, and their genetically improved variety, but a limited genetic diversity in the European accessions. These findings add new perspectives to the historical processes of crop diffusion within and across agroclimatic zones of America, Asia, and Europe. GWAS revealed 9 quantitative trait loci (QTLs) for lignin content, harboring 184 genes. These genes were significantly enriched into 7 major gene ontology (GO) terms involved in plant-type cell wall organization or bioenergy. The alleles of 9 QTLs in the 206 accessions were geographically distributed. The findings provide us with an understanding of the origin and spread of haplotypes linked to lignin content. The findings will allow improvements to feed quality and adaptation to stresses in sorghum, through the rapid increase of genetic gains for lignin content.

Keywords *Sorghum bicolor* · Quantitative trait locus · Lignin content · Genome-wide association studies

H. Niu · Y. Han
College of Agriculture, Shanxi Agricultural University,
Taigu 030801, China

H. Niu · J. Ping · Y. Wang · X. Lv · H. Li · F. Zhang ·
J. Chu
Sorghum Institute of Shanxi Academy of Agricultural Sciences,
Shanxi Agricultural University, Jinzhong 030600, China

H. Niu · J. Ping · Y. Wang · X. Lv · H. Li · F. Zhang ·
J. Chu
Shanxi Key Laboratory of Sorghum Genetic and Germplasm
Innovation, Jinzhong 030600, China

H. Niu · J. Ping (✉) · Y. Wang · X. Lv · H. Li ·
F. Zhang · J. Chu
Key Laboratory of Crop Gene Resources and Germplasm
Enhancement on Loess Plateau, Ministry of Agriculture,
Taiyuan 030031, China
e-mail: pingja1029@163.com

Introduction

Sorghum (*Sorghum bicolor* (L.) Moench) is the fifth leading cereal crop worldwide in both subsistence and commercial agriculture systems, particularly in semiarid regions such as sub-Saharan Africa and South Asia (National Research Council 1996; Mundia et al. 2019). *Sorghum bicolor* is divided into four groups according to their production characteristics and usage, including grain, forage, sweet sorghum, and broom. Compared to

grain sorghum, course forage sorghum, often used for green chop, hay, silage, and pasture (Shoemaker and Bransby 2010), grows quickly, is morphologically taller and leafier, and matures later with a higher biomass and a juicy stem, but relatively lower seed yield (Harlan and De Wet 1972; Getachew et al. 2016). Recently, it has been identified as a potential target for sugar as well as the production of lignocellulosic biofuels (Mathur et al. 2017). Efforts to develop multipurpose forage sorghum cultivars or hybrids with superior stem quality, standability, high sugar levels, high dry matter production, and low prussic acid levels have been initiated using both classical and biotechnological approaches.

Forage sorghum typically grows over 150 cm tall (Getachew et al. 2016), primarily due to lignin, which improves strength and rigidity in cell walls, as well as mechanical support for leaf blades and stems (Frei 2013). Lignin is a phenylpropanoid polymer which possesses complex structure and composition, as well as a high molecular weight (Frei 2013; Liu et al. 2018). It mainly deposits in fibers, such as parenchyma bundle-sheath cells (Frei 2013; Liu et al. 2018). As a result of this trait, together with its tolerance to stresses and relatively high nutrient uptake efficiency, forage sorghum is globally cultivated in diverse geographical areas, even barren regions unsuitable for the production of food and feed (National Research Council 1996; Vermerris 2011; Liu et al. 2018; Mundia et al. 2019). However, lignin affects the digestibility of forage in animals and is considered both an unfavorable component of forages and an antinutritive in animals (Frei 2013). Reducing lignin while increasing sugar and cellulose content are current priorities in breeding, in order to improve digestibility by livestock. *Sorghum bicolor* is highly adaptable across an extensive geographic range; sources for the improvement of traits may be found in traditional varieties from Africa and Asia (National Research Council 1996; Casa et al. 2008; Upadhyaya et al. 2009; Vermerris 2011; Morris et al. 2013; Mathur et al. 2017; Mundia et al. 2019). However, the complex genetics for these remarkable traits of forage sorghum have not been well elucidated.

In-depth sequencing of whole genomes of diverse *Sorghum bicolor* populations is increasingly being used to identify genes and loci of underlying complex traits (Casa et al. 2008; Upadhyaya et al. 2009; Vermerris 2011; Morris et al. 2013; Mathur et al. 2017) at an unprecedented rate (The International HapMap

Consortium 2007; The Wellcome Trust Case Control Consortium 2007). Genome-wide single nucleotide diversity scans (Casa et al. 2006; Zheng et al. 2011; Bouchet et al. 2012) and association studies (Morris et al. 2013) have been performed to investigate genome-wide and intraspecific variation for dissecting the genetic basis of important traits, and for tailor-designed breeding in sorghum. However, compared to maize (Jiao et al. 2012) and rice (Huang et al. 2010), less is known. Brown midrib (bmr) mutants in *Sorghum bicolor* (Porter et al. 1978) possess a reduced concentration and altered composition of lignin, as well as improved digestibility of cell walls, and are demonstrated to have desirable qualities for livestock forage development as well as the rapidly growing lignocellulosic biofuel industry (Cherney et al. 1991; Oliver et al. 2004). Several bmr mutants are involved in specific enzymatic steps of the monolignol biosynthetic pathway, such as cinnamyl alcohol dehydrogenase (CAD) (Pillonel et al. 1991; Saballos et al. 2012; Scully et al. 2016), among which a CAD allele (SbCAD2) with an 8-bp deletion in its 5′-untranslated region (UTR) was mapped and cloned in the sorghum germplasm line PI 595743 (Li and Huang 2017). Barcode multiplexing and high-throughput sequencing (Elshire et al. 2011) were applied to develop thousands of markers for rapid genotyping of four classical dwarfing loci (Quinby 1975) in sorghum at a low cost. Recently, Zheng et al. (2011) sequenced the inbred lines of two sweet and one grain sorghum and identified an arrangement of almost 1500 genes differentiating the two sorghums. Morris et al. (2013) analyzed the diversity of sorghum and mapped several classical loci and genes for plant height and inflorescence architecture with genome-wide association analysis (GWAS) by characterizing 971 sorghum accessions at 265,487 SNPs with genotyping-by-sequencing (GBS).

The genetic diversity among forage sorghum accessions has been studied with simple sequence repeats (SSR) (Ping et al. 2018) and sequence-related amplified polymorphism (SRAP) (Zhang et al. 2018). Zhang et al. (2018) revealed the narrow SRAP genetic diversity of 21 introduced *Sorghum bicolor* bmr sterile lines, whereas Ping et al. (2018) reported relatively abundant genetic diversity among 48 forage sorghum germplasms by 109 SSRs. To better understand the genetic diversity of forage sorghum, we characterized 206 lines collected worldwide by genome sequencing a much larger sample population than those used by previous studies (Ping

et al. 2018; Zhang et al. 2018). GWAS was applied to identify genes underlying lignin content in forage sorghum.

