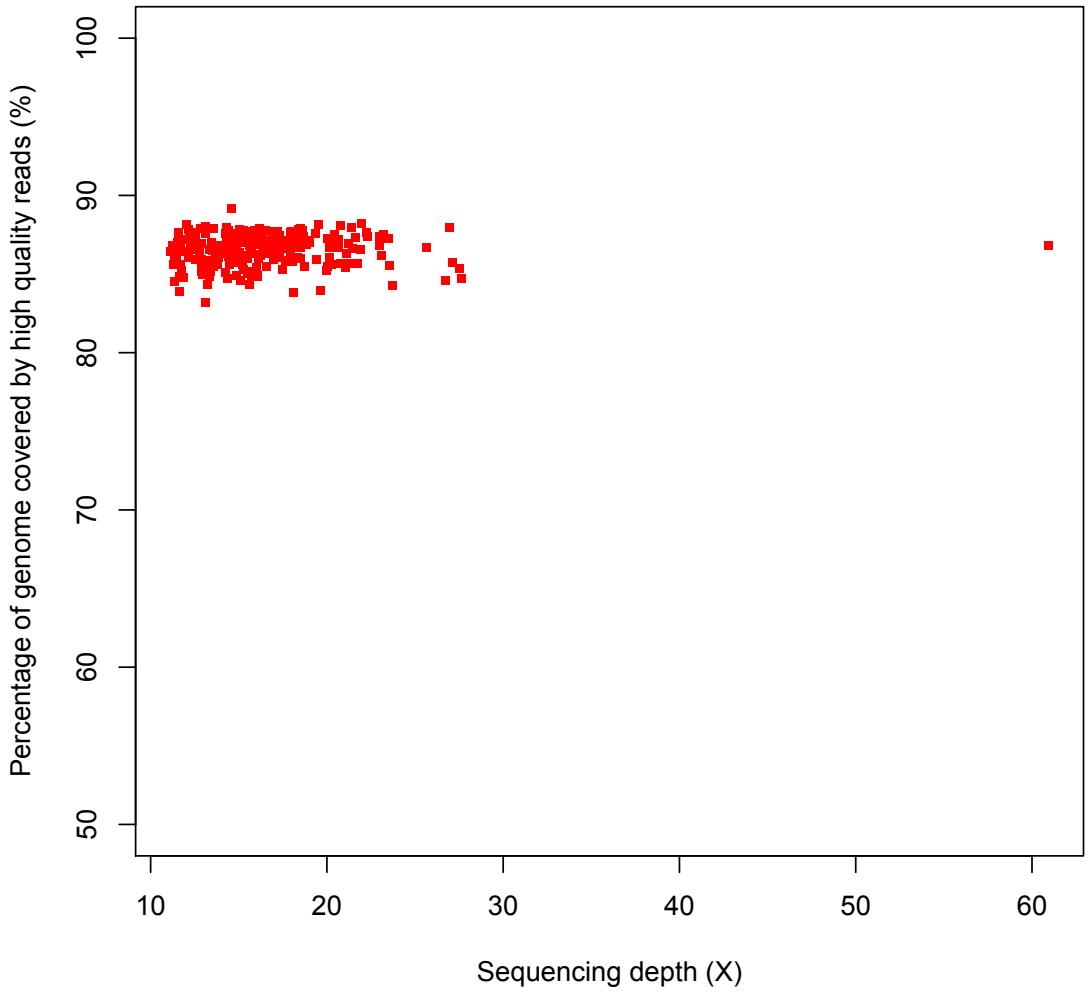
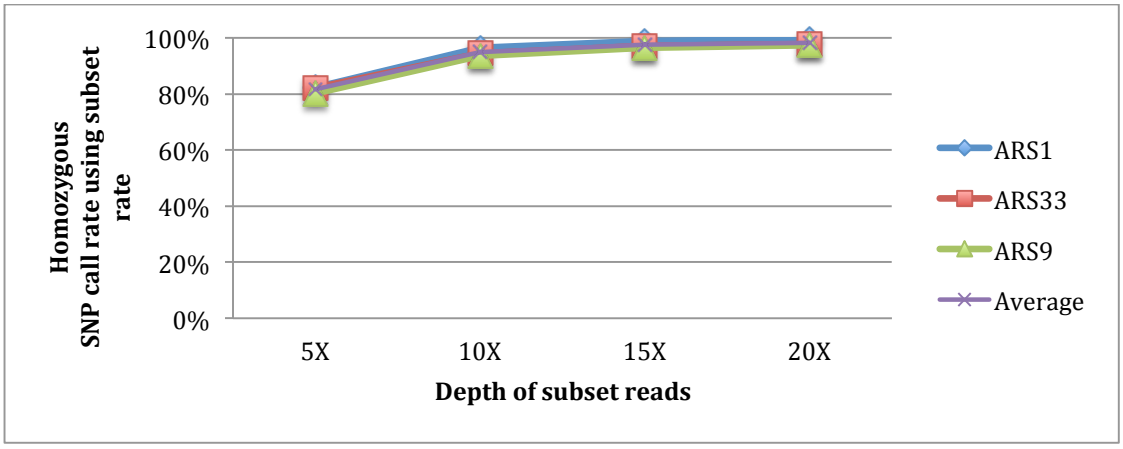
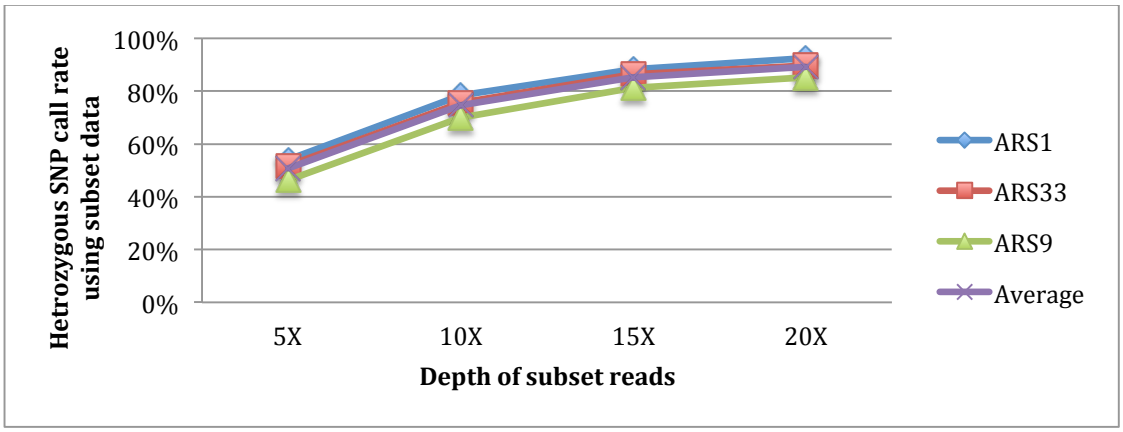


