



# Efficient Identification of Causal Mutations through Sequencing of Bulk F<sub>2</sub> from Two Allelic Bloomless Mutants of *Sorghum bicolor*

Yinping Jiao<sup>1,2</sup>, Gloria Burow<sup>1</sup>, Nicholas Gladman<sup>1</sup>, Veronica Acosta-Martinez<sup>1</sup>, Junping Chen<sup>1</sup>, John Burke<sup>1</sup>, Doreen Ware<sup>2,3\*</sup> and Zhanguo Xin<sup>1\*</sup>

<sup>1</sup> Cropping Systems Research Laboratory, Agricultural Research Service (USDA), Lubbock, TX, United States, <sup>2</sup> Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, United States, <sup>3</sup> USDA-ARS NAA Plant, Soil and Nutrition Laboratory Research Unit, Cornell University, Ithaca, NY, United States

## OPEN ACCESS

### Edited by:

Paul Gepts,  
University of California, Davis,  
United States

### Reviewed by:

Liming Xiong,  
Texas A&M University, United States  
Alfred (Heqiang) Huo,  
University of Florida, United States

### \*Correspondence:

Doreen Ware  
ware@cshl.edu  
Zhanguo Xin  
zhanguo.xin@ars.usda.gov

### Specialty section:

This article was submitted to  
Plant Genetics and Genomics,  
a section of the journal  
Frontiers in Plant Science

**Received:** 09 October 2017

**Accepted:** 27 December 2017

**Published:** 12 January 2018

### Citation:

Jiao Y, Burow G, Gladman N,  
Acosta-Martinez V, Chen J, Burke J,  
Ware D and Xin Z (2018) Efficient  
Identification of Causal Mutations  
through Sequencing of Bulk F<sub>2</sub>  
from Two Allelic Bloomless Mutants of  
*Sorghum bicolor*.  
Front. Plant Sci. 8:2267.  
doi: 10.3389/fpls.2017.02267

*Sorghum* (*Sorghum bicolor* Moench, L.) plant accumulates copious layers of epi-cuticular wax (EW) on its aerial surfaces, to a greater extent than most other crops. EW provides a vapor barrier that reduces water loss, and is therefore considered to be a major determinant of sorghum's drought tolerance. However, little is known about the genes responsible for wax accumulation in sorghum. We isolated two allelic mutants, *bloomless40-1* (*bm40-1*) and *bm40-2*, from a mutant library constructed from ethyl methane sulfonate (EMS) treated seeds of an inbred, BTx623. Both mutants were nearly devoid of the EW layer. Each *bm* mutant was crossed to the un-mutated BTx623 to generate F<sub>2</sub> populations that segregated for the *bm* phenotype. Genomic DNA from 20 *bm* F<sub>2</sub> plants from each population was bulked for whole genome sequencing. A single gene, Sobic.001G228100, encoding a GDSL-like lipase/acylhydrolase, had unique homozygous mutations in each bulked F<sub>2</sub> population. Mutant *bm40-1* harbored a missense mutation in the gene, whereas *bm40-2* had a splice donor site mutation. Our findings thus provide strong evidence that mutation in this GDSL-like lipase gene causes the *bm* phenotype, and further demonstrate that this approach of sequencing two independent allelic mutant populations is an efficient method for identifying causal mutations. Combined with allelic mutants, MutMap provides powerful method to identify all causal genes for the large collection of *bm* mutants in sorghum, which will provide insight into how sorghum plants accumulate such abundant EW on their aerial surface. This knowledge may facilitate the development of tools for engineering drought-tolerant crops with reduced water loss.

**Keywords:** *Sorghum bicolor*, mutation mapping, bloomless, epicuticular wax, cutin synthase

## INTRODUCTION

Forward genetics is a powerful approach to identifying gene mutations that cause phenotypes of interest, and ultimately elucidating the biochemical and molecular mechanisms and signaling processes underlying plant development and adaptation. The forward genetic approach begins with physical or chemical mutagenesis of a pure (inbred) line, followed by screening for mutants with

the desired phenotype. The resultant mutants are outcrossed to another pure line with extensive DNA polymorphism relative to the mutagenized parental line. Linkage of the mutant phenotype with DNA markers is analyzed, using as many markers as possible, until the mutation is mapped to a very small region flanked by two markers. The success of this process, also referred to as map-based cloning, depends on two critical factors: a large segregating mapping population and dense DNA markers with the resolution to delimit the mutation into a small region. Several hundreds to thousands of accurately phenotyped  $F_2$  plants must be analyzed with large number of DNA markers in order to narrow the mutation to a region harboring only a few genes (Jander et al., 2002). The entire region is then sequenced in order to identify the gene that carries the causal mutation. To confirm the identity of the gene, the wild-type gene is introduced into the mutant through transformation to determine whether it can complement the mutant phenotype. This last step can be bypassed if two or more independent mutant alleles are identified that carry unique mutations in the same gene because the odd of finding two independent mutations by chance on the same gene that lead to the same phenotype is near zero (Jander et al., 2002).

Next-generation sequencing (NGS) techniques provide massive high-resolution genotypic data very quickly (Metzker, 2010), and can thus replace the costly and time-consuming linkage analysis required by conventional map-based cloning (Schneeberger and Weigel, 2011). Two general approaches have been developed for using NGS to identify causal mutations. The first strategy is represented by two similar approaches, mapping-by-sequencing (ShoreMap) (Schneeberger et al., 2009; Hartwig et al., 2012) and next-generation mapping (NGM) (Austin et al., 2011). These approaches are very similar to conventional map-based cloning. First, a mapping population is generated by outcrossing the mutant to a divergent line with sufficient DNA polymorphism, and homozygous mutants are selected from the segregating  $F_2$  population. Then, equal quantities of genomic DNA from each of the selected lines are pooled for NGS analysis, according to the principle of bulk segregant analysis (BSA) (Michelmore et al., 1991). Because only homozygous mutants selected from the mapping  $F_2$  population are subjected to whole-genome sequencing, the single-nucleotide polymorphic (SNP) markers surrounding the causal mutation will be derived predominantly from the mutant parent. Consequently, the region containing the causal mutation will form a “valley” devoid of polymorphic markers from the divergent parent. The region containing the causal mutation can be narrowed further with sophisticated bioinformatics tools, sometimes, to a single gene (Austin et al., 2011; Schneeberger and Weigel, 2011).

The second approach is represented by isogenic mapping-by-sequencing (MutMap), a variation of ShoreMap, designed to map mutant phenotypes subject to modification by genetic background (Abe et al., 2012; Hartwig et al., 2012; Zhu et al., 2012). In MutMap, the mutant is crossed to a wild-type plant of the same genetic background as the mutant, rather than to a divergent line as in ShoreMap, to avoid interference from a divergent genetic background. Homozygous mutants are selected from the backcrossed  $F_2$  population and bulked for whole-genome sequencing. After identification of SNPs, the SNP ratio

(i.e., the proportion of all short reads that contain the mutated SNPs) is plotted along each chromosome. The causal mutation for the phenotype is expected to have a SNP ratio of 1 (i.e., 100% mutant), whereas unlinked background mutations are expected to have ratios around 0.5. For mutations close to the causal mutation, the SNP ratio should vary from  $<0.5$  to 1 depending on genetic distance. This approach is particularly useful for mapping induced mutations (Abe et al., 2012; Zhu et al., 2012). Because major mutations capable of causing phenotypes are rare in induced mutant populations (Greene et al., 2003; Henry et al., 2014; Jiao et al., 2016; Krasileva et al., 2017), it is often possible to narrow down the candidates to one or a few mutations using 20–50 homozygous mutants selected from a segregating  $F_2$  population (Abe et al., 2012; Takagi et al., 2015).