Materials and methods

Plant materials

A set of 206 diverse forage sorghum accessions were collected worldwide, including 30 male sterile lines, 30 maintainer lines, and 146 fertility restorer lines, representing various geographical origins and states of improvement.

All 206 sorghum accessions were grown for phenotypic evaluation of agronomic traits and quality in the experimental station of Jinzhong City, Shanxi Province, China (112° 42′ N and 37° 36′ E) from 2016 to 2018 cropping seasons. Using a randomized complete block design with three replications, seeds were planted in 1236-row plots. Plots were 5 m in length with a row spacing of 0.5 m. Sowing density was approximately 20 seeds/row. At the end of August 2018, in order to measure quality-related components (including lignin), 206 forage sorghum accessions were sampled in batches according to their heading dates. The fresh samples were ground using a pulverizer into particles 1–2 cm in size, and evenly divided into 500 g in Kraft paper bags. The samples were first dried at 65 °C for 72 h after inactivation at 105 °C for 1 h in a programmable oven. The dried samples were then analyzed for quality-related components by the Beijing Quality Testing Center, China.

Determination of lignin content

The 100-g samples per accession were dried to less than 10% moisture and milled using a cyclone mill to a 1-mm particle size. The sieved powder was collected for spectral scanning with FOSS 5000 near infrared (NIR) analyzer (FOSS, Hillerød, Denmark). Thirty-two scans were performed for each sample using the wavelength range of 1100 to 2500 with 2-nm data spacing. Spectral data was collected using Vision software (version 3.5.0.0). Each sample was loaded and scanned 2 times, and the average spectrum was regarded as the sample spectrum used for NIR analysis.

DNA isolation and genome sequencing

Following 3 weeks of growth, genomic DNA was extracted from the young leaves of one sorghum plant from each accession. DNA extraction was performed using the cetyltrimethylammonium bromide (CTAB) method (Murray and Thompson 1980) for genome sequencing on the Illumina HiSeq system (PersonalBio, Shanghai, China). Briefly, following quality assessment, genomic DNA was randomly fragmented using sonification and size-fractionated through electrophoresis, and DNA fragments of the desired length (400 bp) were gel purified. Using an insert size of approximately 400 bp, Illumina paired-end sequencing libraries were constructed, following the manufacturer's instructions (Illumina, San Diego, USA). Libraries were sequenced as 2X150 bp paired reads.

Read alignment and variation calling

First, raw data (raw reads) in fastq format were cleaned using in-house perl scripts to remove reads containing adapter, ploy-N, and low-quality reads. Meanwhile, high-quality clean data was calculated for Q20, Q40, and GC contents, and all downstream analyses were performed. Clean reads were mapped to the reference genome BTx623 (v3.1) (McCormick et al. 2018) (ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/plant/Sorghum_bicolor/latest_assembly_versions/GCA_000003195.3_Sorghum_bicolor_NCBIv3) with Burrows-Wheeler Alignment (BWA) software (Li and Durbin 2009) (version 0.7.12-r1039) using the default parameters. SNP and INDEL were called by the Genome Analysis Toolkit (McKenna et al. 2010) (GATK, version: 3.8). All raw SNPs were filtrated with a Fisher test of Strand Bias (FS) ≤ 60, HaplotypeScore ≤ 13.0, Mapping Quality (MQ) ≥ 40, Quality Depth (QD) ≥ 2, ReadPosRankSum ≥ -8.0, MQRankSum > -12.5, alternative allele called in ≥ 4 reads, and removal of SNPcluster sites. Indels were filtrated with a Fisher Test of Strand Bias (FS) ≤ 200, Quality Depth (QD) ≥ 2, and ReadPosRankSum ≥ -20. Annotations of SNP and INDEL were performed based on the gene model set v2.0 from Phytozome v11.0 using ANNOVAR (Wang et al. 2010) (version: 2015-03-22).

Population genetics analysis and GWAS

In the population genetics analysis and GWAS, only SNPs with minor allele frequency (MAF) ≥ 0.05 were

used. Hierarchical population structure was developed using the MEGA 6.0 software (http://www.megasoftware.net/) with bootstrap replications of 1000. Principal component analysis (PCA) of the population was performed using R package. Genetic relatedness (additive genetic effect matrix) was determined using the genome-wide SNP data (Endelman and Jannink 2012). Whole-genome linkage disequilibrium (LD) was separately calculated (r^2) for all accessions and each subgroup using TASSEL 3.0 (Bradbury et al. 2007). GWAS was performed using the Efficient Mixed-Model Association eXpedited (EMMAX, beta version) (Kang et al. 2010) software package, and in order to determine the threshold, distribution of the most significant P values across the 1000 replicates was used. $-\log10(P \text{ value}) > 6$ was used as the threshold for all binary-like quantitative traits. The matrix of pairwise genetic distances, which were derived from the simple matching coefficients, as the variance-covariance matrix of the random effects, was also calculated by EMMAX.

Results

Phylogenetic relationships of 206 diverse sorghum accessions

For the 206 accessions, whole-genome resequencing (WGRS) data was generated. This includes 89 Chinese core collection accessions, 4 landraces, and 113 other accessions from different sorghum-producing nations and areas, representing the genetic, geographic, and morphological diversity of forage sorghum. In total, 7.31 billion paired-end reads were generated with sequencing depths ranging from 3.56× to 14.06× and genome coverage of approximately 90.01%. Following mapping against the reference genome and the SNP calling, a total of 14,570,430 SNPs with MAF ≥ 0.05 and 1,967,033 indels were identified. In the PCA, based on all collected SNPs, the first and second principal coordinate (PC) explained 17.87% and 5.16% of the molecular variance for SNPs, respectively (Fig. 1a). Genetic relationships among 206 accessions reflected a more diverse spread of sorghum accessions, as well as breeding lines from Asia and America, Asia × America, Asia × Australia, and a limited genetic diversity of European accessions (Fig. 1b).

Analyses based on pairwise dissimilarity of all the SNPs in the collection using neighbor joining revealed two distinct groups (Fig. 1c). Group I contained 60 American accessions and one Asia × America, whereas group II included all breeding lines and accessions. Accessions from Europe, America, and Asia, and breeding lines of Asia × America and Asia × Australia were predominantly but not exclusively divided into several subgroups or minor groups within group II, showing the prevalence of genetic mixtures of breeding lines and landraces in sorghum breeding.

