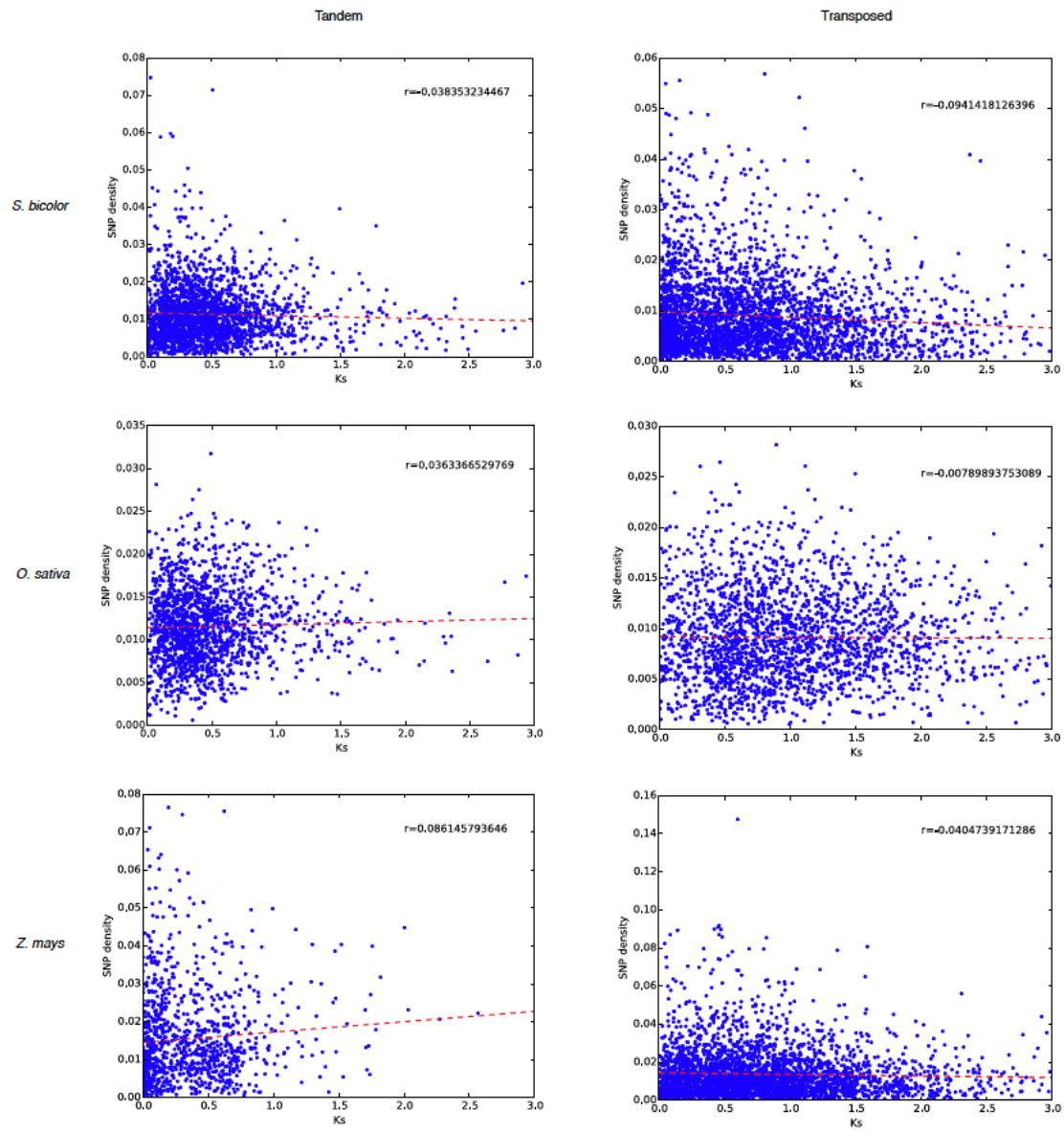
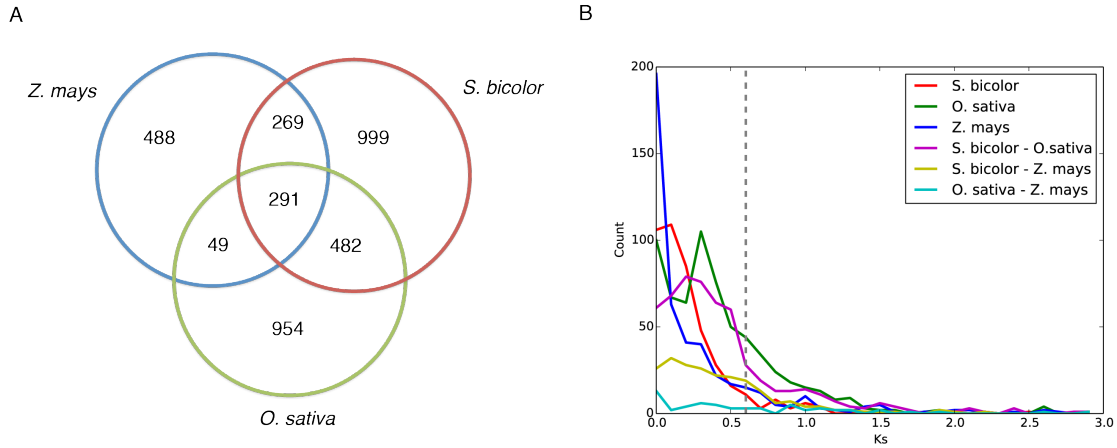