Sorghum is the fifth most important crop in the world, providing food, feed, and forage to humans and agricultural animals in many areas around the world (<http://www.fao.org/home/en/>). In addition, due to its excellent tolerance of drought and high-temperature stress, high water use efficiency, and high productivity, it has become an increasingly important crop for bioenergy production (Rooney et al., 2007; Vermerris, 2011; de Siqueira Ferreira et al., 2013; Mullet et al., 2014). Sorghum accumulates higher levels of epicuticular wax (EW) on its aerial surface than most other crops, and this trait is considered to make an important contribution to sorghum's high water use efficiency and drought tolerance (Jordan et al., 1984; Jenks et al., 1992; Premachandra et al., 1994; Samdur et al., 2003; Burrow et al., 2008; Uttam et al., 2017). Many sorghum mutants that are devoid of EW, named “*bloomless*” (*bm*), have been isolated over the years but few genes responsible for the deposit of massive EW on its aerial surface have been identified (Weibel, 1986a,b; Peters et al., 2009; Jiao et al., 2016).

Here, we report an efficient method for identifying causal gene mutations by whole genome re-sequencing of bulked  $F_2$  populations, and demonstrate the approach using two independent alleles of sorghum *bm* mutants devoid of EW. We chose the *bm* mutants for allelic mutant mapping because we have isolated a large collection of *bm* mutants from the sorghum mutant library generated with ethyl methane sulfonate (EMS) in an elite inbred line BTx623 and the characteristic shining green sheath that can be easily distinguished from the sheath of wild type sorghum (Burrow et al., 2008, 2009; Peters et al., 2009). Our method offers a tool to identify most *Bm* genes in sorghum. These *Bm* genes may facilitate improvement of water use efficiency and drought tolerance, not only in sorghum, but also in other crops (Weibel, 1986a,b; Peters et al., 2009; Jiao et al., 2016; Uttam et al., 2017).

## MATERIALS AND METHODS

### Generation of Sorghum Pedigreed Mutant Library and Selection of *bm* Mutants

The sorghum pedigreed mutant library was established as described (Xin et al., 2008). During the development of the mutant library, each head row at the  $M_3$  stage was evaluated for mutants that were devoid of EW on the sheath of the flag leaf

at the boot stage. The *bm* mutants could be easily identified from the shiny boot when the flag leaf was fully expanded. Over the last few years, over 100 independent *bm* mutants have been identified from this mutant population (Jiao et al., 2016). The putative *bm* mutants were crossed to BTx623<sub>ms8</sub>, a near-isogenic line of BTx623 with a nuclear male sterile mutation (Xin et al., 2017). The nature of the mutation (recessive/dominant) was determined in the F<sub>1</sub> plants from the cross: if the F<sub>1</sub> plants produced high levels of EW, the mutation would be recessive, whereas if the F<sub>1</sub> plants had the *bm* phenotype, the mutation would be dominant. All *bm* mutants isolated so far are recessive. Only *bm* mutants that exhibited a single Mendelian segregation ratio (1 *bm*: 3 WT) in F<sub>2</sub> progeny were considered to be confirmed *bm* mutants. Efforts to determine the allelic groups of these *bm* mutants are ongoing, but are far from complete due to the large number of complementation crosses required.

## Phenotype Analysis

EW coverage was analyzed by scanning electron microscopy (SEM) at the College of Arts & Sciences Microscopy of Texas Tech University. Fresh leaf and sheath tissues were mounted on a sample holder and imaged directly at 15 KeV accelerating voltage under air pressure of 30–90 Pa on a Hitachi S-4300 field emission environmental scanning electron microscope. To determine the wax load, BTx623 and two *bm* mutants were planted in a greenhouse at 28/20°C (day/night) temperatures on May 25, 2017. At boot stage, when the flag leaf was fully expanded, the plants were photographed and subjected to EW analysis. EW from the leaf and sheath of the second leaf from the top (the leaf right below the flag leaf) was extracted by dipping a cut sample of ~6 cm<sup>2</sup> in 6 ml of hexane for 60 s. Samples were dried under nitrogen, subjected to esterification using the MIDI protocol (<http://www.midi-inc.com>), dried again under nitrogen, and reconstituted in hexane containing an internal standard. Wax analysis was performed using a gas chromatograph coupled with a flame ionization detector (FID).

## Allelic Bulk Segregant Analysis

Two allelic *bm* mutants were selected for this study. Previously, 39 complementation groups of sorghum *bm* mutant have been reported (Weibel, 1986a,b; Peters et al., 2009; Uttam et al., 2017). Therefore, we tentatively named the two allelic *bm* mutants as *bm40-1* and *bm40-2* because their allelic relationship to previously reported *bm* mutants has not been tested. Two F<sub>2</sub> populations were developed by crossing each *bm* mutant to BTx623<sub>ms8</sub>, a nuclear male sterile near-isogenic line (NIL) of BTx623 (Xin et al., 2017). From the F<sub>2</sub> populations, 20 *bm* mutants were selected for preparation of genomic DNA according to the method as described in Xin and Chen (2012). Equal amounts of DNA were bulked and subjected to 150-bp paired-end sequencing on an Illumina X-10 instrument (<https://en.novogene.com>). About 10 Gb of high-quality sequence, corresponding to ~15× coverage of the whole genome, was obtained for each bulked F<sub>2</sub> population. The data from the two F<sub>2</sub> population were first analyzed separately using our highly efficient pipeline, *ems\_mutation*, in the Cyverse Discovery App (Merchant

et al., 2016) (<https://pods.iplantcollaborative.org/wiki/display/TUT/EMS+mutant+sites+identification+Workflow>). Briefly, the reads were first aligned to sorghum reference genome version 2 using Bowtie2 in “very-sensitive” mode. Samtools was used to sort and change the format of the alignment file (Li et al., 2009; Paterson et al., 2009). PCR duplications in the reads were masked using Picard-tools-1.113 (<http://broadinstitute.github.io/picard/>). Variation call process was performed by Samtools and Bcftools (Li, 2011). Only EMS-induced type GC→AT single-nucleotide changes supported by 5–100 reads were retained for downstream filtering. Background variations from the parental line were filtered according to two datasets: the variations called from whole-genome sequencing of the parental line BTx623, and the high-frequency (≥0.05) allele variations in the sequenced mutants from the same population (Jiao et al., 2016). Based on the high accuracy of phenotyping, the SNP ratio ([number of reads supporting the mutation allele]/[number of all reads mapped to the loci]) cutoff was set to 1. SNPs with large effects on genes (missense, nonsense, splice\_site\_acceptor, splice\_site\_donor) predicted by SNPeff (Cingolani et al., 2012) were retained as candidate causal mutations. The candidate mutations from two allelic F<sub>2</sub> populations were then compared to determine the overlap.