Whole-genome patterns of linkage disequilibrium

Whole-genome patterns of linkage disequilibrium (LD) were analyzed separately for Asia and America, Asia × America, Asia × Australia, and European accessions using the SNP data. The LD decay rate was measured as the chromosomal distance at which the average pairwise correlation coefficient (r^2) dropped to half its maximum value. On average, LD decays to 50% of its initial value by 1 kb and to background levels ($r^2 < 0.1$) within 150 kb. A rapid decline in LD was observed with increasing physical distance on subgroups of Asia and America, Asia × America, Asia × Australia, and European accessions (Fig. 1d). The decay rate varied widely across the genomes of Asia and America, Asia × America, Asia × Australia, and European subgroups, which will presumably lead to differential resolutions of association mapping at different genomic regions. European accessions had the lowest LD decay rate, which might be a result of self-fertilization coupled with a relatively small effective population size.

Genome-wide associations with agronomic traits

In total, 19 agronomic traits and 18 feedable quality traits were phenotyped in the population of 206 accessions in three different experiments from 2016 to 2018 cropping seasons using a randomized complete block design with three replications. Using the pooled phenotyping data, a total of 14,570,430 SNP markers with MAF ≥ 0.05 from 206 accessions were analyzed for 19 agronomic traits and 18 feedable quality traits. EMMAX was used to estimate marker-trait associations (MTAs). As shown in Fig. 2a, the lignin content values across all 206 sorghum accessions showed continuous frequency distribution with two peaks, suggesting polygenic and quantitative inheritance of lignin content. A total of 9 candidate regions ($-\log(P) > 6$) spanning 100 kb were significantly associated with lignin content
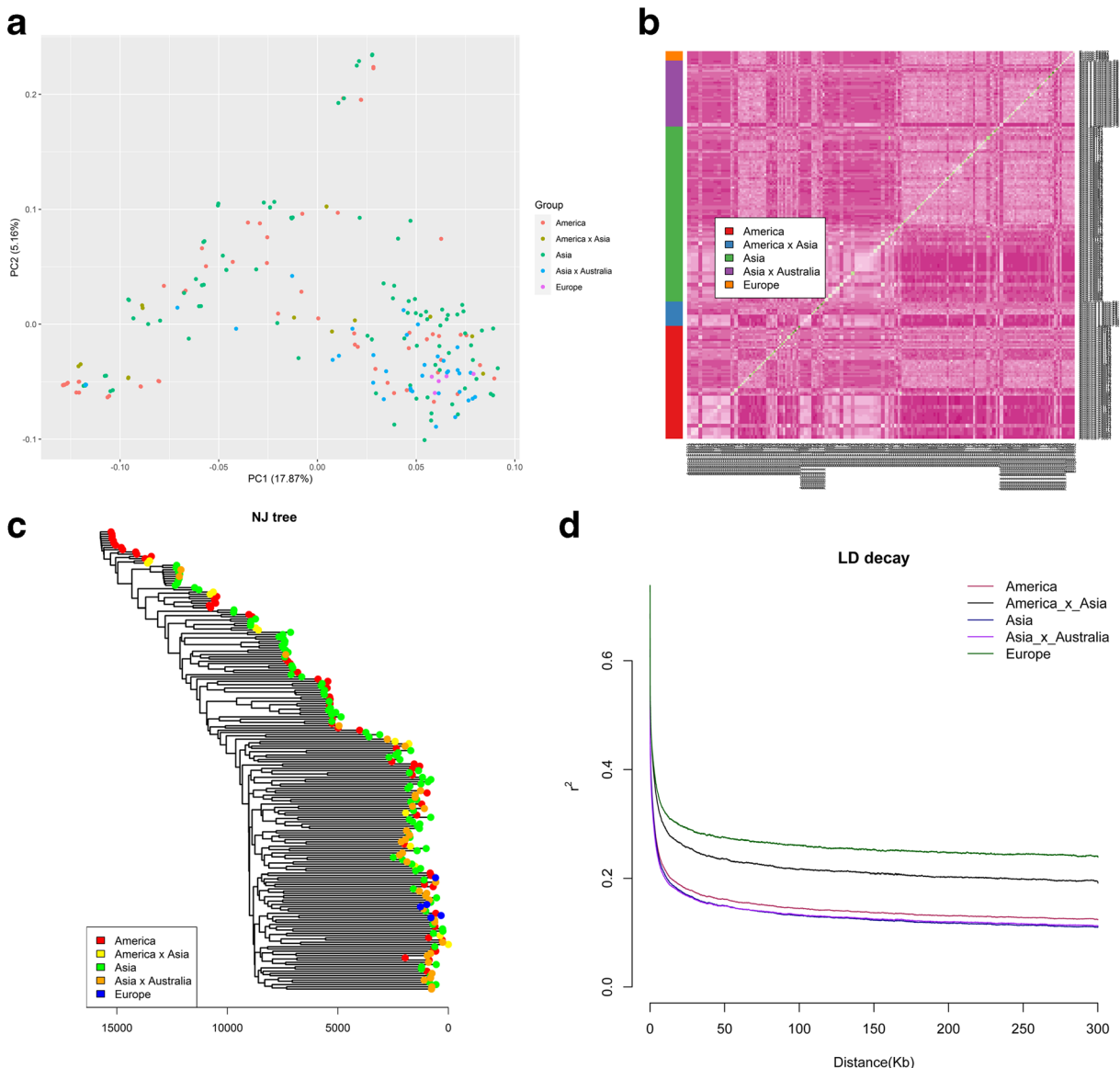
a

b

c

d

Fig. 1 Genetic structure of 206 worldwide sorghum accessions using SNPs detected in whole-genome resequencing data. **a** PCA, **b** genetic relatedness (additive genetic effect matrix) using SNPs detected in whole-genome resequencing data, **c** neighbor joining tree constructed from simple matching distance of all SNPs, and **d** genome-wide average LD decay

on chromosomes 1, 4, 6, 7, and 9 (Fig. 2b and c), supported by the strong LD among associated SNPs.