Supplemental Figure 1. Average sequencing depth and genome coverage in 256 samples. The whole genome sequencing depth was in a range from 11X to 60X coverage of Sorghum genome. The genome coverage (averagely 86.6%) did not improve with increased sequencing depth.

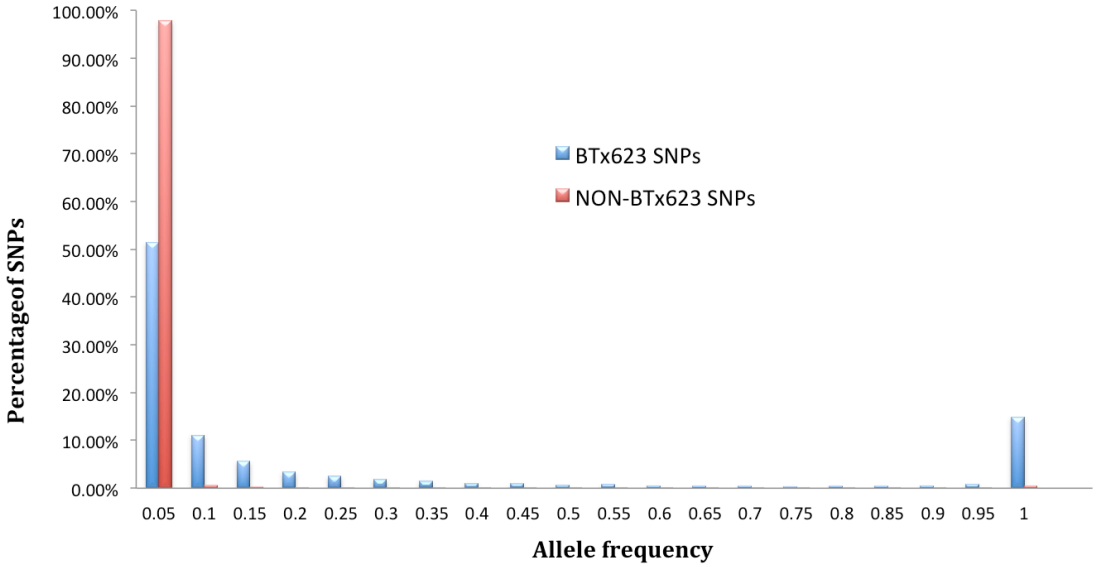
A.