To analyze the phylogenetic relationship of GDSL lipases, phylogenetic tree was constructed using Mega 7 (Kumar et al., 2016) with ClustalW alignment and the maximum likelihood method from all sequences and orthologs extracted by Gramene BioMart (Tello-Ruiz et al., 2016). Multiple alignments from ClustalW were visualized with MSAviewer (Yachdav et al., 2016).

## RESULTS

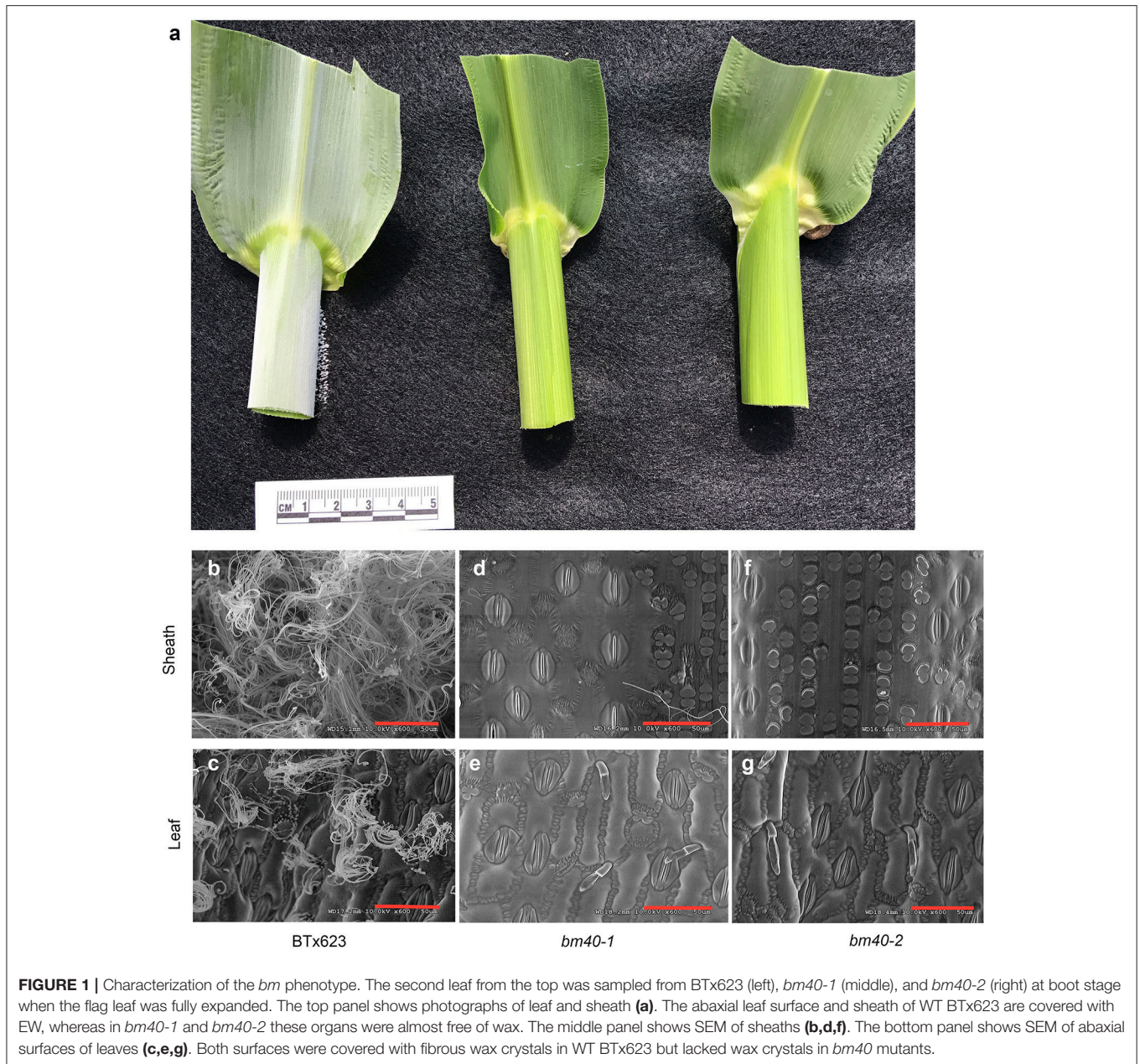
### Deficiency of EW in *bm40* Mutants

Two recessive allelic *bm* mutants, *bm40-1* and *bm40-2*, isolated from the sorghum pedigreed mutant library (Xin et al., 2008; Jiao et al., 2016) were used to identify the causal gene for *bm40* mutants. The aerial surfaces of plants of the inbred line BTx623, which was used to generate the pedigreed mutant population, were covered with conspicuous white crystals of EW, especially on the surface of the leaf sheath (Figure 1). In BTx623, the sheath obviously accumulated more EW than leaves (Figures 1a–c). SEM of the WT sheath revealed that the stomata on the sheath were completely covered with fibrous wax crystals, whereas the abaxial surface of leaf blade was only partially covered (Figures 1b,c). In the *bm40-1* and *bm40-2* mutants, the surfaces of both leaves and sheaths were devoid of wax crystals, and the stomata were completely exposed (Figures 1a,d,e,f,g).

To measure wax load in BTx623 and the *bm* mutants, pieces of sheaths and leaves, ~6 cm<sup>2</sup>, were dipped quickly to hexane to extract wax. The extracted EW was quantified by weight after hexane evaporated. In comparison with the wild type BTx623, the wax loads of leaves and sheath surfaces in the mutant were reduced by 80–84% and >92%, respectively (Table 1).

To examine the fatty acid composition of the extracted EW, the lipids extracted from the surfaces of the WT and *bm* mutants were esterified and analyzed by GC. Consistent with





**TABLE 1 |** Analysis of leaf and sheath epicuticular wax (EW) content of WT and bloomless sorghum mutant plants at boot stage, and estimation of the amount of EW reduction in mutants vs. WT.