Geographic distribution of QTL alleles and lignin content variation between haplotypes

The geographic distribution for SNPs of 9 QTLs in 206 accessions was characterized in order to acquire increased knowledge of haplotype spread and origin linked to lignin content (Fig. 3). Allelic distribution for

C1_21143564 showed that European accessions carried haplotypes CT and TT only, whereas the America × Asia breeding lines contained haplotypes CC and TT. The lignin content in haplotype CC was significantly lower than those found in both haplotypes TT and CT (Fig. 3a). For C1_S24250037, European sorghum only carried haplotype TT, whereas the America × Asia breeding lines did not contain haplotype TC. Haplotype CC was significantly associated with lower lignin content compared with haplotypes TC and TT (Fig. 3b). For
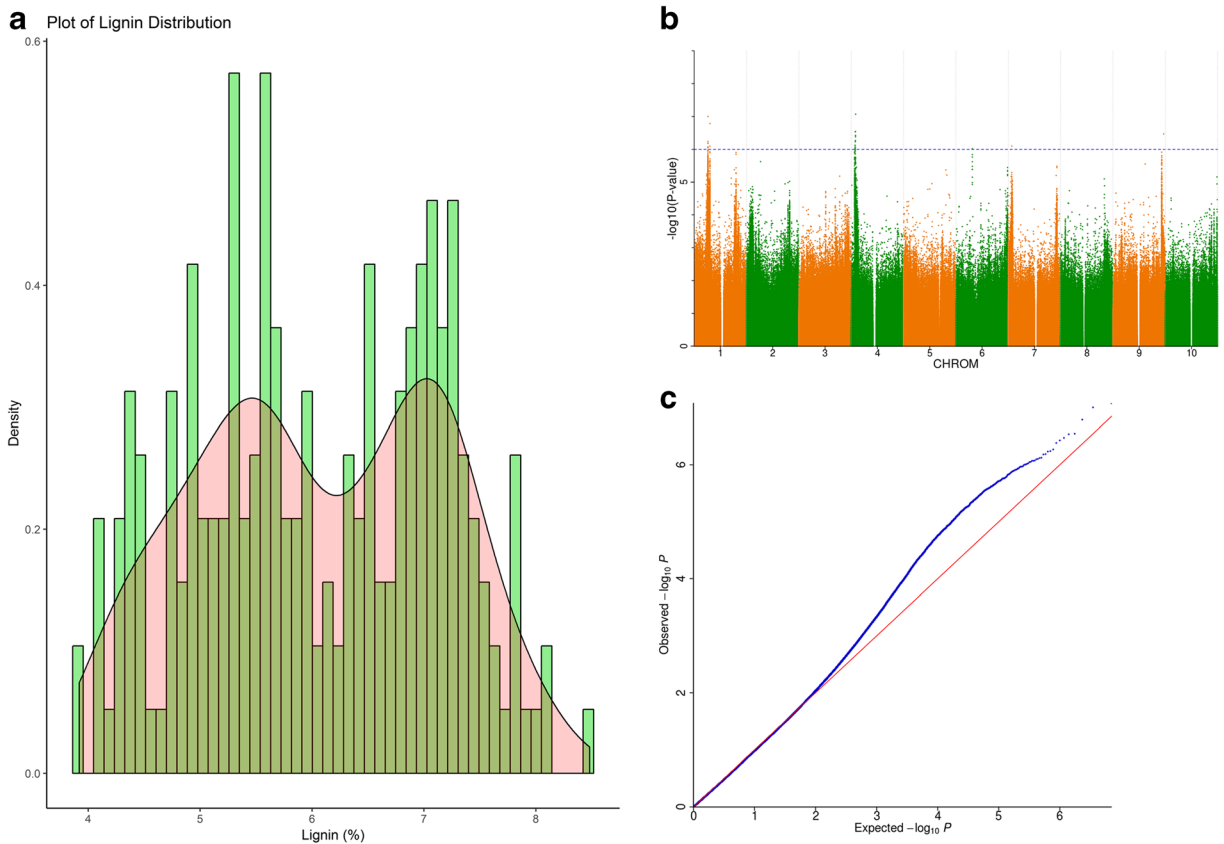
a Plot of Lignin Distribution

b

c

Fig. 2 GWAS of the sorghum lignin content. **a** Distribution of the lignin content values across 206 sorghum accessions. **b** Manhattan plot for lignin content. The −log10 P values from a genome-wide scan are plotted against the position on each of the 10 chromosomes. The horizontal dashed line indicates the genome-wide significance threshold ($P = 5 \times 10^{-8}$). **c** Quantile plot for lignin content. The horizontal axis shows −log10 transformed expected P values, and the vertical axis indicates −log10 transformed observed P values

C4_S4831311, no European accession carried haplotype TT that was significantly related to lower lignin content (Fig. 3c). European accessions carried haplotype AA only, whereas the America × Asia breeding lines did not contain haplotype AG for C4_S5402582 (Fig. 3d). The lignin content in haplotype GG was significantly lower than that in both haplotypes AG and AA. For C6_S19087442, only haplotypes TC and TT existed in all accessions with no significant effect on lignin content (Fig. 3e). For C7_S4285582, European accessions carried haplotype GG only, whereas America × Asia breeding lines did not contain haplotype AG. Haplotype AA was significantly associated with lower lignin content compared to haplotypes AG and GG (Fig. 3f). The America × Asia breeding lines did not contain haplotypes AA and TA, whereas the European accession did not carry haplotype AA for C9_S57393274. The lignin content in haplotype TT was significantly

lower than that of haplotype TA but insignificantly different from haplotype AA (Fig. 3g).