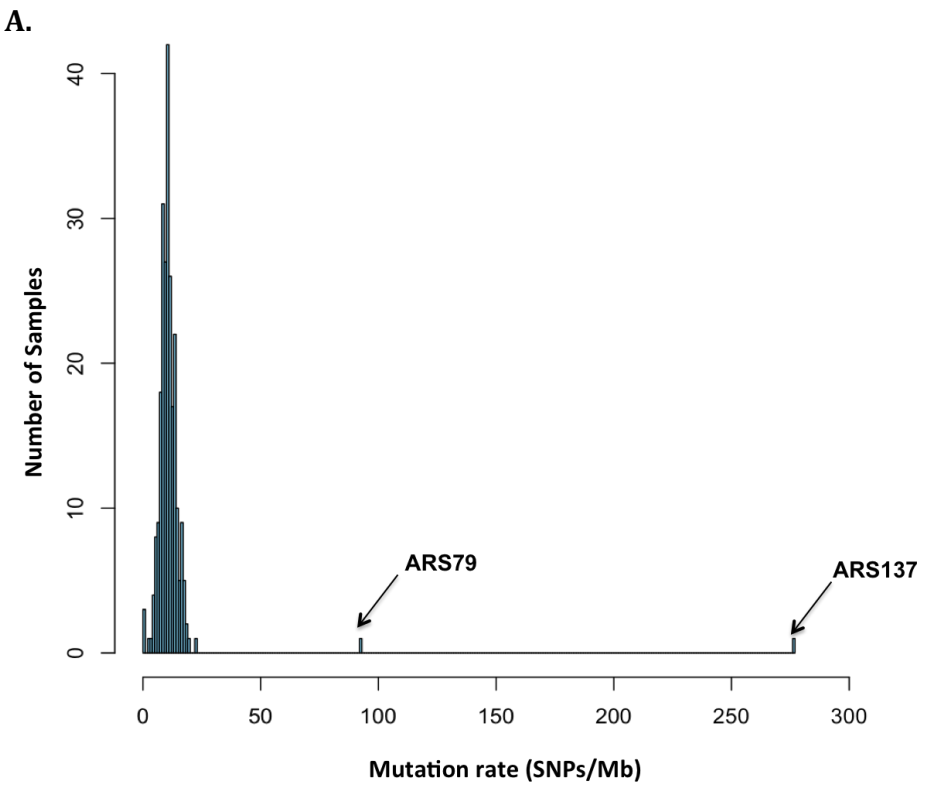
B.