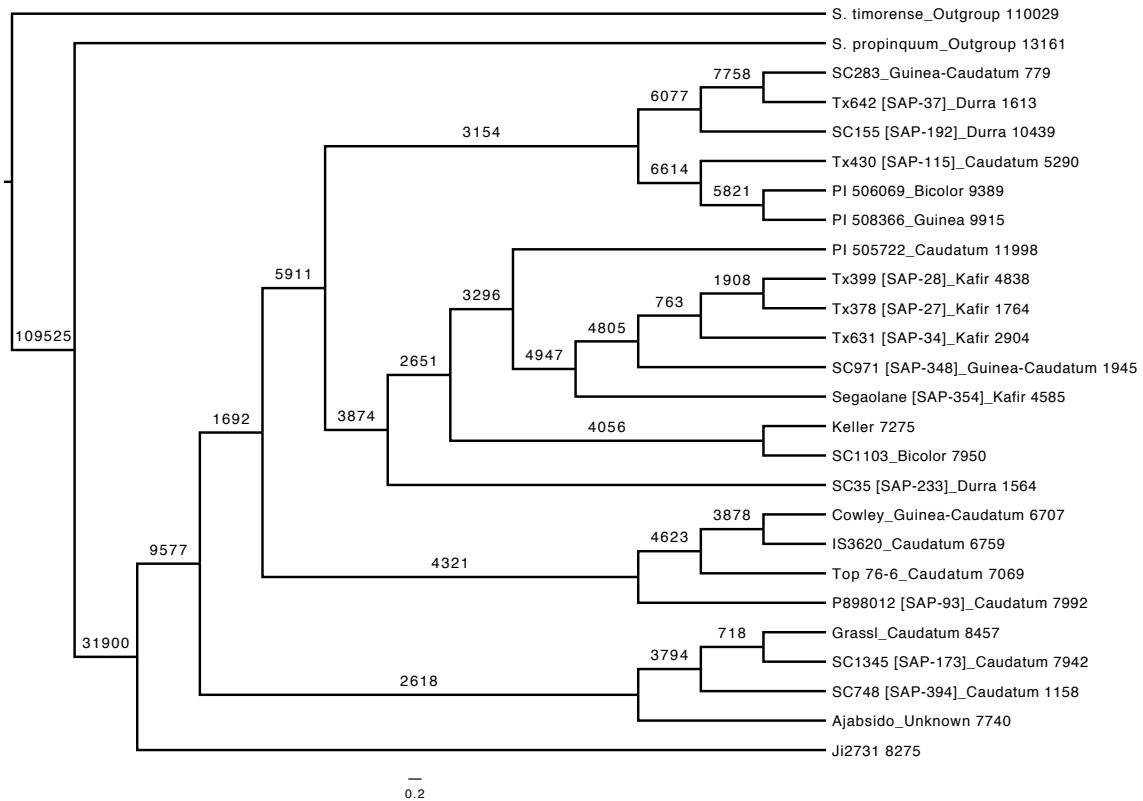
Supplemental materials



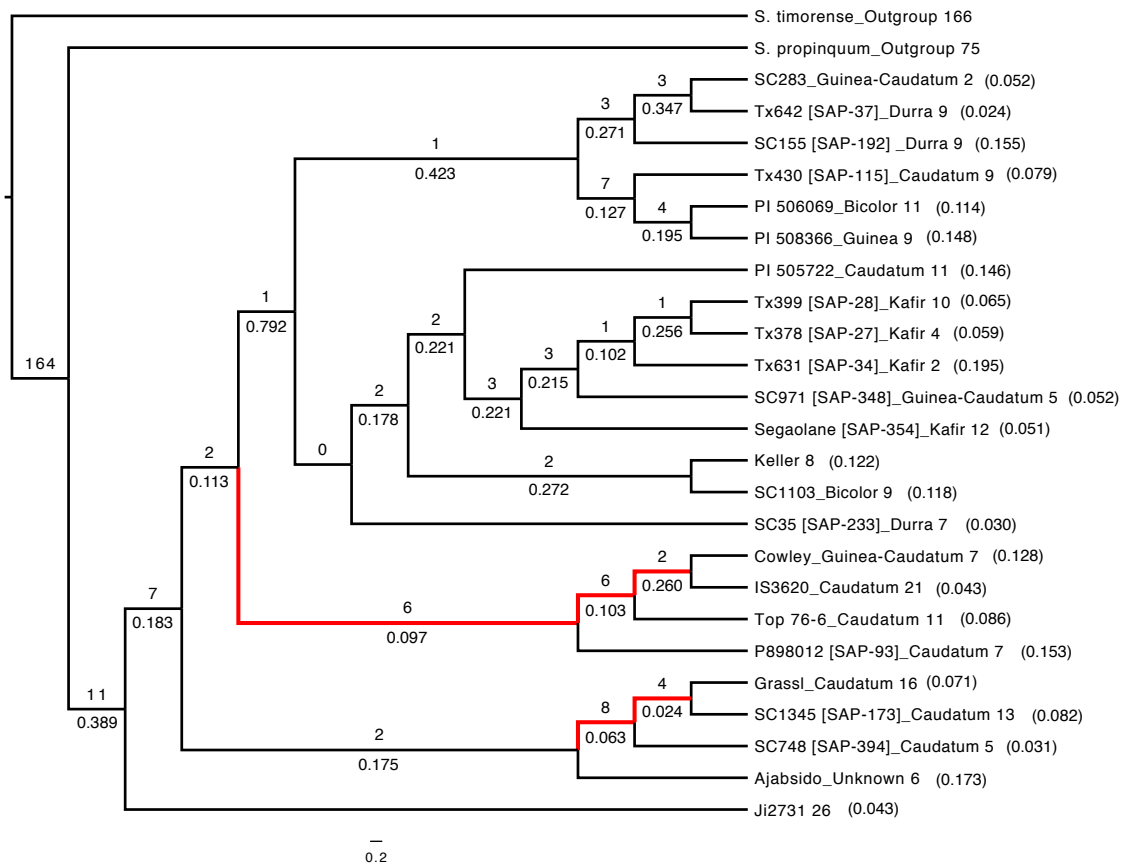
Supplemental Figure S1. Pearson correlation between the age of duplication indicated by Ks values and SNP densities. No strong correlation is observed between the nucleotide divergence of the duplicated genes and the time of origin.



