

SorghumBase

Building Partnerships and Integrating Genetic Knowledge for
the Sorghum Community

Marcela Karey Tello-Ruiz, PhD

Ware Laboratory, USDA-ARS
Cold Spring Harbor Laboratory

May 10, 2023

Webinar Outline

1. Introduction to SorghumBase
2. Questions & Answers
3. Explore SB with Your Favorite Gene (YFG)
 - Demo example => Take screenshots, URLs
4. Participants share their examples, findings, issues
5. Gather feedback & survey

Workshop Slideshow Access Information

Slides and other info will be placed here: <http://ftp.sorghumbase.org/>

Live demo folder:

<https://drive.google.com/drive/folders/1SmorbRVzSnrK451Q6IB36CRLjo7LIT0E>

Gramene Protocol Book Chapter:

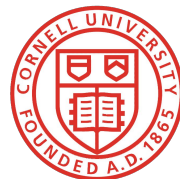
https://link.springer.com/protocol/10.1007/978-1-4939-3167-5_7



YouTube: <https://www.youtube.com/@gramenedatabase3929>



Community Collaborations



SorghumBase Team



<https://sorghumbase.org/>



Doreen Ware
Leadership



Kapeel Chougule
Genome Annotation &
Data Resources



Marcela K. Tello-Ruiz
Communication &
Data Resources



Andrew Olson
Architecture &
Infrastructure



Sharon Wei
Infrastructure &
Data Resources



Zhenyuan Lu
Workflows &
Data Resources



Audra Olson
Communication



Vivek Kumar
Infrastructure &
Data Resources



Nick Gladman
Communication &
Leadership



Sunita Kumari
Communication &
Data Resources



Peter Van Buren
Systems Engineer

What is SorghumBase?

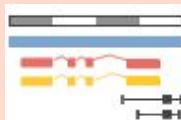
- USDA funded database with regular releases, currently on data release #5
- Pangenome browser with 28 sorghum genomes (10 CP-NAM, 6 SAP, and 11 BAP) & 8 outspecies
- Gene family trees with functional protein domains, DNA/protein alignments (orthologs/paralogs), synteny maps
- Gene expression, genetic variation, limited QTLs & metabolic pathways for BTx623 (anchor)

SB Release #5

Genomes

28 sorghum genomes
10 CP-NAM, 6 SAP, 11 BAP

8 outspecies



1.5M gene annotations



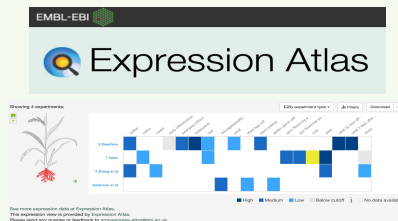
44K gene trees

83 synteny maps

```
ATCGAGCT  
ATCCAGCT  
ATCGAGAT
```

15 whole-genome alignments

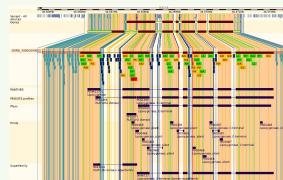
Expression*



235 baseline experiments (8 studies)

Variation*

~61M SNPs
SAP, TERRA-REF, TERRA-MEPP
4.3M EMS



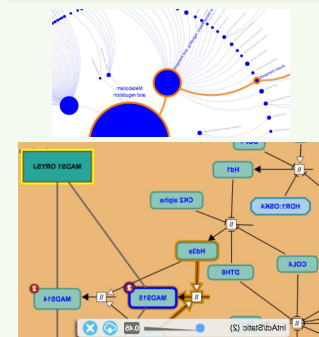
6K QTLs

Sorghum QTL Atlas

Pathways*

269 orthology-based projections (1,248 genes)

Plant REACTOME



* BTx623 only

BTx623 is the anchor genome

- Genetic variation with Variant Effect Predictions (missense, stop-gained)
 - SNPs in population panels (SAP, TERRA-REF, TERRA-MEPP)
 - EMS-induced mutations
- QTLs (Sorghum QTL Atlas)
- Baseline gene expression (EBI-Atlas)
- Orthology-based pathway projections (Gramene Plant Reactome)