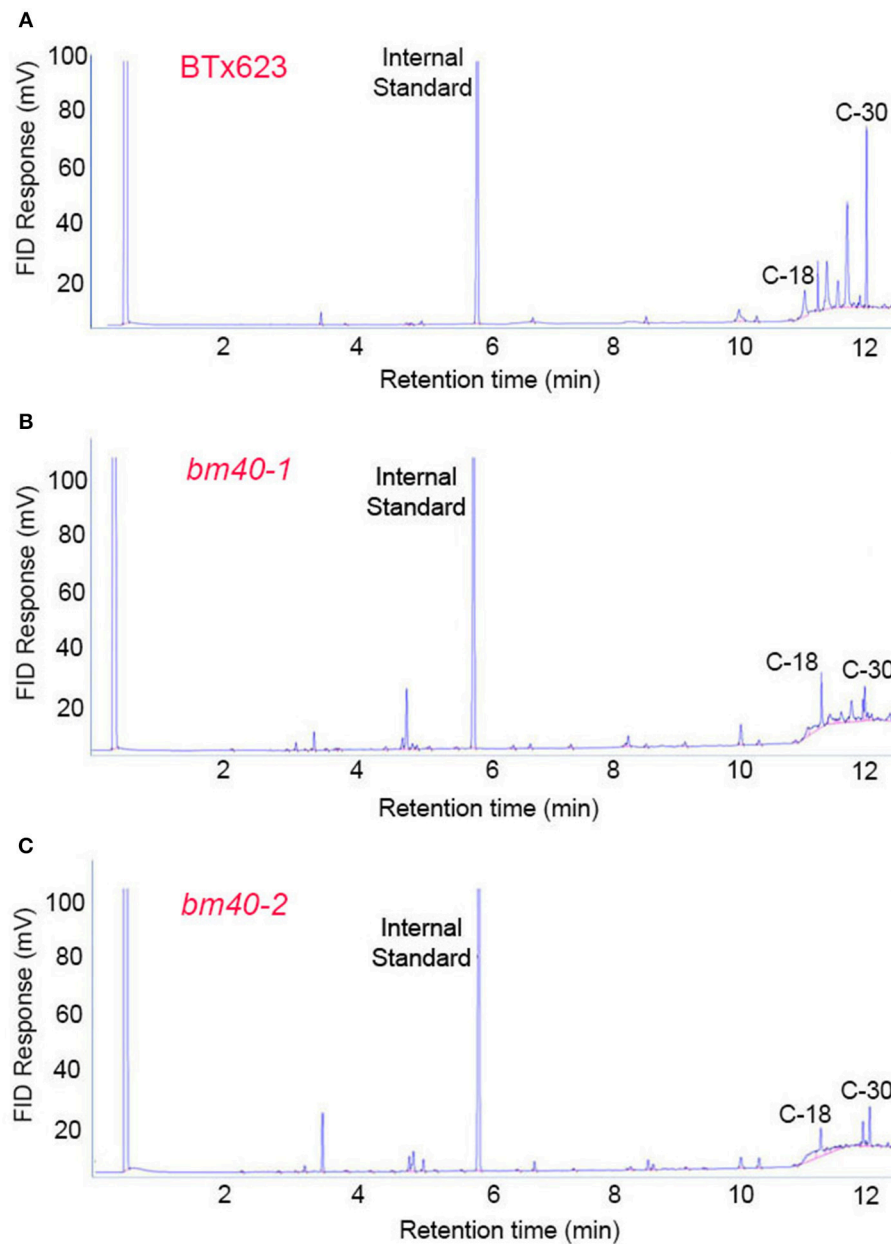
Genotype	Leaf		Sheath	
	EW (mg/cm <sup>2</sup> )	%EW reduction*	EW (mg/cm <sup>2</sup> )	%EW reduction*
WT	2.12 (0.15)	0	5.67 (0.33)	0
<i>bm40-1</i>	0.34 (0.15)	84.00	0.44 (0.19)	92.16
<i>bm40-2</i>	0.42 (0.15)	80.00	0.44 (0.19)	92.16

Values inside the parentheses are standard deviation. \*%EW reduction was estimated as:  $=[(WT\ EW) - Bloomless1(EW)]/WT\ (EW)] \times 100$ .

the wax load analysis, BTx623 accumulated large quantity of C30 and other long chain fatty acids (Figure 2A) on the surface of leaf sheath, which were greatly reduced in the *bm* mutants (Figures 2B,C).

### The Causal Gene in *bm40* Encodes a Cutin Synthase

To identify the causal mutation, both mutant strains were subjected to BSA (Figure 3A). For each F<sub>2</sub> population, a total of 20 F<sub>2</sub> *bm* type individuals were pooled for sequencing, yielding 12 and 18 Gb 150-bp paired-end data for *bm40-1* and *bm40-2*, respectively. After background mutations were