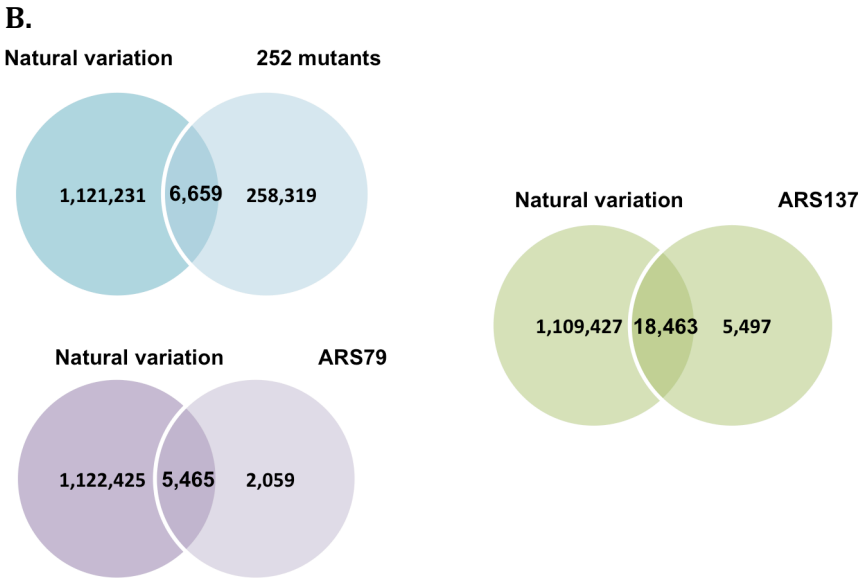


Supplemental Figure 2. SNP discovery rate at various sequencing depths. ARS1, ARS9, and ARS33 had sequencing coverage > 27x; the average coverage for all 256 lines was 16.4X. These three lines were used to determine how many SNPs could be recovered at the average sequencing depth, assuming that the SNP recovery rate at 27x was 100%. Below, the SNP discovery rate is plotted against sequencing depth after the order of short sequences was randomized. Over 98% of homozygous SNPs and over 85% of heterozygous SNPs discovered at 27x coverage were identified at 15x coverage. The standard deviation is too small to be shown on the graph. **A)** Homozygous SNP capture rate. **B)** Heterozygous SNP capture rate.



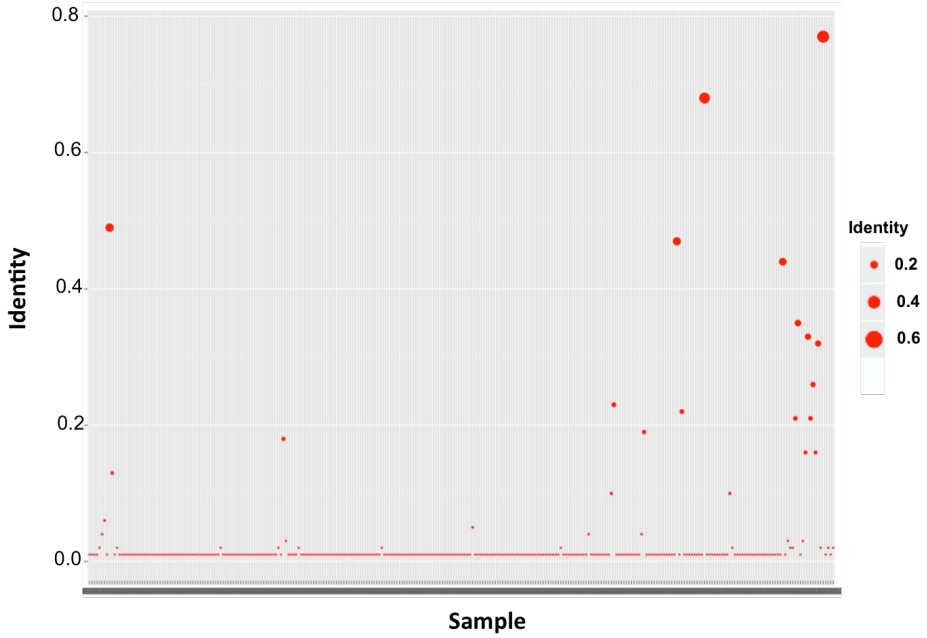
Supplemental Figure 3. Allele frequency of BTx623 SNPs and the NON-BTx623 SNPs in the 256 samples. The background SNPs identified by comparing with the whole genome sequencing of the parental line BTx623 had higher allele frequency than the rest.