Genetic Variation

[Explore the Impact of Variants on Gene Structure](#)



Gene Expression

[Access Curated Gene Expression Data](#)



Pathways

[Plant Reactome Pathways on Community Reference Genomes](#)



News

[Relevant developments in Sorghum research.](#)



Tools

[Gene Search, Genome Browser, BLAST](#)



Events

[Upcoming Meetings and Conferences](#)



Interactive snapshots



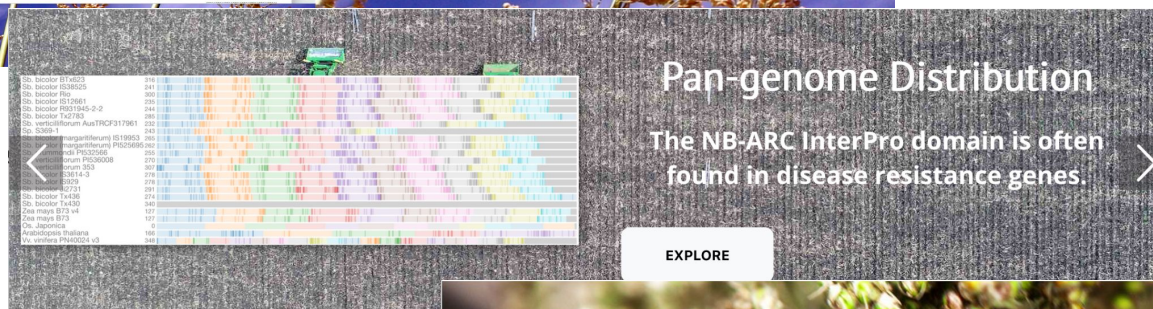
<https://sorghumbase.org/>

Type-ahead Search
Select filters to search or refine a search

Pan-genome Distribution

The NB-ARC InterPro domain is often found in disease resistance genes.

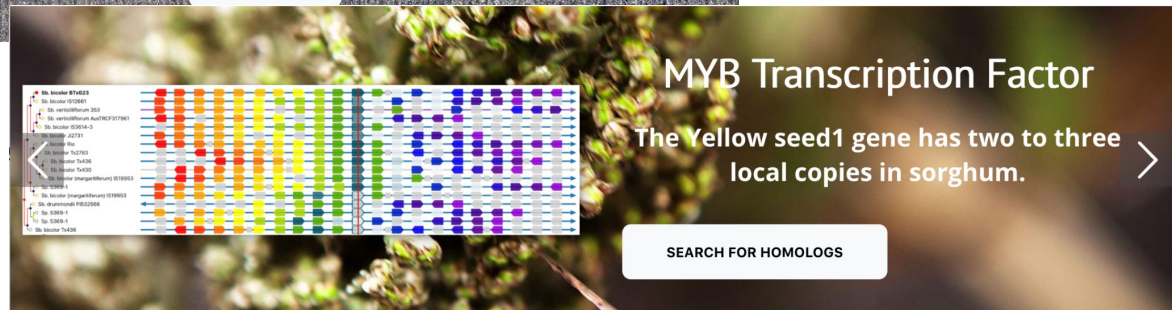
EXPLORE



MYB Transcription Factor

The Yellow seed1 gene has two to three local copies in sorghum.

SEARCH FOR HOMOLOGS



SB Homepage: Events



<https://sorghumbase.org/>



NEWS ENGAGE GENOMES TOOLS RESEARCH ABOUT

Search



Genetic Variation

[Explore the Impact of Variants on Gene Structure](#)



Gene Expression

[Access Curated Gene Expression Data](#)



Pathways

[Plant Reactome Pathways on Community Reference Genomes](#)



News

[Relevant developments in Sorghum research.](#)



Tools

[Gene Search, Genome Browser, BLAST](#)



Events

[Upcoming Meetings and Conferences](#)




SB Homepage: Events



UPCOMING EVENTS



[View past events](#)