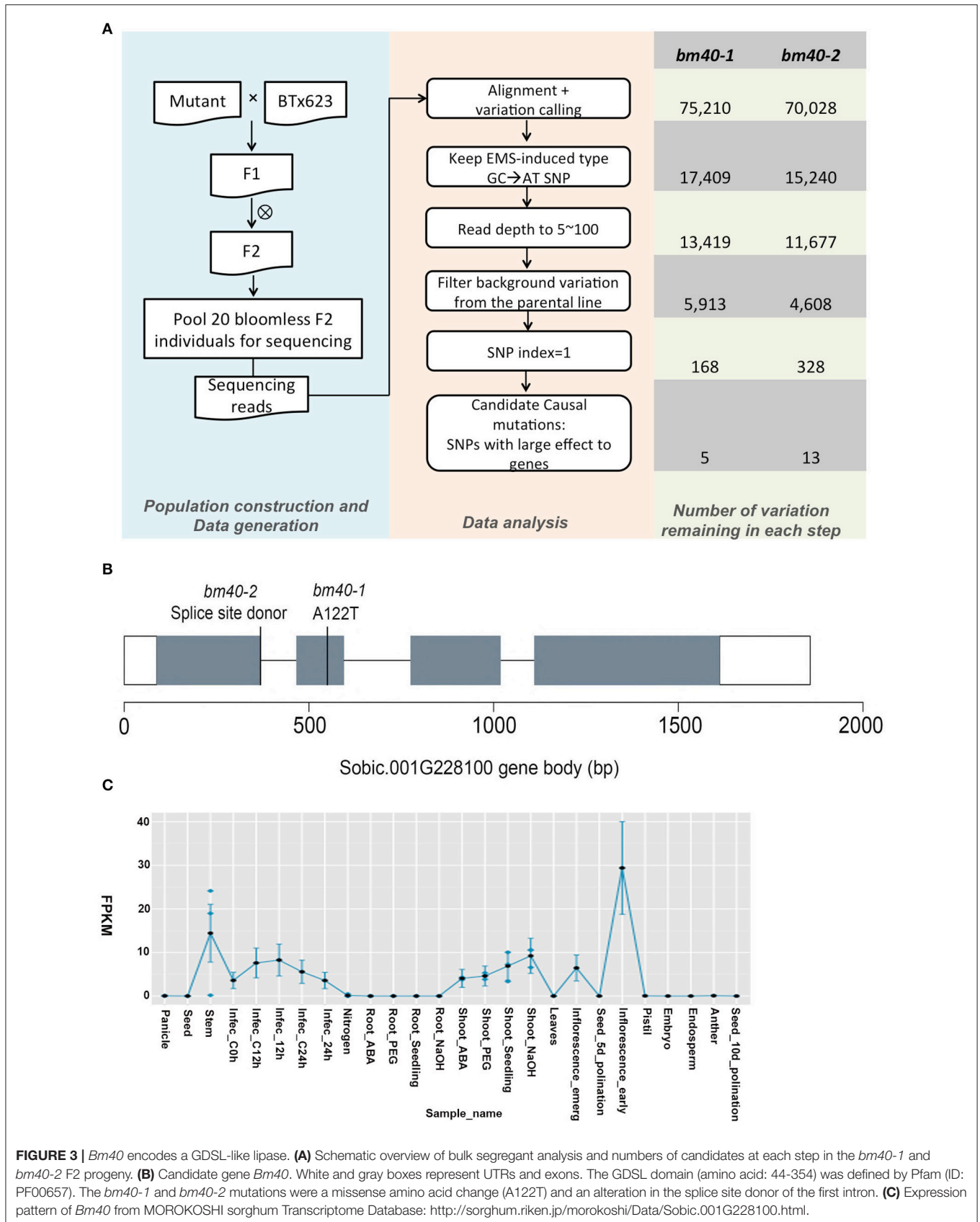


**FIGURE 2** | Fatty acid profile of surface wax from wild-type BTx623 (A) and the two bloomless mutants (B,C). Epicuticular waxes were extracted by quick immersion in hexane. After esterification, the fatty acids were analyzed by GC (with a FID detector). The levels of long-chain fatty acids (C-30) were greatly reduced in both *bm* mutants in comparison with BTx623.

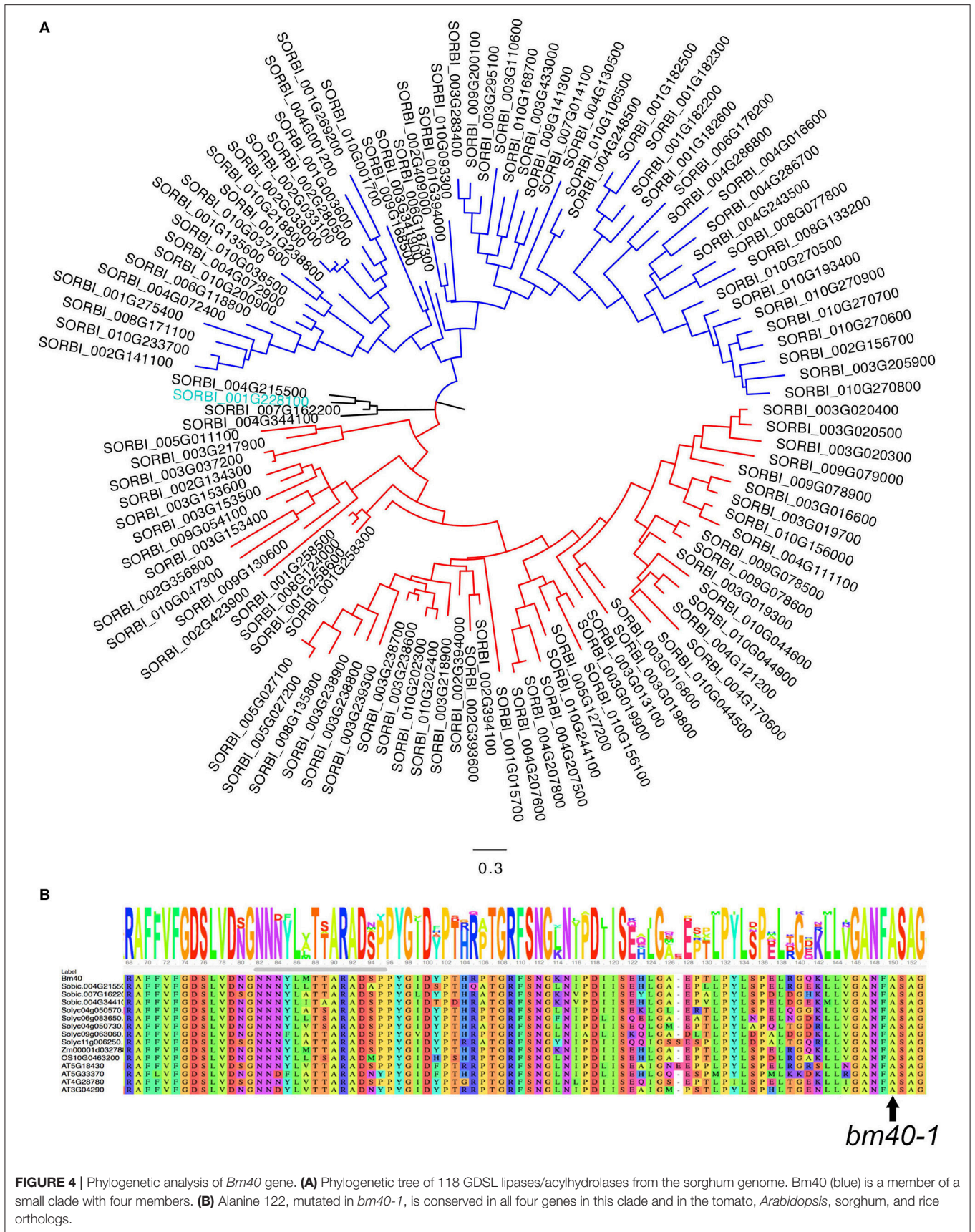
filtered out using our high-efficiency data analysis pipeline, 5 and 13 homozygous mutations remained for bulked F<sub>2</sub> of *bm40-1* and *bm40-2*, respectively. Comparison of the two bulked F<sub>2</sub> populations revealed that only one gene, Sobic.001G228100, was present in both populations. In *bm40-1*, a G→A mutation at genomic DNA position chr01\_21842959 results in substitution of alanine at amino acid position 122 with threonine (A122T), whereas in *bm40-2*, a G→A mutation

at position chr01\_21843140 alters a splice site after the first exon (Figure 3B).

The *bm40* gene encodes a GDSL-like lipase/acylhydrolase with 383 amino acids, and both mutations are in the GDSL-like lipase domain (Pfam ID: PF00657). The protein is similar to cutin synthase 1 (CD1) in tomato (Yeats et al., 2014). According to the public sorghum expression atlas data (Makita et al., 2015), the gene is highly expressed in aerial tissues but undetectable in roots







**FIGURE 4 |** Phylogenetic analysis of *Bm40* gene. **(A)** Phylogenetic tree of 118 GDSL lipases/acylhydrolases from the sorghum genome. *Bm40* (blue) is a member of a small clade with four members. **(B)** Alanine 122, mutated in *bm40-1*, is conserved in all four genes in this clade and in the tomato, *Arabidopsis*, sorghum, and rice orthologs.

and seeds, consistent with its role in cutin synthesis on aerial surfaces (**Figure 3C**).

Querying the sorghum genome with the Pfam domain of GDSL returned a total of 118 genes that could be grouped into three clades (**Figure 4A**); *Bm40* belongs to a small clade with four members. Alignment of these four proteins with the *Arabidopsis* and tomato orthologs of BM40 revealed that the site of the missense mutation in *bm40-1*, A122T, is conserved across the sorghum paralogs in this clade and *Bm40* orthologs across species (**Figure 4B**).