Genetic collinearity analysis

Nine candidate regions for lignin content revealed 255,689 and 253,400 SNPs in the regions on chromosomes 1, 433, 000, and 435, 667 SNPs within the regions on chromosome 4, and 200,000 SNPs in the regions on chromosomes 6, 7, and 9 (Table 1), thereby covering 184 genes. Collinearity analysis showed a high degree of collinearity in 13 of 184 sorghum lignin-related loci with previously reported gene families in grasses, including corn and rice (Table 2). SORBI_3001G222700 is highly collinear with *Wda1* (Os10g0471100) in rice, encoding a protein similar to those involved in the production of wax.
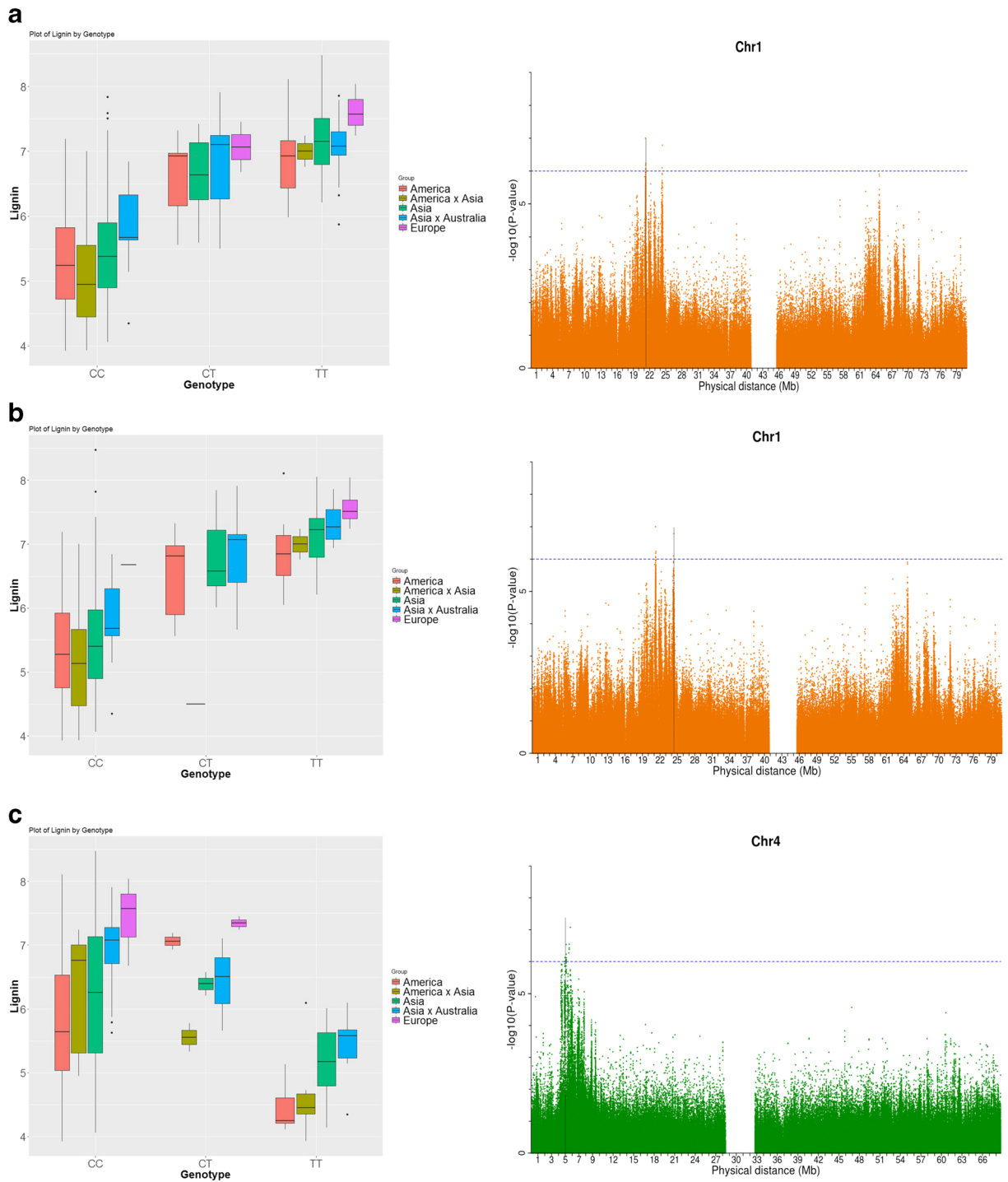
Fig. 3 The geographic distribution of alleles and lignin content variation among different haplotypes in each QTL. **a** C1_S21143564. **b** C1_S24250037. **c** C4_S4831311. **d** C4_S5402582. **e** C6_S19087442. **f** C7_S4285582. **g** C9_S57393274
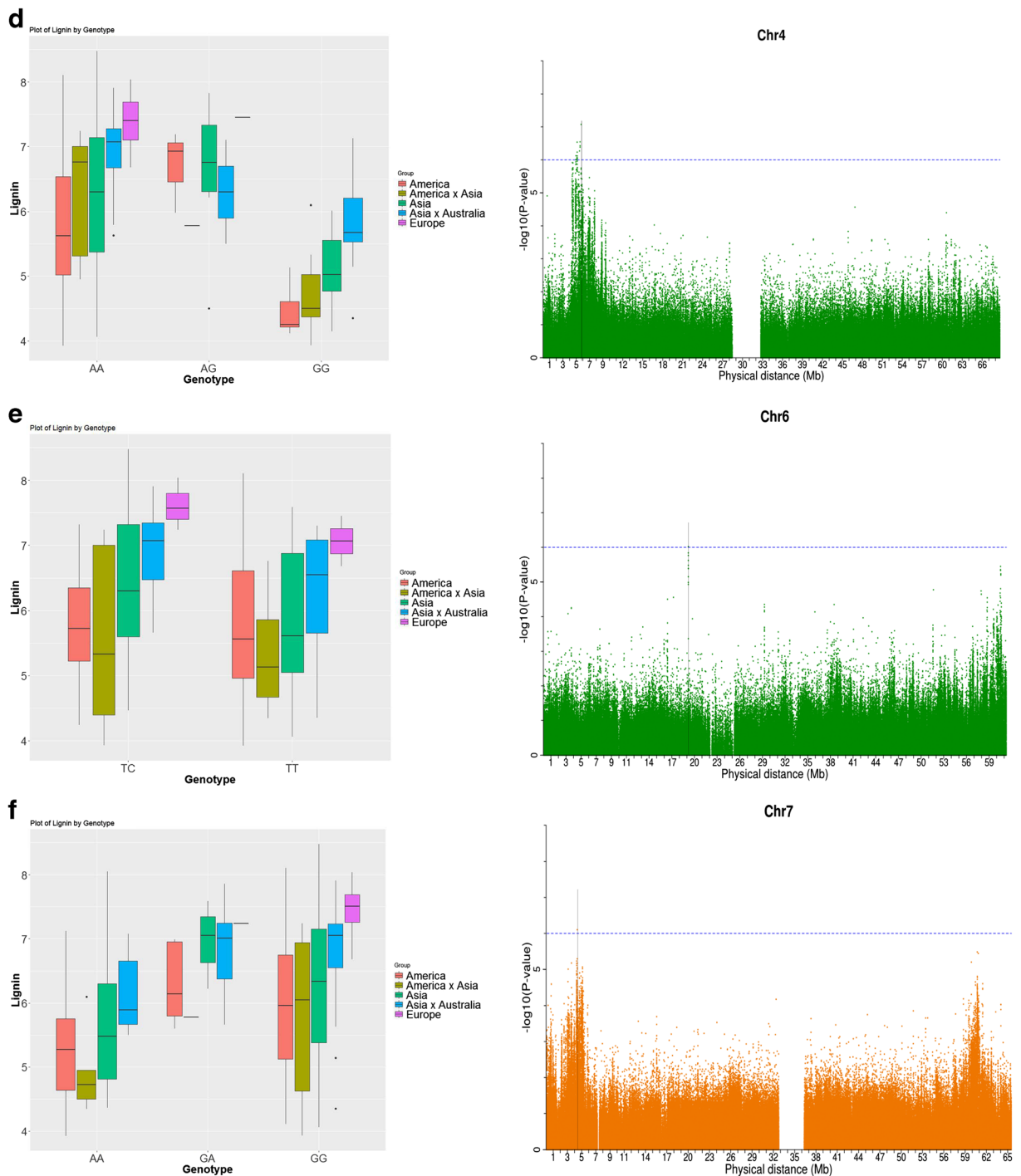
Fig. 3 (continued)