- 5 June 2023 

Sorghum in the 21st Century Global Sorghum Conference

5 June 2023 - 9 June 2023

[Event details](#)




- 13 June 2023 

Center for Sorghum Improvement Webinar

John Mullet

13 June 2023



[Event details](#)


- 15 June 2023 

AG2PI conference, Mapping the Future of Agricultural Genome to Phenome Research

15 June 2023 - 16 June 2023

[Event details](#)



- 11 July 2023 

Center for Sorghum Improvement Webinar

SorghumBase

11 July 2023

[Event details](#)



SB Homepage: News



<https://sorghumbase.org/>



NEWS ▾ ENGAGE ▾ GENOMES ▾ TOOLS ▾ RESEARCH ▾ ABOUT ▾

Search 🔍



Genetic Variation

[Explore the Impact of Variants on Gene Structure](#)



Gene Expression

[Access Curated Gene Expression Data](#)



Pathways

[Plant Reactome Pathways on Community Reference Genomes](#)



News

[Relevant developments in Sorghum research.](#)



Tools

[Gene Search, Genome Browser, BLAST](#)



Events

[Upcoming Meetings and Conferences](#)



NEWS & RESEARCH-HIGHLIGHTS

Collaborative Sorghum Investment Program Seeks Proposals In Systems Data And Modeling, And Trait Discovery, Submissions Due June 16

👤 SorghumBase Team / 📅 04 May 2023

The Collaborative Sorghum Investment Program (CSIP) is requesting proposals for two projects they will be funding.

[READ MORE >](#)

Phenotypic And Genetic Diversity In Sweet Sorghum Presents Opportunities For Strengthening Valued Traits Through Cross-Breeding

👤 SorghumBase Team / 📅 01 May 2023

Researchers investigated 75 lines of sweet sorghum for phenotypic and genetic variability and found significant differences, which present opportunities for heterotic cross-breeding.

[READ MORE >](#)

PAG Australia, Sept 20-22, Perth, Poster Abstracts Due Aug. 25

👤 SorghumBase Team / 📅 24 April 2023

The International Plant and Animal Genome Conference will be held in Perth, Australia from September 20-22, 2023.

[READ MORE >](#)

Identification And Promoter Analysis Of 18 Sorghum JAZ Genes With Critical Functions In The JA Signaling Pathway

👤 SorghumBase Team / 📅 24 April 2023

Scientists identified 18 sorghum JAZ genes that have critical functions in the JA signaling pathway and analyzed their promoter regions to identify stress-related cis-acting regulatory elements and transcription factors.

[READ MORE >](#)

Genetic Mapping Behind Low Hydrocyanic Acid In A Sorghum-Sudangrass Hybrid

👤 SorghumBase Team / 📅 16 April 2023

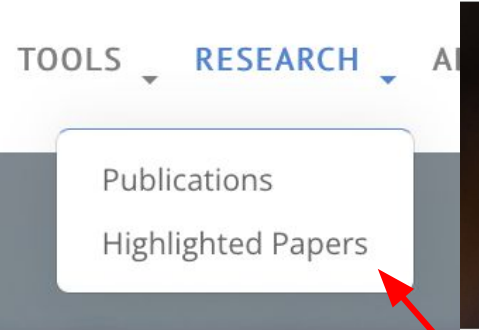
Wu et al. investigated the underlying genetic mechanism behind low HCN content in sorghum and screened 11 simple sequence repeats (SSR) polymorphic primers and, using bulked segregant analysis (BSA), developed four SSR markers associated with low HCN content.

Chromatin, Epigenetics & Gene Expression, CSHL July 26 – Aug. 15

👤 SorghumBase Team / 📅 12 April 2023

Cold Spring Harbor Laboratory is holding the “Chromatin, Epigenetics & Gene Expression” advanced course from July 26 – August 15, 2023.

[READ MORE >](#)



Filter by year:

- 2023
- 2022
- 2021
- 2020
- 2019
- 2018
- 2017
- 2016
- 2015
- 2014
- 2013
- 2012

Structure and genetic regulation of starch formation in sorghum (*Sorghum bicolor* (L.) Moench) endosperm: A review.