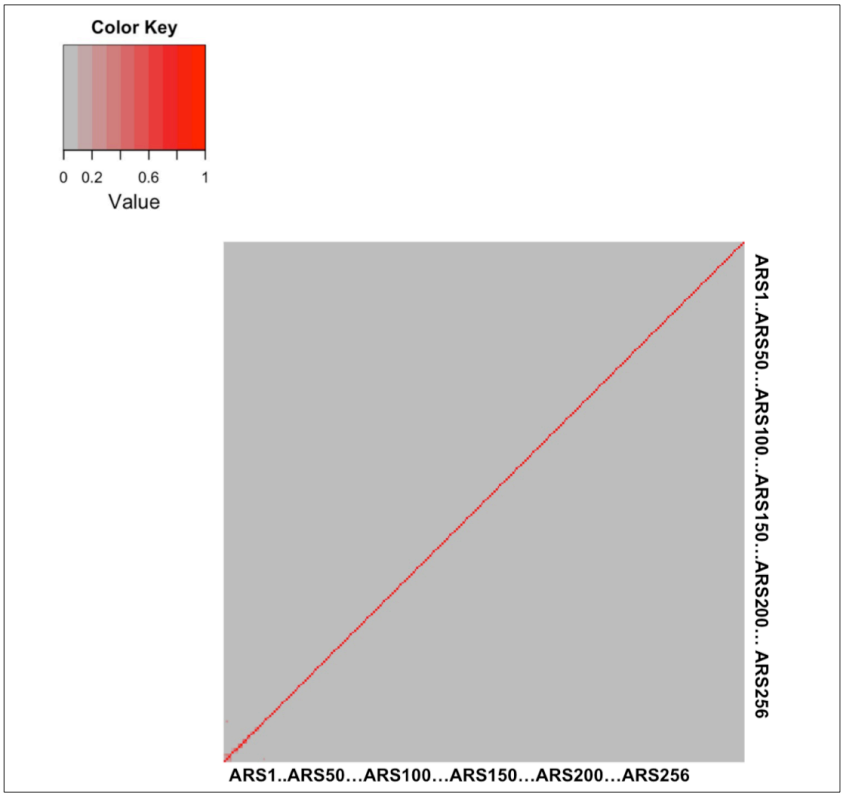


Supplemental Figure 4. Detection of two contamination lines. A) Extremely high mutation rate of the two lines. **B)** Overlap between natural variations and EMS mutations.

A.

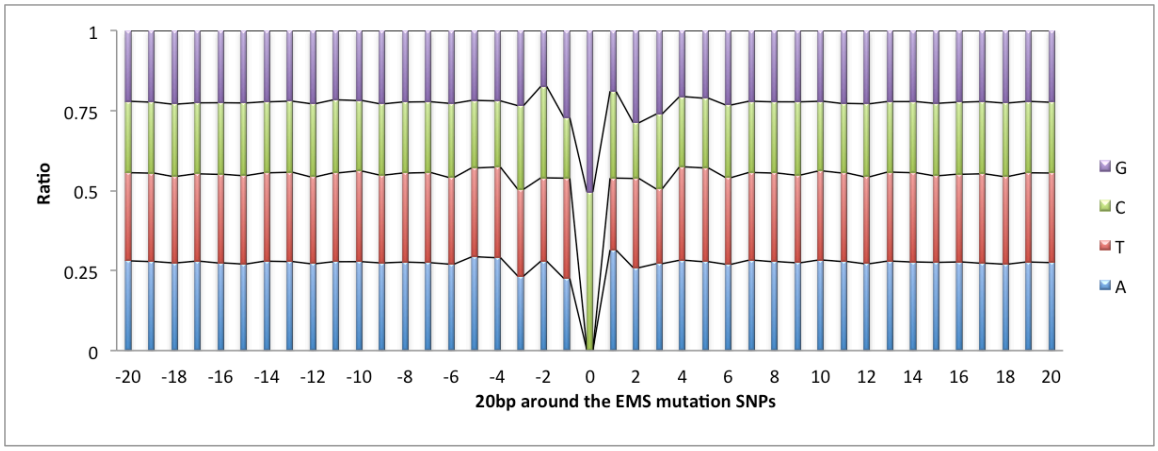


B.