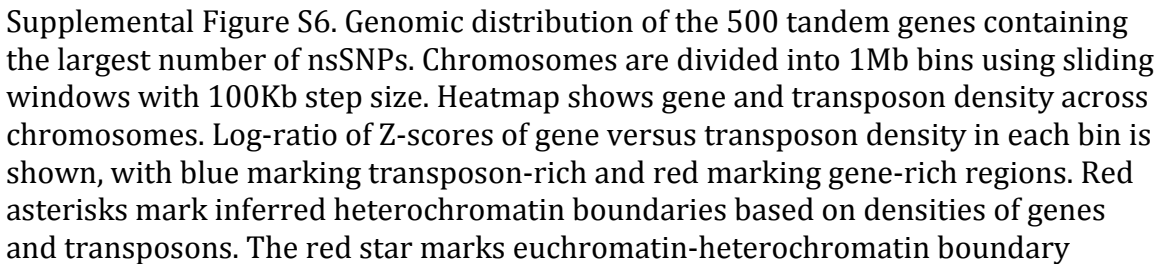
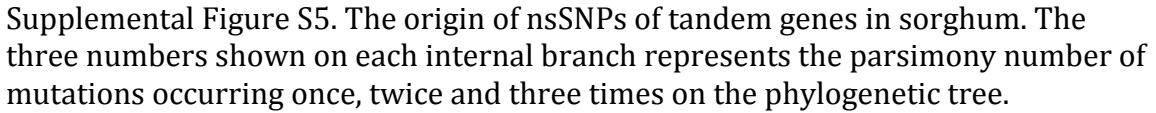
Supplemental Figure S2. Tandem genes. (A) Number of gene families retaining tandem duplication in sorghum, maize and rice. (B) Distribution of tandem gene pairs with different synonymous substitution rate (Ks) within and between species.