Kang X, Gao W, Cui B, Abd El-Aty AM

Published: 7 April 2023 in *International journal of biological macromolecules*

Keywords: [Enzyme and genetic regulation, sorghum, Starch formation, Starch structures](#)

Pubmed ID: [37023877](#)

DOI: [10.1016/j.ijbiomac.2023.124315](#)

[Details](#)

Large and stable genome edits at the sorghum alpha kafirin locus result in changes in chromatin accessibility and globally increased expression of genes encoding lysine enrichment.

Hurst JP, Yobi A, Li A, Sato S, Clemente TE, Angelovici R, Holding DR

Published: 1 April 2023 in *Frontiers in plant science*

Keywords: [ATACseq, CRISPR/Cas9, Grain quality, lysine, RNASeq, seed protein, sorghum](#)

Pubmed ID: [36998682](#)

DOI: [10.3389/fpls.2023.1116886](#)

[Details](#)

Genome-wide identification and comparative analyses of key genes involved in C4 photosynthesis in five main grainaceous crops

Research Highlights



Evidence Found Supporting Stay-Green Alleles Involvement in Drought Adaptation in Sorghum

SorghumBase Team / 26 May 2022

Sorghum (*Sorghum bicolor* (L.) Moench) is a drought resistant crop of great importance as a staple cereal in developing countries in the semi-arid tropics. Periodic and more continuous drought conditions can severely limit plant productivity. The physiological and genetic basis of sorghum's drought resistance remains poorly understood. Gaining a better understanding of this would open up opportunities for breeders to use selection to increase drought adaptability and production under these adverse conditions and would be of great interest, not only to breeders, but to sorghum geneticists as well.

A multi-institutional team, led by scientists from Kansas State University and Institut Sénégalais de Recherches Agricoles studied 590 sorghum accessions of the West African sorghum association panel (WASAP) grown in 10 different environments and conducted genome-wide association studies (GWAS) using the previous genotyping-by-sequencing (GBS) single-nucleotide polymorphism (SNP) dataset. Sorghum plants were exposed to managed water stress in four possible conditions over the course of 4 years: pre-flowering water stress (WSt), post-flowering water stress (WS2), well-watered (WW), and rainfed (RF). Days to 50% flowering, aboveground dry biomass, plant height, and plant grain yield components (including grain weight), panicle weight and grain number per panicle, and 1000-grain weight were measured. Overall, across the tested environments, plant grain yield and biomass heritability was 33-92%. In the WSt and WS2 conditions there was a significant correlation between the grain weight per panicle and the stress tolerance index.

Based on the GWAS studies, it was concluded that SbZTF1 and SbCNT2, orthologs of maize (*Zea mays* L.) flowering genes, are strong candidates for the regulation of flowering time variation under water stress. These genes were not previously pinpointed as sources of genetic variation in U.S. germplasm. Flowering time is a potential adaptive characteristic, with early flowering time allowing plants under water stress to produce flowers and seeds earlier, thereby escaping the drought. There was a lack of overlap between the pre- and post-flowering drought tolerance, which indicates a potential trade-off between the two mechanisms. The fact that both flowering conditions have a moderate positive correlation with grain yield suggests that either mechanism confers drought tolerance.

Associations between the stress tolerance index and drought effects were found through GWAS; these included 16 putative pleiotropic associations. Thirty of the associations colocalized with Stg1, Stg2, Stg3, and Stg4 loci and had large effects. These stay-green (Stg) loci (Stg1–Stg4) have been linked to post-flowering drought tolerance. The Stg alleles were first found in Ethiopian sorghum lines (Stx642). The study provides preliminary support for the hypothesis that Stg alleles are involved in drought adaptation not just in Ethiopia, but across Africa as well. In addition, the study finds evidence supporting natural genetic variation for drought in West African sorghum.