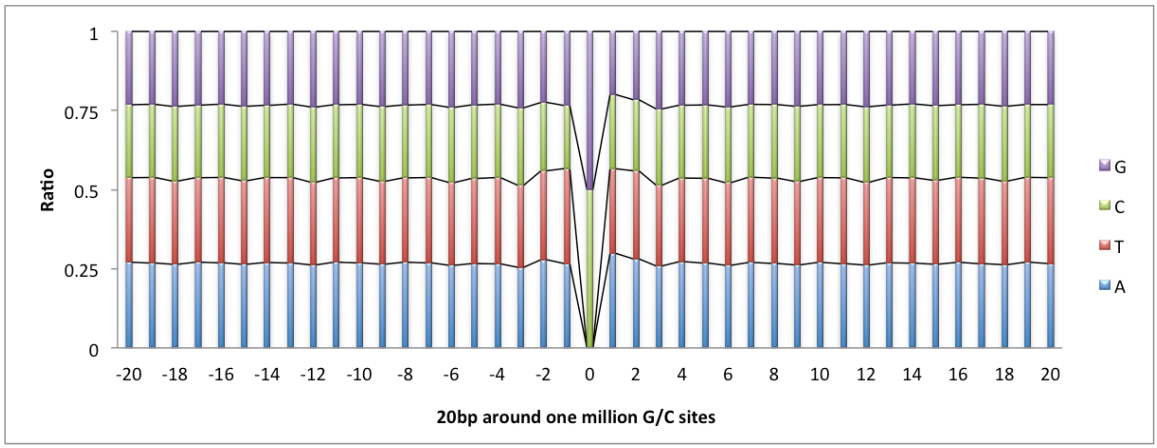


Supplemental Figure 5. Sibling line detection in the population. A) Similarity of 256 samples before filtering. **B)** Heatmap of the similarity of 252 samples after removing the contaminating and sibling lines.

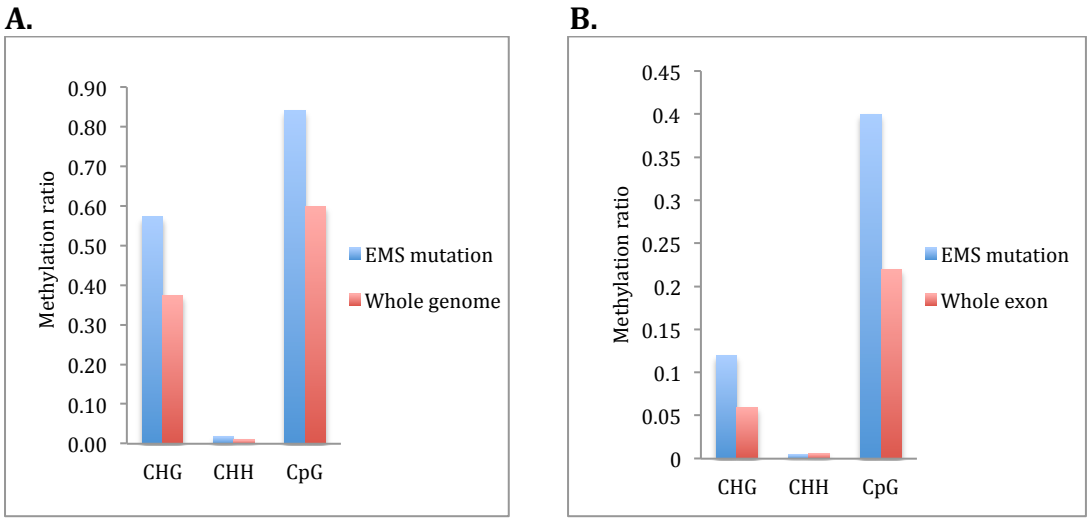
A.



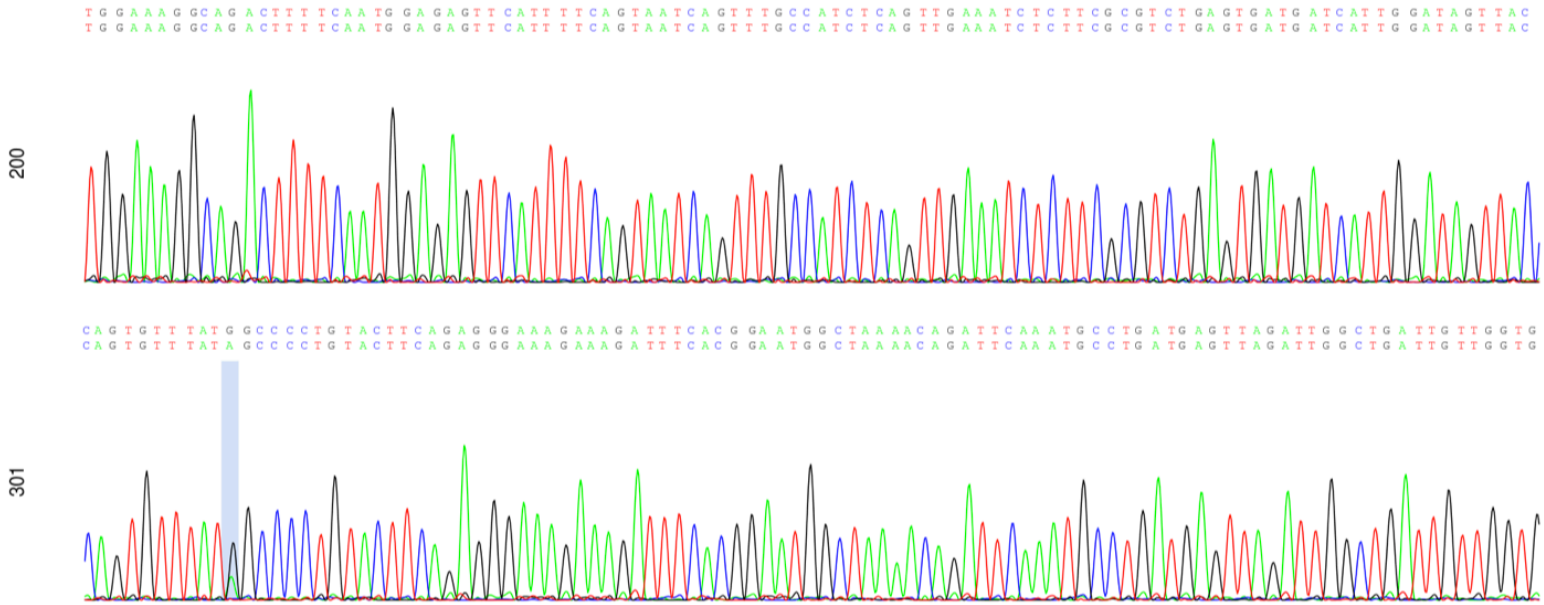
B.