Discussion

The improvement of crop productivity has been greatly facilitated by pyramiding several advantageous traits into a singular type. However, the availability of genomic resources for the genetic dissection of complex traits in *Sorghum bicolor* is limited in comparison to those in rice (Huang et al. 2010) and maize (Jiao et al.
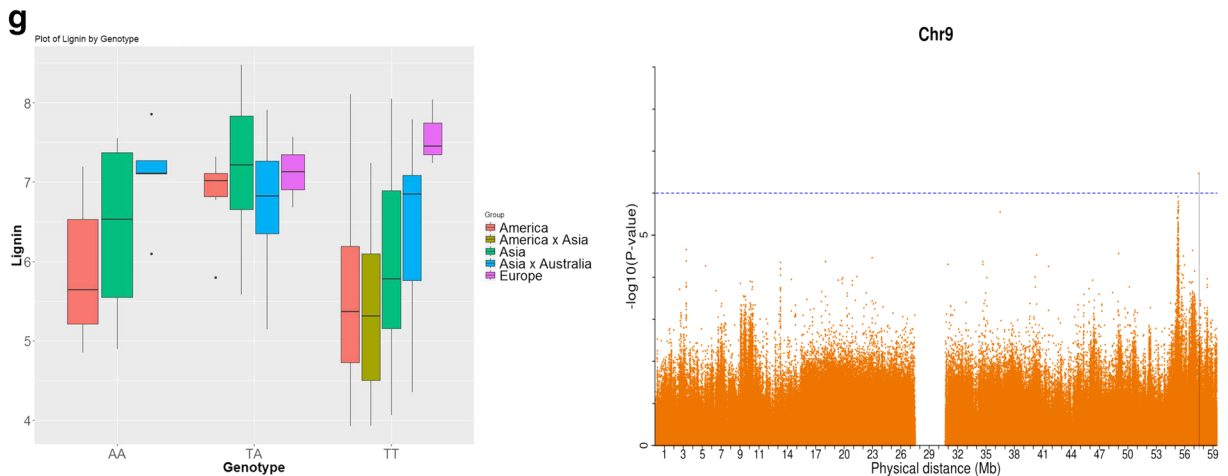
Fig. 3 (continued)

2012). Targeted efforts to exploit the tremendous genetic diversity in *Sorghum bicolor* germplasm collections desirable traits are underway (Zheng et al. 2011; Morris et al. 2013; Ordonia et al. 2016). Systematic assessment and application of available germplasm is important for the improvement of forage sorghum, through access to allelic variations that affect essential agronomic traits such as biomass, and feedable quality traits such as lignin content. In the present study, we characterized 14,570,430 SNPs and 1,967,033 indels, creating a valuable resource for both future gene-phenotype studies and molecular breeding of *Sorghum bicolor*, in addition to related species. The population structure of 206 *Sorghum bicolor* germplasm was characterized using genome-wide SNPs, which are a priority in decisions regarding selection strategies and plant breeding options for elite forage sorghum varieties and hybrids. Genetic relationships among 206 accessions revealed a more diverse spread of sorghum accessions and breeding lines

from Asia and America and their genetically improved variety, and a limited genetic diversity in European accessions, adding to the knowledge of historical crop diffusion processes, within and across the agroclimatic zones of America, Asia, and Europe.

The quantity of genomic data increases understanding of the allelic variation in the genetic resource collections; it will also encourage breeders to propose an efficient path for improvement of varieties by design. Morris et al. (2013) traced the independent spread of several haplotypes containing alleles for short stature and long inflorescence branches. In total, we identified 9 QTLs for lignin content and observed evidence of the independent spread of multiple alleles controlling lignin content (Fig. 3a–g). In general, favorable alleles for lower lignin content are observed in Asian and American accessions and their genetically improved variety, whereas unfavorable alleles were found in the European sorghum population. Interestingly, none of the alleles for lower lignin content at loci C1_S21143564, C1_S24250037, C4_S5402582, and C7_S4285582 were restricted to European accessions (Fig. 3a–g). This suggests that it is not QTL that is responsible for the lower lignin content phenotype in European accessions. The alleles for lignin content aid in reductions to molecular diversity and are found in specific genomic regions, such as on C1_S21143564, C1_S24250037, C4_S5402582, and C7_S4285582; specifically, those that have experienced selective sweeps. The lignin-related gene sets were significantly enriched in the lignin-associated biological process of plant-type cell wall organization ($P < 0.01$). In addition to the

Table 1 Candidate QTL regions significantly associated with lignin content

Chr	QTL	Start	End
1	C1_S21143564	21,143,564	21,399,253
1	C1_S24250037	24,250,037	24,503,437
4	C4_S4831311	4,831,311	5,264,311
4	C4_S5402582	5,402,582	5,838,249
6	C6_S19087442	19,087,442	19,287,442
7	C7_S4285582	4,285,582	4,485,582
9	C9_S57393274	57,393,274	57,593,274

Table 2 Gene collinearity of lignin content between sorghum and crops (corn and rice)