"Jacques study was an important contribution because it connects a drought tolerance mechanism that was well characterized in biparental families to broader questions on genomic basis of drought adaptation in diverse global germplasm." – Morris

SorghumBase examples

SbZTF1 = Sobic.0060201600 and SbCNT2 = Sobic.0030295300 genes are orthologs of maize (*Zea mays* L.) flowering genes were found to be involved in flowering time variation under water stress in sorghum.

A.

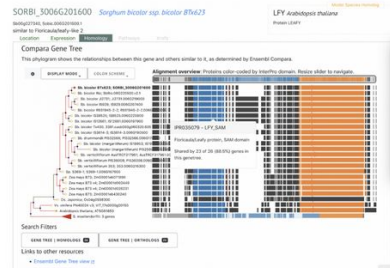


Figure 1: SbZTF1 (SORBI_30060201600) is an ortholog of a maize (*Zea mays* L.) flowering gene found to be involved in flowering time variation under water stress to sorghum in this study. The above image (A) shows conservation of this gene product across sorghum variants, maize, rice and grapevine. Protein domains are displayed in colors, and by clicking on them, the name, identifier, description and some statistics for that given domain are provided as shown in the inset for IP053070, Folic acid early protein, SAM domain (LFP_SAM). The image below (B) shows the QTLs available in the gene's region, and which include four of them associated with days to flowering (DraT).

B.



Figure 2: https://sorghumbase.org/Sorghum_bicolor/Gene/Summary?g=SORBI_30060201600-r=655289735-65293320-t=EE511382

Reference

Faye JM, Akata EA, Sine B, Diatta C, Cisse N, Fonceka D, Morris GP. Quantitative and population genomics suggest a broad role of stay-green/tpg loci in the drought adaptation of sorghum. *Plant Genome*. 2022 Mar;15(1):e20176. PMID: 34817181. DOI: 10.1002/tpg2.20176. [Read more](#).

Related Project Websites:

Morris lab: <https://www.morrislab.org/home>



Image 1: There is abundant sorghum diversity in West Africa, such as these landraces in a farmer's field in Senegal. Photo credit Geoff Morris (Colorado State).



<https://sorghumbase.org/>

- News
- Meetings & Events
- Release Notes

RELEASE NOTES

Release 5

Releases

- Release 5
- Release 4
- Release 3
- Release 2
- Initial Release

Overall Highlights

New Genomes

- [Populus trichocarpa](#) (poplar tree)
- [Sorghum bicolor CHINESE AMBER](#) (Chinese Amber [PI 22913])
- [Sorghum bicolor GRASSL](#) (Grassl [PI 154844])
- [Sorghum bicolor LEOTI](#) (Leoti [PI 586454])
- [Sorghum bicolor IS2382](#) (IS2382 [PI 229841])
- [Sorghum bicolor IS13633](#) (IS13633 [PI 297155])
- [Sorghum bicolor IS11069](#) (IS11069 [PI 329311])
- [Sorghum bicolor MBONOU](#) (Mbonou [PI 506069])
- [Sorghum bicolor AP79-714](#) (AP79-714 [PI 510757])
- [Sorghum bicolor PINK KAFIR](#) (Pink Kafir [PI 655972])
- [Sorghum bicolor RIO](#) (Rio [PI 563295])


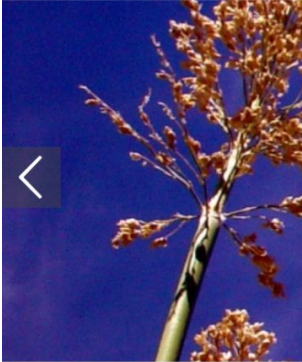
New Data

- Nearly 44 million genetic variants including ~38 M SNPs and ~5 M indels genotyped in 400 SAP accessions (Boatwright *et al.*, 2022).
- Eighty-three new synteny maps including those between each of the 28 sorghum genomes in SorghumBase, and maize B73 v5 or Japonica rice IRGSPv1, and the sorghum BTx623 reference and each of the other 27 sorghum varieties.
- Whole-genome alignments between the sorghum BTx623 reference and Arabidopsis TAIR10, Japonica rice IRGSPv1, or maize B73 v5.