The GDSL lipase/hydrolases are a large newly discovered gene superfamily present in many kingdoms of life, including plants (Ling, 2008; Chepyshko et al., 2012; Lai et al., 2017). These enzymes have a broad substrate spectrum and may be involved in a wide range of metabolic pathways. Although large numbers of this family are present in plants, few genes have been assigned specific functions. One member of the superfamily was shown to be the gene encoding cutin synthase, *CD1*, in tomato (Yeats et al., 2012b). *CD1*, which is expressed between the cell wall and cuticle layer, is responsible for synthesis of the polyester component of the cuticular network. *Arabidopsis Cus2*, a homolog of the tomato *CD1*, is involved in synthesis and maintenance of cuticle ridges on flower petals. Sorghum *Bm40* is closely related to both genes. Moreover, the large reduction in EW in *bm40* mutants was also similar to the tomato *cd1* lines lack of EW (Yeats et al., 2012a).

## DISCUSSION

Here, we demonstrated that sequencing of two independent allelic mutants represents a powerful approach to identifying causal gene mutations in sorghum. Our approach is very similar to MutMap through bulked segregant analysis, described in rice (Abe et al., 2012). Conceptually, MutMap is a straightforward way to identify the causal homozygous mutation in bulked  $F_2$  plants selected for a phenotype of interest. Because the mutant phenotype is preselected, the causal mutation is expected to present in all short sequence reads covering the mutation site, whereas the ratio of mutant to WT SNPs will be 0.5 for all unlinked mutations. For linked mutations, this ratio varies from 1 to  $<0.5$  depending the distance to the causal mutation. In practice, however, the situation is more complicated. Sequencing errors in the reference genome and pre-existing SNPs in BTx623 prior to mutagenesis will all be identified as homozygous mutations (**Figure 3A**), making it challenging to determine which homozygous mutations are causal. In a previous study, we sequenced BTx623 and 256 selected lines from a mutant population generated in this line (Jiao et al., 2016). From sequenced BTx623 and mutant lines, we annotated over 80,000 SNPs that differ from the sorghum reference genome sequence (Paterson et al., 2009). These SNPs serve as a powerful filter for removing background SNPs from BSA. For example, during the mapping of both *bm40-1* and *bm40-2*, huge number of mutations (13,419 in *bm40-1* and 11,677 in *bm40-2*) mutations were identified in the bulked *bm40-1* and *bm40-2* populations. After removing background mutations

by subtracting the 80,000 SNPs annotated from the sequenced mutant lines, the numbers of raw mutations were greatly reduced (**Figure 3A**).

Another challenge of BSA is identification of the causal mutation from closely linked mutations, but again, this hurdle can be overcome by sequencing two or more allelic mutants. In our study, we identified 5 linked homozygous mutations in *bm40-1* and 13 in *bm40-2*. Comparison of these two populations revealed that only one gene, Sobic.001G228100, had homozygous mutations in both bulked  $F_2$  sequences. Furthermore, the mutation site in *bm40-1* bulked  $F_2$  was a missense mutation in the conserved GDSL domain, whereas the mutation in *bm40-2* bulked  $F_2$  altered the splice junction from the donor side. It is less likely for two allelic mutants with same phenotypes to have mutations in different genes. Thus, this gene is most likely the candidate for *Bm40*. Based on the average number of deleterious mutations of 147 per line in the sequenced mutant library (Jiao et al., 2016), the gene mutation rate per line per generation would about 0.004 (147/34,000) by chance. For two mutations occurred randomly in one gene would be  $1.6 \times 10^{-5}$ , a nearly impossible event. Similar approach has been reported in identifying the causal mutations that affect germination of lettuce seeds under high temperatures (Huo et al., 2016). Because of the advantage of sequencing two independent alleles by BSA in overcoming the limitations of MutMap, This method should be named allelic MutMap. This method is an efficient and cost-effective approach to identifying causal gene mutations when an allelic mutant series is available. Based on current sequencing prices, the cost to sequence two bulked samples to 10 Gb data or  $\sim 15x$  coverage is  $< \$500$  in sorghum.

For a large number of mutants with similar phenotypes, such as the sorghum *bm* mutants, it is impractical to perform pairwise allelic test. It may be possible to sequence many bulked  $F_2$  populations first without prior knowledge of allelic relationship. If two or more bulked  $F_2$  population harbor unique mutations on a common gene, complementation tests can then be performed only among mutants with the same mutated gene to confirm that they are allelic. Reciprocal crosses and cosegregation analysis maybe required to further confirm the identification of the causal gene. In this manner, the time and expense required for complementation testing among large numbers of mutants could be reduced substantially. In light of the continuing decrease in sequencing costs and increase in data output and quality, it will soon be feasible to use this approach to sequence large numbers of bulked  $F_2$  populations. It is likely that some bulked  $F_2$  populations possess several linked gene mutations but no common mutated genes are found from other bulked  $F_2$ . In these cases, complementation of the mutant phenotype with the wild type gene through transformation will be needed to confirm the identification of the causal mutation.

The importance of epicuticular wax to the drought tolerance in sorghum has long been recognized. Many *bm* mutants have been registered and genetically tested (Weibel, 1986a,b; Peters et al., 2009). A total of 38 *bm* loci were identified through genetic complementation (Peters et al., 2009), and a *bm* locus from a natural mutation was recently mapped (Uttam et al., 2017), bringing the total number of *bm* loci to 39. From the



pedigreed mutant library, we identified 107 independent *bm* mutants (Jiao et al., 2016). Despite the large number of sorghum *bm* mutants collected thus far, few causal genes have been identified. Based on the *Arabidopsis* gene networks involved in EW accumulation (Kunst and Samuels, 2003; Bernard and Joubès, 2013; Fich et al., 2016), we searched the 256 sequenced core mutant lines and identified seven sorghum homologs in which mutations cause *bm* phenotypes (Jiao et al., 2016). Due to the massive accumulation of EW in sorghum, it is reasonable to speculate that sorghum employs multiple genes to regulate wax accumulation in its aerial surface. Identification of these genes and elucidation of their regulation may facilitate manipulation of surface wax accumulation in other crops with the goal of increasing their drought tolerance. To identify the genes involved in EW accumulation in sorghum, we established an efficient forward genetic method to identify the causal mutations defined by the vast collection of sorghum *bm* mutants, which can potentially identify all causal mutations of sorghum *bm* mutants in the near future.

Another GDSL-like lipase gene, Sobic.001G269200, was implicated in a *bm* QTL mapped from a natural accession RS647 (Uttam et al., 2017). However, because Sobic.001G269200 is more than 13 Mb from the *bm40* on chromosome 1 and belongs to a different clade (Figure 4A), these two GDSL lipases may represent two separate genes. To determine whether this gene is also involved in EW accumulation, we searched the sequenced mutant database (Jiao et al., 2016). We identified 10 mutations in the gene, including one stop-gain mutation in the sequenced mutant line ARS192. However, none of these 10 lines segregated for the *bm* phenotype (Supplementary Table 1). We also identified a heterozygous mutation in *Bm40*, Sobic.001G228100, in ARS29. This line was shown to segregate for the *bm* phenotype as a single recessive gene mutation (Supplementary Table 1). It is unclear if the *bm* QTL mapped from RS647 is related to *Bm40*. Nevertheless, *Bm40* is likely to be a functional orthology of the tomato *CD1* based on phenotypic similarity and protein sequence identity, but further evidence, such as enzyme assays or lipidomic profiling, is required to firmly establish whether *Bm40* is indeed a cutin synthase. The allelic mutants and identification of the causal gene provide critical materials for such future studies.