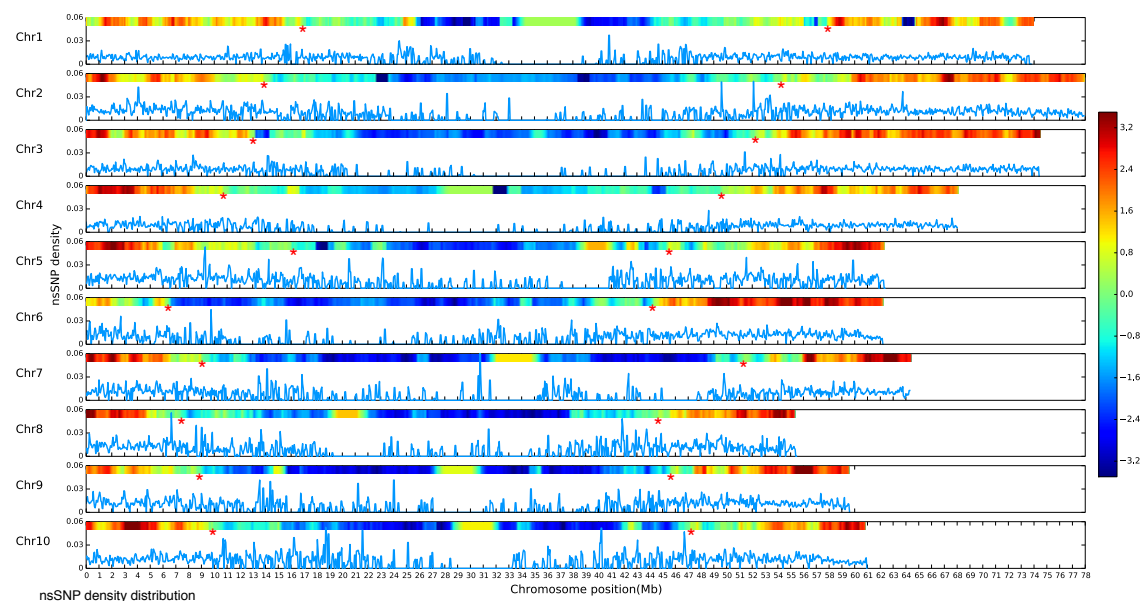
Supplemental Figure S3. The origin of synonymous nucleotide variations in sorghum population. The number shown on each internal branch represents the parsimony number of synonymous mutations occurred on the branch.



Supplemental Figure S4. The origin of tandem duplications in sorghum. A maximum-likelihood phylogenetic tree was constructed based on alignment of single copy genes from all genotypes, using PhyML v3.0 software (Guindon, Dufayard et al. 2010). Above each internal branch is the number of tandem duplications occurring on that branch of the phylogenetic tree, inferred based on maximum parsimony. Under each internal branch is the average time (in million years) required for tandem duplication to occur once on that branch. The numbers and rates on each leaf follow the accession ID. The internal branches marked in red indicate the top five with highest tandem duplication rate among all the internal branches and are clustered to the ancestors of caudatum clades.



defined by the first encounter of log-ratio of Z-score is less than zero for 15 continuous bins (1.5Mb).



Supplemental Figure S7. Genomic distribution of the nsSNP density across the sorghum genome. Chromosomes are divided into 1Mb bins using sliding windows with 100Kb step size. Heatmap shows gene and transposon density across chromosomes. Log-ratio of Z-scores of gene versus transposon density in each bin is shown, with blue marking transposon-rich and red marking gene-rich regions. Red asterisks mark inferred heterochromatin boundaries based on densities of genes and transposons. The red star marks euchromatin-heterochromatin boundary defined by the first encounter of log-ratio of Z-score is less than zero for 15 continuous bins (1.5Mb).