SorghumBase Release 5.0

Released: December 2022

Top bar: Genomes drop-down menu



CP-NAM	REFERENCE	REFERENCE	DATA ACCESS
ChineseAmber	IS12661	BTx623	Species Table
Grassl	IS36143	Tx430	Phylogenetic Overview
Leoti	IS8525	Tx436	FTP Site
Rio (UNCC)	IS929	Tx2783	
PI 229841	Ji2731	Rio	
PI 297155	R931945-2-2		
PI 329311	IS19953		
PI 506069	PI 525695		
PI 510757	PI 532566		
PI 655972	PI 536008		
	AusTRCF317961		
	353		
	S369-1		



Genetic Variator

[Explore the Impact of Variants on Gene Structure](#)



[Access Curated Gene Expression Data](#)



[Plant Reactome Pathways on Community Reference Genomes](#)



Sorghum bicolor ssp. bicolor BTx623 Assembly and Gene Annotation

Genome Browser Entry Page

About Sorghum bicolor BTx623

Sorghum bicolor (L.) Moench subsp. *bicolor*, is a widely grown cereal crop, particularly in Africa, ranking 5th in global cereal production (FAO STAT 2008; <http://www.fao.org/in-action/inpho/crop-compendium/cereals-grains/>). It is a C4 grass also used for sugar production, brewing, feedstock, and as a biofuel crop. Its diploid genome (~730 Mbp) has a haploid chromosome number of 10. The inbred variety 'BTx623' is the current reference genome for sorghum. It has short stature and an early maturing genotype used primarily to produce grain sorghum hybrids. It is a line susceptible to sugarcane aphid and sensitive to low nitrogen, and therefore often used in functional comparative studies.

Germplasm

U.S. National Plant Germplasm System (GRIN - Global Identifier for BTx623: [PI_564163](#)).

This germplasm is part of the following population panels:

- [Sorghum Association Panel \(SAP\)](#) - 407 accessions ([Casa et al. 2008](#))
- [Sorghum Bioenergy Association Panel \(BAP\)](#) - 386 accessions ([Brenton et al. 2016](#))

Image



Image source: [The GRIN database](#)

Assembly

The genome assembly of *Sorghum bicolor* cv. Moench was published in 2009 (Paterson et al. 2009). The present assembly corresponds to v3.1.1 at the US Department of Energy Joint Genome Institute (JGI) described in (McCormick et al. 2018), and is also known as the NCBIV3 assembly. Sequencing by the JGI's [Community Sequencing Program](#) in collaboration with the Plant Genome Mapping Laboratory at the University of Georgia, followed a whole-genome shotgun strategy reaching 8X coverage (where possible- being assigned to the genetic map. JGI did two additional rounds of improvements. The most recent update of released v3.0 included ~351 Mb of finished sorghum sequence. A total of 349 clones were manually inspected, then finished and validated using a variety of technologies including Sanger, 454 and Illumina. They were integrated into chromosomes by aligning to v1.0 assembly. As a result, 4,426 gaps were closed, and a total of 4.96 Mb of sequence was added to the assembly. Overall continuity (contig N50) increased by a factor of 5.8X from 204.5 Kb to 1.2 Mb. For more details, see [Phytozome](#).

NCBI accession: [GCA_000003195.3](#)

Annotation

Gene predictions resulted from combining homology-based and *ab initio* methods with expressed sequences from sorghum, maize and sugarcane, using the JGI annotation pipeline (Goodstein et al. 2012). The SorghumBase browser presents data from the current JGI v3.1.1 release, which comprises the v3.0.1 assembly and v3.1.1 gene set (February 2017). Read more at [Phytozome](#).

This is a modern annotation using resources used in the original v1.0 release (Sb1 assembly and Sb1.4 gene set) and geneAtlas RNA-seq data. The main genome is in 10 chromosomes with small unmapped pieces, some of which contain annotated genes. The NCBIV3 release (Phytozome v3.1.1) is essentially the same as Phytozome v3.1 except for 82 genes/loci that were inactivated due to 4 scaffolds entirely present in chromosome(s) that were removed.