## CONCLUSION

Using allelic MutMap, we showed that the causal gene in *bm40* mutants encodes a GDSL-like lipase, similar to tomato cutin

## REFERENCES

- Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* 30, 174–178. doi: 10.1038/nbt.2095
- Austin, R. S., Vidaurre, D., Stamatiou, G., Breit, R., Provart, N. J., Bonetta, D., et al. (2011). Next-generation mapping of *Arabidopsis* genes. *Plant J.* 67, 715–725. doi: 10.1111/j.1365-3113.2011.04619.x
- Bernard, A., and Joubès, J. (2013). *Arabidopsis* cuticular waxes: advances in synthesis, export and regulation. *Prog. Lipid Res.* 52, 110–129. doi: 10.1016/j.plipres.2012.10.002
- Burow, G. B., Franks, C. D., Acosta-Martinez, V., and Xin, Z. (2009). Molecular mapping and characterization of BLMC, a locus for profuse wax (bloom) and enhanced cuticular features of sorghum [*Sorghum bicolor* (L.) Moench.]. *Theor. Appl. Genet.* 118, 423–431 doi: 10.1007/s00122-008-0908-y
- Burow, G. B., Franks, C. D., and Xin, Z. (2008). Genetic and physiological analysis of an irradiated bloomless mutant (epicuticular wax mutant) of sorghum. *Crop Sci.* 48, 41–48. doi: 10.2135/cropsci2007.02.0119
- Chepyshko, H., Lai, C. P., Huang, L. M., Liu, J. H., and Shaw, J. F. (2012). Multifunctionality and diversity of GDSL esterase/lipase gene family in rice (*Oryza sativa* L. *japonica*) genome: new insights from

synthase 1. The identity of the *Bm40* gene is supported by two independent allelic mutations and a nonsynonymous mutation from a sequenced mutant library. Our results demonstrate that the allelic MutMap is an efficient and cost-effective approach for identifying causal gene mutations. This approach could be used to identify causal mutations for other phenotypes for which allelic mutants are available. Identification of the causal gene for the *bm* phenotype provides critical insight into the accumulation of massive amounts of EW on the aerial surface of sorghum plant and the superior drought tolerance of this crop.

## DISCLAIMER

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

## AUTHOR CONTRIBUTIONS

ZX and DW: Conceived the idea of this work; ZX: Isolated the mutants; ZX, GB, NG, VA-M, JC, and JB: Conducted the phenotype analysis; YJ: Analyzed the genomic data for the identification of the gene; YJ and ZX: Drafted the manuscript. All authors participated in the revision and agree with the final manuscript.

## ACKNOWLEDGMENTS

YJ, GB, JB, and ZX would gratefully acknowledge support from the United Sorghum Checkoff program and USDA ARS 3096-21000-020-00D. YJ, NG, and DW were partly supported by USDA ARS 8062-21000-041-00D. NG is supported by an ARS-funded postdoc fellowship. We thank Bo Zhao from the College of Arts & Sciences Microscopy of Texas Tech University for assistance with SEM, and Lan Liu-Gitz, Halee Hughes, and John Cotton for technical support.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.02267/full#supplementary-material>