Sorghum	High nsSNP	Tandem	>	Transposed	>	Syntenic
	Whole genome	Tandem	>	Transposed	>	Syntenic
Maize	High nsSNP	Transposed	>	Syntenic	>	Tandem
	Whole genome	Transposed	>	Tandem	>	Syntenic

Supplemental Figure S8. Ranking of gene duplication types by the proportion of non-neutral evolutionary genes. “Whole genome” indicates that all the genes in each

duplication type are considered. “high nsSNP” indicates a subgroup of genes with top 5% nsSNP densities are considered. “*” Marginally significant ($0.1 < p < 0.05$); “**” Significant ($0.01 < p < 0.05$); “***” Highly significant ($p < 0.01$).



Supplemental Figure S9. Protein structure of *Sb10G13800* predicted by I-TASSER web service.

Supplemental Table S1. Number of duplicated genes under different selective pressure.

		Transpose	Specific	Syteny	Tandem
Sorghum	High nsSNP*	16*/28*	21/17	3/3	20/7
	Whole genome*	43/108	22/90	47/178	43/124
Maize	High nsSNP	4/6	25/94	1/10	0/4
	Whole genome	378/110	452/428	66/459	21/105

*Note: "Whole genome" indicates all the genes in each duplication type are considered. "High nsSNP" indicates a subgroup of genes with top 5% highest nsSNP density are considered. The numbers in each cell indicates the number of genes showing balancing/negative selection indicated by Tajima's D test.

Supplemental Table S2. Primers information for Sb010G138100 expression analyses.

No.	Primer name	Froward (5'→3')	Reverse(5'→3')	Product size (bp)
1	RT-1	GAGAGCGGCCTCATGTTAAG	GAGCCAGAAATTGCTTTCCA	293
2	RT-2	TAATCCAACCTTGCGCTTTC	TCCACTGACATCGCTTCATC	207
3	RT-3	ACTAGCCCAAGTGGCAAGTC	ATGTTGCTTCTTGCCCAATC	269
4	SbActin	ACATTGCCCTGGACTACGAC	AATGAAGGATGGCTGGAAGA	149

Supplemental Table S3. Sorghum sequencing information.

Bio Classification Name	Type	Library	NCBI accession
Sorghum bicolor SC301	RNA-seq	BSUC	SRP024578
Sorghum bicolor B 35	RNA-seq	BSWY	SRP024579
Sorghum bicolor Grassl	RNA-seq	BSWZ	SRP024580
Sorghum bicolor IS25702	RNA-seq	BUUB	SRP024581
Sorghum bicolor SC52	RNA-seq	BUUC	SRP024582
Sorghum bicolor SC110	RNA-seq	BXYU	SRP024583
		BXY	
Sorghum bicolor SC650	RNA-seq	W	SRP024584
Sorghum bicolor SC566	RNA-seq	BXYX	SRP024585
Sorghum bicolor SC55	RNA-seq	BXYX	SRP024586
Sorghum bicolor BTx631	RNA-seq	CIUA	SRP015654
	RNA-seq		SRP105466
Sorghum bicolor PI567946	RNA-seq	CIUC	SRP116944

Sorghum bicolor PI610727	RNA-seq	CIUF	SRP116976
Sorghum bicolor Liang Tang Ai	RNA-seq	CNGO	SRP015362
Sorghum bicolor bicolor SSM29	RNA-seq	CNXF	SRP116939
Sorghum bicolor bicolor Tiandougou	RNA-seq	CNXG	SRP116938
Sorghum bicolor verticilliflorum virgatum	RNA-seq	CNXH	SRP116940
Sorghum bicolor B963676	RNA-seq	CNXN	SRP116931
Sorghum bicolor bicolor IS9830	RNA-seq	CNXO	SRP116994
Sorghum bicolor bicolor N13	RNA-seq	CNXP	SRP116993
Sorghum bicolor drummondii IS21691	RNA-seq	CNXS	SRP116974
Sorghum bicolor verticilliflorum IS14216	RNA-seq	CNXT	SRP116973
Sorghum bicolor bicolor S35	RNA-seq	CNXU	SRP117002
Sorghum bicolor verticilliflorum IS18868	RNA-seq	CNX W	SRP116978
Sorghum bicolor bicolor CSM63E	RNA-seq	CNZG	SRP117001
Sorghum bicolor bicolor E36-1	RNA-seq	CNZH	SRP116991
Sorghum bicolor bicolor IS1054(M35-1)	RNA-seq	CNZI	SRP116995
Sorghum bicolor bicolor IS15401	RNA-seq	CNZN	SRP116992
Sorghum bicolor bicolor IS2205	RNA-seq	CNZO	SRP116981
Sorghum bicolor bicolor IS22632	RNA-seq	CNZP	SRP116972