Repeats

Repeats were annotated with the Ensembl Genomes repeat feature pipeline (Aken et al. 2016), which uses six classes of repeats loaded from ENA.

Repeat feature	Frequency	Coverage (Mb)	% of the genome covered
Low complexity (Dust) features	685,783	29	4
RepeatMasker (with RepBase library)	455,749	451	62.1
RepeatMasker (with REdat library)	392,778	409	56.2
Tandem repeats (TRF) features	245,654	41	5.7

Baseline Gene Expression (Atlas)

Baseline gene expression data from seven sorghum BTx623 datasets curated and processed by the [EMBL-EBI Expression Atlas](#).

Click here for an [example of baseline gene expression](#) for the msd2 gene.

Variation

Variation in SorghumBase is available for short variants (genetic variation, which in turn may be naturally occurring or chemically induced) and QTL variants associated with physical traits.

Genetic Variation

Genetic variation data for a sorghum gene is available graphically and in tabular form, and for each variant, a Variant page provides more detailed information. Below are examples of each of these data representations.

- Click here for an [image of the genetic variants](#) mapping to the smaller transcript of the msd2 gene.
- Click here for a [table of the genetic variants](#) mapping to the smaller transcript of the msd2 gene.
- Click here for a [summary page of a chemically induced variant](#) (mp_6_47487513_G_A).

Naturally occurring genetic variation

Single Nucleotide Polymorphisms (SNPs). Currently in SorghumBase, there is a single SNP data set for sorghum BTx623:

- The [Lozane et al \(2021\)](#) data set includes early 13 million naturally occurring SNPs in 499 sorghum accessions (including 14 duplicate samples). Included are accessions from the TERRA-MEPP and TERRA-REF population panels, and lines previously genotyped by Emma Mace and collaborators in 2013.
- The [Boatwright SNP dataset](#) (Boatwright et al. 2022) consists of almost 44 million genetic variants including nearly 38 million SNPs and 5 million indels genotyped in 400 Sorghum Association Panel (SAP) accessions via whole-genome sequencing (WGS).

Chemically induced variation

Ethyl methane sulfonate (EMS)-induced mutations. Currently in SorghumBase, there are two collections of EMS-induced mutant lines. EMS is a chemical commonly used to cause point mutations, that is, to change single nucleotides in the DNA of a plant seed. The Jiao dataset was recalled and the Addo-Quaye dataset was newly added.

- The [Addo-Quaye et al \(2018\)](#) data set includes over 2.6 million point mutations identified in 486 sorghum accessions corresponding to the M3 generation of an EMS-mutagenized sorghum population.
- The Xin EMS dataset ([Jiao et al. 2016](#)) features over 1.7 million variants recalled from the original EMS-induced G/C to A/T transition mutations data set annotated from 252 M3 families selected from the 6,400 sorghum mutant library in BTx623 background described by Xin and colleagues (Xin et al. 2008). Genomic DNA used for sequencing was pooled from 20 M3 plants per M2 family ([Jiao et al. 2016](#)).

Phenotypic Variation

Quantitative Trait Locus (QTLs). Data corresponding to 5,843 QTL features for 220 sorghum traits were imported from [Sorghum QTL Atlas](#) and are provided with predicted synthetic locations in maize and rice.

Example region with QTLs associated with multiple traits including greenbug resistance, fresh biomass, and flag leaf height. Hint: For additional regions with QTL data in the current sorghum assembly (v3), use the [physical or genetic \(cM\) coordinates](#) kindly provided by the Sorghum QTL Atlas team.

References

Addo-Quaye C, Tuinstra M, Carraro N, Weil C, Dilkes BP. Whole-Genome Sequence Accuracy Is Improved by Replication in a Population of Mutagenized Sorghum. *Genome*. 2018;8: 1079–1094. doi: [10.1534/g3.117.300301](#).