Gene ID	Ortholog name	Description
SORBI_3004G069850	Os02g0187100	Similar to cyclase
SORBI_3004G070000	Os06g0638700	Putative cyclase family protein
SORBI_3001G237900	Os10g0439200	Similar to expansin-A28
SORBI_3001G238000	Os03g0428700	Similar to expansin-A31
SORBI_3001G238000	Os10g0439100	Similar to cDNA clone:002-180-G03, full insert sequence
SORBI_3001G238100	Os03g0428700	Similar to expansin-A31
SORBI_3001G238100	Os10g0439100	Similar to cDNA clone:002-180-G03, full insert sequence
SORBI_3001G238200	Os03g0428700	Similar to expansin-A31
SORBI_3001G238200	Os10g0439100	Similar to cDNA clone:002-180-G03, full insert sequence
SORBI_3001G238300	Os03g0428700	Similar to expansin-A31
SORBI_3001G238300	Os10g0439100	Similar to cDNA clone:002-180-G03, full insert sequence
SORBI_3001G238400	Os03g0428700	Similar to expansin-A31
SORBI_3001G238400	Os10g0439100	Similar to cDNA clone:002-180-G03, full insert sequence
SORBI_3001G222700	Os10g0471100	Protein with high similarity to proteins involved in wax production, cuticular wax biosynthesis
SORBI_3004G060700	Os04g0353600	Similar to OSIGBa0092G14.1 protein
SORBI_3004G060700	Os04g0354400	Similar to OSIGBa0092G14.8 protein
SORBI_3004G060700	Os04g0354600	Similar to acyl-CoA reductase-like protein
SORBI_3004G060700	Os04g0355500	Similar to OSIGBa0147J02.3 protein
SORBI_3004G064000	Os02g0178800	Glossy1(GL1) homolog, cuticular wax biosynthesis, drought resistance
SORBI_3004G069850	Zm00001d015599	Cyclase family protein
SORBI_3001G237800	Zm00001d032876	Expansin-A11
SORBI_3001G237900	Zm00001d032881	Expansin-A11
SORBI_3001G238000	Zm00001d032882	Expansin-A11
SORBI_3001G238200	Zm00001d032883	Expansin-A11
SORBI_3001G238300	Zm00001d032884	Expansin-A11
SORBI_3001G238300	Zm00001d032885	Expansin-A11
SORBI_3001G238400	Zm00001d032886	Expansin-A11
SORBI_3001G222700	Zm00001d014055	Protein ECERIFERUM 1
SORBI_3004G064000	Zm00001d015477	Protein ECERIFERUM 3
SORBI_3001G222700	MTR_4g054150	Fatty acid hydroxylase superfamily protein
SORBI_3001G222700	MTR_4g054290	Fatty acid hydroxylase superfamily protein
SORBI_3001G222700	MTR_7g090100	Fatty acid hydroxylase superfamily protein
SORBI_3001G222700	MTR_7g090120	Fatty acid hydroxylase superfamily protein
SORBI_3001G222700	MTR_7g090140	Fatty acid hydroxylase superfamily protein
SORBI_3001G222700	MTR_8g009560	Fatty acid hydroxylase superfamily protein
SORBI_3001G222700	MTR_8g009590	Fatty acid hydroxylase superfamily protein
SORBI_3004G060700	MTR_2g086070	Gland-specific fatty acyl-CoA reductase
SORBI_3004G060700	MTR_2g086090	Gland-specific fatty acyl-CoA reductase
SORBI_3004G060700	MTR_4g053785	Gland-specific fatty acyl-CoA reductase
SORBI_3004G060700	MTR_4g053880	Gland-specific fatty acyl-CoA reductase
SORBI_3004G060700	MTR_4g053900	Gland-specific fatty acyl-CoA reductase
SORBI_3004G060700	MTR_4g054220	Fatty acyl-CoA reductase
SORBI_3004G060700	MTR_4g054240	Fatty acyl-CoA reductase-like protein, putative
SORBI_3004G060700	MTR_4g054820	Gland-specific fatty acyl-CoA reductase
SORBI_3004G060700	MTR_5g045330	Fatty acyl-CoA reductase
SORBI_3004G060700	MTR_5g072540	Gland-specific fatty acyl-CoA reductase
SORBI_3004G060700	MTR_5g072550	Fatty acyl-CoA reductase-like protein
SORBI_3004G060700	MTR_6g034620	Fatty acyl-CoA reductase
SORBI_3004G060700	MTR_6g034660	Fatty acyl-CoA reductase
SORBI_3004G060700	MTR_6g039280	Fatty acyl-CoA reductase-like protein, putative
SORBI_3004G064000	MTR_4g129630	Fatty acid hydroxylase superfamily protein

identification of 9 QTLs for lignin content, we revealed a high collinearity in 13 of 184 sorghum lignin-related loci with previously reported genes in crops such as maize and rice (Table 2). SORBI_3001G222700 might act like Wda1 (Os10g0471100) (Jung et al. 2006) in rice to encode proteins highly similar to those involved in the production of wax. This information will guide breeders in developing a better strategy to improve forage sorghum.

Sorghum bicolor has demonstrated qualities such as high biomass yield, superior drought tolerance, and efficient use of nutrients, and serves as one of the leading biomass feedstocks for livestock. Lignin imparts strength and rigidity to plant cell walls as well as mechanical support for stems and leaf blades (Buxton and Redfearn 1997; Frei 2013; Liu et al. 2018). Research has demonstrated that lignin and lignin-like compounds, alongside other constituents of cell walls, provide resistance to disease, insects, cold temperatures, and other biotic and abiotic stresses (Rest et al. 2006; Tripathi et al. 2003; Ithal et al. 2007; Shadle et al. 2007). Lignin is an important component of the secondary cell wall, but is also a major impediment to biomass feedable utilization for feedstocks (Frei 2013), due to interference with microbial degradation of fiber polysaccharides. It acts as a physical barrier and is cross-linked to polysaccharides by ferulate bridges (Buxton and Redfearn 1997). Thus, forage sorghum breeding practices for improving digestibility without negative effects on growth and survival in field environments are challenging, due to the level of reduction available for lignin and other cell wall components. Selection for low lignin concentration in alfalfa resulted in less growth vigor and poorer field survival (Buxton and Casler 1993). Therefore, an improvement strategy for lignin that minimizes developmental defects in lignin-modified plants involves a set of genes that coordinate composition of lignin synthesis, deposition, and integration into the cell wall network. MTAs and accessions containing desirable alleles, as identified in this study, will help speed up the rate of improvement of forage sorghum varieties and hybrids, with a desirable range of lignin content using genomics-assisted breeding with sequence variation. Further studies are needed to define the causal genes for these and other MTAs, in order to improve understanding of the underlying regulatory mechanisms of lignin production. This in turn will aid in the long-term goal of manipulation of cell wall structure in biomass feedstocks.

In conclusion, genomic data for the worldwide collection of forage sorghum, inferences drawn from their analysis, and MTAs for agronomically relevant traits collectively provide valuable resources to accelerate genetic gains in programs dedicated to the improvement of forage sorghum. The candidate genes participating in lignin content lay the foundation for studying the potential use of functional genomics methodologies to validate the function and effect of candidate genes in regulating lignin content.

Compliance with ethical standards

References

Bouchet S, Pot D, Deu M, Rami JF, Billot C, Perrier X, Rivallan R, Gardes L, Xia L, Wenzl P, Kilian A, Glaszmann JC (2012) Genetic structure, linkage disequilibrium and signature of selection in Sorghum: lessons from physically anchored DarT markers. PLoS One 7:e33470

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Buxton DR, Casler MD (1993) Environmental and genetic effects on cell wall composition and digestibility. In: Jung HG, Buxton DR, Hatfield RD, Ralph J (eds) Forage cell wall structure and digestibility. ASA-CSA-SSSA, Madison, pp 685–714

Buxton DR, Redfearn DD (1997) Plant limitations to fiber digestion and utilization. J Nutr 127:814S–818S

Casa AM, Mitchell SE, Jensen J, Hamblin MT, Paterson AH, Aquadro CF, Kresovich S (2006) Evidence for a selective sweep on chromosome 1 of cultivated sorghum. Crop Sci 46:S27–S40