Sorghum bicolor bicolor Guinea roxburguii IS14317	RNA-seq	CNZS			SRP116941
Sorghum bicolor bicolor IS24887	RNA-seq	CNZT			SRP116943
Sorghum bicolor bicolor ratoon	RNA-seq	CNZU			SRP116942
Sorghum bicolor verticilliflorum IS27587	RNA-seq	CNZW			SRP116975
Sorghum bicolor SAP156	RNA-seq	COCW			SRP117008
Sorghum bicolor SAP165	RNA-seq	COCX			SRP117009
Sorghum bicolor SAP183	RNA-seq	COCY			SRP117010
Sorghum bicolor SAP207	RNA-seq	COCZ			SRP117006
Sorghum bicolor SAP317	RNA-seq	COFA			SRP117013
Sorghum bicolor SAP336	RNA-seq	COFB			SRP117015
Sorghum bicolor SAP348	RNA-seq	COFC			SRP117016
	RNA-seq				SRP116994
Sorghum bicolor Ajabsido	Resequencing	IACF	SRR562213	SRS361620	SRP015636
Sorghum bicolor Cowley	Resequencing	ITZS	SRR4028805	SRS1615372	SRP081858
Sorghum bicolor Grassl	Resequencing	INIB	SRR486615	SRS309422	SRP012224
Sorghum bicolor IS3620c	Resequencing	IACI	SRR562491	SRS361627	SRP015637
Sorghum bicolor Macia	Resequencing	ITZI	SRR4028749	SRS1615348	SRP081837
Sorghum bicolor P898012	Resequencing	ITZH	SRR4028763	SRS1615359	SRP081847
Sorghum bicolor PI 505722	Resequencing	INIH	SRR485885	SRS308944	SRP012181
Sorghum bicolor PI 506069	Resequencing	INIG	SRR4030046	SRS1616273	SRP082069
Sorghum bicolor PI 508366	Resequencing	INIC	SRR485949	SRS309196	SRP012183
Sorghum bicolor RTx430	Resequencing	IBAI	SRR4028129	SRS1615291	SRP081800
Sorghum bicolor SC1103	Resequencing	IACP	SRR562875	SRS361645	SRP015641
Sorghum bicolor SC1345	Resequencing	ITZN	SRR4028801	SRS1615370	SRP081856
Sorghum bicolor SC155	Resequencing	IACT	SRR561755	SRS361608	SRP015631
Sorghum bicolor SC265	Resequencing	IACX	SRR561908	SRS361613	SRP015632
Sorghum bicolor SC283	Resequencing	IAGN	SRR4028746	SRS1615347	SRP081835
Sorghum bicolor SC35	Resequencing	HZZG	SRR562954	SRS361646	SRP015642
Sorghum bicolor SC748	Resequencing	ITZT	SRR4028760	SRS1615357	SRP081845
Sorghum bicolor SC971	Resequencing	IACB	SRR562030	SRS361618	SRP015634
Sorghum bicolor Segalane	Resequencing	IUPX	SRR4028748	SRS1615349	SRP081836
Sorghum bicolor Top76-6	Resequencing	INIA	SRR485950	SRS309197	SRP012184
Sorghum bicolor Tx378	Resequencing	IACU	SRR561245	SRS361463	SRP015516
Sorghum bicolor Tx399	Resequencing	IACC			SRX2020510
Sorghum bicolor Tx631	Resequencing	IACW	SRR562643	SRS361632	SRP015639
Sorghum bicolor Tx642	Resequencing	ITZO	SRR4028151	SRS1615303	SRP081812
Sorghum propinquum	Resequencing	GTAH	SRR072055	SRS120942	SRP004259
Sorghum propinquum	Resequencing	GUXX	SRR072053	SRS120942	SRP004259
Sorghum propinquum	Resequencing	GUXX	SRR072054	SRS120942	SRP004259
Sorghum propinquum	Resequencing		SRR072056	SRS120942	SRP004259
Sorghum propinquum	Resequencing		SRR090571	SRS120942	SRP004259
Sorghum propinquum	Resequencing		SRR090572	SRS120942	SRP004259
Sorghum timorense	Resequencing	IIUB	SRR424217	SRS295746	SRP011156
Sorghum versicolor	Draft	HYYN	SRR427176	SRS297258	SRP011266