- bioinformatics analysis. *BMC Genomics* 13:309. doi: 10.1186/1471-2164-13-309
- Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- de Siqueira Ferreira, S., Nishiyama, M. Y. Jr., Paterson, A. H., and Souza, G. M. (2013). Biofuel and energy crops: high-yield Saccharinae take center stage in the post-genomics era. *Genome Biol.* 14:210. doi: 10.1186/gb-2013-14-6-210
- Fich, E. A., Segerson, N. A., and Rose, J. K. (2016). The plant polyester cutin: biosynthesis, structure, and biological roles. *Annu. Rev. Plant Biol.* 67, 207–233. doi: 10.1146/annurev-arplant-043015-111929
- Greene, E. A., Codomo, C. A., Taylor, N. E., Henikoff, J. G., Till, B. J., Reynolds, S. H., et al. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. *Genetics* 164, 731–740. Available online at: <http://www.genetics.org/content/164/2/731>
- Hartwig, B., James, G. V., Konrad, K., Schneeberger, K., and Turck, F. (2012). Fast isogenic mapping-by-sequencing of ethyl methanesulfonate-induced mutant bulks. *Plant Physiol.* 160, 591–600 doi: 10.1104/pp.112.2.00311
- Henry, I. M., Nagalakshmi, U., Lieberman, M. C., Ngo, K. J., Krasileva, K. V., Vasquez-Gross, H., et al. (2014). Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. *Plant Cell* 26, 1382–1397. doi: 10.1105/tpc.113.1.21590
- Huo, H., Henry, I. M., Coppoolse, E. R., Verhoef-Post, M., Schut, J. W., de Rooij, H., et al. (2016). Rapid identification of lettuce seed germination mutants by bulked segregant analysis and whole genome sequencing. *Plant J.* 88, 345–360. doi: 10.1111/tpj.13267
- Jander, G., Norris, S. R., Rounsley, S. D., Bush, D. F., Levin, I. M., and Last, R. L. (2002). Arabidopsis map-based cloning in the post-genome era. *Plant Physiol.* 129, 440–450 doi: 10.1104/pp.003533
- Jenks, M. A., Rich, P. J., Peters, P. J., Axtell, J. D., and Ashworth, E. N. (1992). Epicuticular wax morphology of bloomless (bm) mutants in *Sorghum bicolor*. *Int. J. Plant Sci.* 153, 311–319. doi: 10.1086/297034
- Jiao, Y., Burke, J. J., Chopra, R., Burrow, G., Chen, J., Wang, B., et al. (2016). A sorghum mutant resource as an efficient platform for gene discovery in grasses. *Plant Cell* 28, 1551–1562. doi: 10.1105/tpc.16.00373
- Jordan, W. R., Shouse, P. J., Blum, A., Miller, F. R., and Monk, R. L. (1984). Environmental physiology of sorghum. II. Epicuticular wax load and cuticular transpiration. *Crop Sci.* 24, 1168–1173. doi: 10.2135/cropsci1984.0011183X002400060038x
- Krasileva, K. V., Vasquez-Gross, H. A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., et al. (2017). Uncovering hidden variation in polyploid wheat. *Proc. Natl. Acad. Sci. U.S.A.* 114, E913–E921. doi: 10.1073/pnas.16192.68114
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Kunst, L., and Samuels, A. L. (2003). Biosynthesis and secretion of plant cuticular wax. *Prog. Lipid Res.* 42, 51–80. doi: 10.1016/S0163-7827(02)00045-0
- Lai, C. P., Huang, L. M., Chen, L. F. O., Chan, M. T., and Shaw, J. F. (2017). Genome-wide analysis of GDSL-type esterases/lipases in Arabidopsis. *Plant Mol. Biol.* 95, 181–197. doi: 10.1007/s11103-017-0648-y
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Ling, H. (2008). Sequence analysis of GDSL lipase gene family in *Arabidopsis thaliana*. *Pak. J. Biol. Sci.* 11, 763–737. doi: 10.3923/pjbs.2008.763.767
- Makita, Y., Shimada, S., Kawashima, M., Kondou-Kuriyama, T., Toyoda, T., and Matsui, M. (2015). MOROKOSHI: transcriptome database in *Sorghum bicolor*. *Plant Cell Physiol.* 56:e6. doi: 10.1093/pcp/pcu187
- Merchant, N., Lyons, E., Goff, S., Vaughn, M., Ware, D., Micklos, D., et al. (2016). The iPlant collaborative: cyberinfrastructure for enabling data to discovery for the life sciences. *PLoS Biol.* 14:e1002342. doi: 10.1371/journal.pbio.1002342
- Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat. Rev. Genet.* 11, 31–46. doi: 10.1038/nrg2626
- Michelmore, R. W., Paran, I., and Kesseli, R. V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. U.S.A.* 88, 9828–9832 doi: 10.1073/pnas.88.2.9828
- Mullet, J., Morishige, D., McCormick, R., Truong, S., Hilley, J., McKinley, B., et al. (2014). Energy Sorghum—a genetic model for the design of C4 grass bioenergy crops. *J. Exp. Bot.* 65, 3479–3489. doi: 10.1093/jxb/eru229
- Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457, 551–556. doi: 10.1038/nature07723
- Peters, P. J., Jenks, M. A., Rich, P. J., Axtell, J. D., and Ejeta, G. (2009). Mutagenesis, selection, and allelic analysis of epicuticular wax mutants in sorghum. *Crop Sci.* 49, 1250–1258. doi: 10.2135/cropsci2008.08.0461
- Premachandra, G. S., Hahn, D. T., Axtell, J. D., and Joly, R. J. (1994). Epicuticular wax load and water-use efficiency in bloomless and sparse-bloom mutants of *Sorghum bicolor* L. *Environ. Exp. Bot.* 34, 293–301. doi: 10.1016/0098-8472(94)90050-7
- Rooney, W. L., Blumenthal, J., Bean, B., and Mullet, J. E. (2007). Designing sorghum as a dedicated bioenergy feedstock. *Biofuels Bioprod. Bioref.* 1, 147–157. doi: 10.1002/bbb.15
- Samdur, M. Y., Manivel, P., Jain, V. K., Chikani, B. M., Gor, H. K., Desai, S., et al. (2003). Genotypic differences and water-deficit induced enhancement in epicuticular wax load in peanut. *Crop Sci.* 43, 1294–1299. doi: 10.2135/cropsci2003.1294
- Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A. H., Nielsen, K. L., et al. (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat. Methods* 6, 550–551. doi: 10.1038/nmeth0809-550
- Schneeberger, K., and Weigel, D. (2011). Fast-forward genetics enabled by new sequencing technologies. *Trends Plant Sci.* 16, 282–288. doi: 10.1016/j.tplants.2011.02.006
- Takagi, H., Tamiru, M., Abe, A., Yoshida, K., Uemura, A., Yaegashi, H., et al. (2015). MutMap accelerates breeding of a salt-tolerant rice cultivar. *Nat. Biotechnol.* 33, 445–449. doi: 10.1038/nbt.3188
- Tello-Ruiz, M. K., Stein, J., Wei, S., Preece, J., Olson, A., Naithani, S., et al. (2016). Gramem 2016: comparative plant genomics and pathway resources. *Nucleic Acids Res.* 44, D1133–D1140. doi: 10.1093/nar/gkv1179
- Uttam, G. A., Praveen, M., Rao, Y. V., Tonapi, V. A., and Madhusudhana, R. (2017). Molecular mapping and candidate gene analysis of a new epicuticular wax locus in sorghum (*Sorghum bicolor* L. Moench). *Theor. Appl. Genet.* 130, 2109–2125. doi: 10.1007/s00122-017-2945-x
- Vermerris, W. (2011). Survey of genomics approaches to improve bioenergy traits in maize, sorghum and sugarcane. *J. Integr. Plant Biol.* 53, 105–119. doi: 10.1111/j.1744-7909.2010.01020.x
- Weibel, D. E. (1986a). Registration of 12 bloomless and four sparse-bloom lines of sorghum. *Crop Sci.* 26, 840–841.
- Weibel, D. E. (1986b). Registration of eight bloomless parental lines of sorghum. *Crop Sci.* 26:842
- Xin, Z., and Chen, J. (2012). A high throughput DNA extraction method with high yield and quality. *Plant Methods* 8:26. doi: 10.1186/1746-4811-8-26
- Xin, Z., Huang, J., Smith, A. R., Chen, J., Burke, J., Sattler, S. E., et al. (2017). Morphological characterization of a new and easily recognizable nuclear male sterile mutant of sorghum (*Sorghum bicolor*). *PLoS ONE* 12:e0165195. doi: 10.1371/journal.pone.0165195
- Xin, Z., Wang, M. L., Barkley, N. A., Burrow, G., Franks, C., Pederson, G., et al. (2008). Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. *BMC Plant Biol.* 8:103. doi: 10.1186/1471-2229-8-103
- Yachdav, G., Wilzbach, S., Rauscher, B., Sheridan, R., Sillitoe, I., Procter, J., et al. (2016). MSASviewer: interactive JavaScript visualization of multiple sequence alignments. *Bioinformatics* 32, 3501–3503. doi: 10.1093/bioinformatics/btw474

- Yeats, T. H., Buda, G. J., Wang, Z., Chehanovsky, N., Moyle, L. C., Jetter, R., et al. (2012a). The fruit cuticles of wild tomato species exhibit architectural and chemical diversity, providing a new model for studying the evolution of cuticle function. *Plant J.* 69, 655–666. doi: 10.1111/j.1365-3113X.2011.04820.x
- Yeats, T. H., Huang, W., Chatterjee, S., Viart, H. M., Clausen, M. H., Stark, R. E., et al. (2014). Tomato cutin deficient 1 (CD1) and putative orthologs comprise an ancient family of cutin synthase-like (CUS) proteins that are conserved among land plants. *Plant J.* 77, 667–675. doi: 10.1111/tpj.12422
- Yeats, T. H., Martin, L. B., Viart, H. M., Isaacson, T., He, Y., Zhao, L., et al. (2012b). The identification of cutin synthase: formation of the plant polyester cutin. *Nat. Chem. Biol.* 8, 609–611. doi: 10.1038/nchembio.960
- Zhu, Y., Mang, H. G., Sun, Q., Qian, J., Hipps, A., and Hua, J. (2012). Gene discovery using mutagen-induced polymorphisms and deep sequencing: application to plant disease resistance. *Genetics* 192, 139–146. doi: 10.1534/genetics.112.141986
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2018 Jiao, Burow, Gladman, Acosta-Martinez, Chen, Burke, Ware and Xin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.