Supplemental Figure 6. Distribution of the flanking nucleotides. **A)** Local nucleotide frequency around EMS mutation sites. **B)** Local nucleotide frequency around one million random selected G/C sites in the sorghum genome.



Supplemental Figure 7. Methylation profiling of EMS mutation. A) Methylation level at EMS mutation sites and in the whole genome. **B)** Methylation level at EMS mutation sites and in the exome.



Supplemental Figure 8. Validation for heterozygous mutations. Heterozygous mutations were determined by the relative signal strength of the two overlapping bases. The shaded area has two overlapping signals. When the weaker signal was over 30% of the stronger signal, this site would be scored as heterozygous.

Supplemental Table 1. SNP validation rates at various sequencing depths.

Sequencing depth	Homozygous SNPs			Heterozygous SNPs		
	Successful primer pairs	Validated SNPs	Validation rate	Successful primer pairs	Validated SNPs	Validation rate
<5	248	246	99.19%	1	1	100.00%
5 - 10	196	196	100.00%	37	35	94.59%
10 - 15	225	225	100.00%	86	82	95.35%
15 -20	74	74	100.00%	67	65	97.01%
>20	26	26	100.00%	32	29	90.63%

Supplemental Table 2. Mutations associated with seed-size QTLs.

Three sequenced lines harbored mutations in genes within seed-size QTL regions identified in genome-wide association studies. All lines exhibited the predicted changes in seed size.

Gene	Mutant	SNP location	Amino acid change	Rep1 (Core)				Rep2			
				% kernel ≥size 6	% kernel ≤ size 7	% kernel small size	Avg. TKW	% kernel ≥size 6	% kernel ≤size 7	% kernel small size	Avg. TKW
Sobic.006G268800	ARS110	6:61092626	L252F	70.4	26.5	3.0	34.1	52.6	45.5	1.9	26.0
Sobic.006G268800	ARS118	6:61093178	A434V	-	-	-	-	48.3	46.7	5.0	26.4
Sobic.010G144900	ARS37	10:40113456	Q172*	24.1	30.3	45.6	14.7	31.4	59.9	8.7	23.1
Wild-type				23.4	63.0	13.6	26.1	39.6	51.8	8.6	29.0

Supplemental Table 3. Primer sequences used for genotyping the F2 population of ARS20.

Gene ID	Gene	Common	Allele X	Allele Y
Sobic.001G453200	SbCER6	CACTACATCCCGCCCAAC	TGACGAGCTGCGCCTC	TGACGAGCTGCGCCTT
Sobic.009G083300	SbCER5	GCGTCACCGAAGTAGACACA	ACTCTTCGATGACCTTTGCC	CACTCTTCGATGACCTTTGCT
Sobic.004G086300	Negative Control	CATATCACCGCGTGTCAAAA	CTAAGGAAGTGGGGCTGG	GCTAAGGAAGTGGGGCTGA
Sobic.006G127800	Negative Control	ATCTCCTCCTGCTGTTGCTG	AGGTGGCGAGCCAGGC	GGTGGCGAGCCAGGT