Casa AM, Pressoir G, Brown PJ, Mitchell SE, Rooney WL, Tuinstra MR, Franks CD, Kresovich S (2008) Community resources and strategies for association mapping in sorghum. Crop Sci 48:30–40

Cherney JH, Cherney DJR, Akin DE, Axtell JD (1991) Potential of brown-midrib, low-lignin mutants for improving forage quality. Adv Agron 46:157–198

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379

Endelman JB, Jannink JL (2012) Shrinkage estimation of the realized relationship matrix. G3 (Bethesda) 2:1405–1413

Frei M (2013) Lignin: characterization of a multifaceted crop component. Sci World J 2013:1–25. https://doi.org/10.1155/2013/436517

Getachew G, Putnam DH, De Ben CH, De Peters EJ (2016) Potential of Sorghum as an alternative to corn forage. Am J Plant Sci 07:1106–1121

Harlan JR, De Wet JMJ (1972) A simplified classification of cultivated sorghum. Crop Sci 12:172–176

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967

Ithal N, Recknor J, Nettleton D, Maier T, Baum TJ, Mitchum MG (2007) Developmental transcript profiling of cyst nematode feeding cells in soybean roots. Mol Plant-Microbe Interact 20:510–525

Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J, Wang B, Liu Z, Chen J, Li W, Zhang M, Xie S, Lai J (2012) Genome-wide genetic changes during modern breeding of maize. Nat Genet 44:812–815

Jung KH, Han MJ, Lee DY, Lee YS, Schreiber L, Franke R, Faust A, Yephremov A, Saedler H, Kim YW, Hwang I, An G (2006) Wax-deficient anther1 is involved in cuticle and wax production in rice anther walls and is required for pollen development. Plant Cell 18(11):3015–3032

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42:348–354

Li H, Durbin R (2009) Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 25:1754–1760

Li H, Huang Y (2017) Expression of brown-midrib in a spontaneous sorghum mutant is linked to a 5′-UTR deletion in lignin biosynthesis gene SbCAD2. Sci Rep 7:11664

Liu Q, Luo L, Zheng L (2018) Lignins: biosynthesis and biological functions in plants. Int J Mol Sci 19:335

Mathur S, Umakanth AV, Tonapi VA, Sharma R, Sharma MK (2017) Sweet sorghum as biofuel feedstock: recent advances and available resources. Biotechnol Biofuels 10:146

McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B, Mattison A, Morishige DT, Grimwood J, Schmutz J, Mullet JE (2018) The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. Plant J 93:338–354

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, Harriman J, Glaubitz JC, Buckler ES, Kresovich S (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci U S A 110:453–458

Mundia CW, Secchi S, Akamani K, Wang G (2019) A regional comparison of factors affecting global sorghum production: the case of North America, Asia and Africa's Sahel. Sustainability 11:2135

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

National Research Council (1996) Lost crops of Africa, vol 1. Natl Acad Press, Washington, DC

Oliver AL, Grant RJ, Pedersen JF, O'Rear J (2004) Comparison of brown midrib-6 and -18 forage sorghum with conventional sorghum and corn silage in diets of lactating dairy cows. J Dairy Sci 87:637–644

Ordonia R, Ito Y, Morinaka Y, Sazuka T, Matsuaka M (2016) Molecular breeding of Sorghum bicolor, a novel energy crop. Int Rev Cell Mol Biol 321:221–257

Pillonel C, Mulder MM, Boon JJ, Forster B, Binder A (1991) Involvement of cinnamyl-alcohol dehydrogenase in the control of lignin formation in Sorghum bicolor L Moench. Planta 185:538–544

Ping J, Zhang F, Niu H, Yang H, Lv X, Du Z, Li H, Wang Y (2018) Genetic diversity analysis of germplasm resources of forage Sorghum based on SSR marker. Mol Plant Breed 14:4663–4670

Porter KS, Axtell JD, Lechtenberg VL, Colenbrander VF (1978) Phenotype, fiber composition, and in-vitro dry-matter disappearance of chemically-induced brown midrib (bmr) mutants of sorghum. Crop Sci 18:205–208

Quinby JR (1975) The genetics of sorghum improvement. J Hered 66:56–62

Rest BVD, Danoun S, Boudet AM, Rochange SF (2006) Down-regulation of cinnamoyl-CoA reductase in tomato (*Solanum lycopersicum* L.) induces dramatic changes in soluble phenolic pools. J Exp Bot 57:1399–1411

Saballos A, Sattler SE, Sanchez E, Foster TP, Xin Z, Kang C, Pedersen JF, Vermerris W (2012) Brown midrib2 (Bmr2) encodes the major 4-coumarate:coenzyme A ligase involved in lignin biosynthesis in sorghum (*Sorghum bicolor* (L.) Moench). Plant J 70:818–830

Scully ED, Gries T, Funnell-Harris DL, Xin Z, Kovacs FA, Vermerris W, Sattler SE (2016) Characterization of novel Brown midrib 6 mutations affecting lignin biosynthesis in sorghum. J Integr Plant Biol 58:136–149

Shadle G, Chen F, Srinivasa Reddy MS, Jackson L, Nakashima J, Dixon RA (2007) Down-regulation of hydroxycinnamoyl CoA: Shikimate hydroxycinnamoyl transferase in transgenic alfalfa affects lignification, development and forage quality. Phytochemistry 68:1521–1529

Shoemaker C, Bransby D (2010) The role of sorghum as a bioenergy feedstock. Sustainable alternative fuel feedstock opportunities, challenges and roadmaps for six US regions. Soil and Water Conservation Society, Atlanta, pp 149–159

The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851–861

The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678

Tripathi SC, Sayre KD, Kaul JN, Narang RS (2003) Growth and morphology of spring wheat (*Triticum aestivum* L.) culms and their association with lodging: effects of genotypes, N levels and ethephon. Field Crop Res 84:271–290

Upadhyaya HD, Pundir RPS, Dwivedi SL, Gowda CL, Reddy VG, Sube S (2009) Developing a mini core collection of sorghum for diversified utilization of germplasm. Crop Sci 49:1769–1780

Vermerris W (2011) Survey of genomics approaches to improve bioenergy traits in maize, sorghum and sugarcane. J Integr Plant Biol 53:105–119

Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38:e164

Zhang S, Jing Z, Wu F, Luan R, Wang J, Zhang Y, Wang X (2018) The agronomic characters and SRAP genetic diversity analysis of 21 introduced male sterile lines of brown midrib forage sorghum. Pratacultural Sci 35:1072–1079

Zheng L, Guo X, He B, Sun L, Peng Y, Dong S, Liu T, Jiang S, Ramachandran S, Liu C, Jing H (2011) Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). Genome Biol 12:R114