Aken, Brownen L., Sarah Ayling, Daniel Barrell, Laura Clarke, Valery Curwen, Susan Fairley, Julio Fernandez Banet, et al. 2016. "The Ensembl Gene Annotation System." *Database: The Journal of Biological Databases and Curation*. PMID: [27337980](#). doi: [10.1093/database/baw093](#).

Brenton, Zachary W, Elizabeth A. Cooper, Matthew T. Myers, Richard E. Boyles, Nadia Shakoor, Kelsey J. Zielinski, Bradley L. Rauh, William C. Bridges, Geoffrey P. Morris, and Stephen Kresovich. 2016. "A Genomic Resource for the Development, Improvement, and Exploitation of Sorghum for Bioenergy." *Genetics* 204(1): 21–33. PMID: [27356613](#). doi: [10.1534/genetics.115.183947](#).

Casa, Alexandra M., Gael Pressoir, Patrick J. Brown, Sharon E. Mitchell, William L. Rooney, Mitchell R. Tuinstra, Clève D. Franks, and Stephen Kresovich. 2008. "Community Resources and Strategies for Association Mapping in Sorghum." *Crop Science* 48(1): 30–40. doi: [10.2135/cropsci2007.02.0080](#).



<https://sorghumbase.org/>

Gene Expression

Genetic Variation
(SNP, indels, EMS)

Phenotypes (QTL)

References





Sorghum bicolor ssp. bicolor IS12661 Assembly and Gene Annotation

About Sorghum IS 12661

IS 12661 is a landrace of *Sorghum bicolor* collected at Dire Dawa-Isa-Gurgur (9.579999, 41.860001), Hararghe province, Ethiopia. Other identifiers are SA 2355 and PI 276837. The accession is classified in the *caudatum* race and considered representative of the zerzera type within that race. IS 12661 and several of its derivatives from the Sorghum Conversion Program (SC170s) have contributed substantially to modern grain sorghum breeding.

Germplasm

U.S. National Plant Germplasm System (GRIN - Global) identifier for IS 12661: [PI 276837](#).

This germplasm is part of the following population panels:

- [Sorghum Association Panel \(SAP\)](#) - 407 accessions ([Casa et al. 2008](#))
- [Sorghum Bioenergy Association Panel \(BAP\)](#) - 386 accessions ([Brenton et al. 2016](#))

Accessions names and identifiers

- ICRISAT: [IS 12661](#)
- USDA GRIN: [PI 276837 \(IS 12661\)](#)
- USDA GRIN converted line: [PI 534157 \(converted\)](#)
- FAO (DOI): [10.18730/NBYG*](#)
- SAP: SAP-391
- CNGB: [CNS0314253](#)
- Other: SA 2355

Passport data

- Passport data for GRIN accession: [PI 276837 \(IS 12661\)](#)
- Passport data for converted GRIN accession: [PI 534157 \(converted\)](#)
- Germplasm page at Genbank ICRISAT: [IS 12661](#)
- Passport summary at ICRISAT: [IS 12661](#)
- Passport summary at FAO: [10.18730/NBYG*](#)
- Passport data at Genesys: [IS 12661](#)

Assembly

The genome assembly of Sorghum IS12661 was reported in [Tao et al. \(2021\)](#). Sequencing was conducted by the Australia Sorghum breeding team in collaboration with BGI-Shenzhen using Illumina HiSeq 4000 platform and the PacBio Sequel platform to achieve 171X and 36X coverage, respectively. The assembly effort generated a genome of 622.1Mb with contigs N50 of 163.7kb.

Annotation

Gene prediction was performed using a hybrid approach combining de novo gene predictors and evidence-based methods ([Tao et al. 2021](#)), which led to the identification of 33,979 genes in the genome.

References

- Casa, Alexandra M., Gael Pressoir, Patrick J. Brown, Sharon E. Mitchell, William L. Rooney, Mitchell R. Tuinstra, Cleve D. Franks, and Stephen Kresovich. 2008. "Community Resources and Strategies for Association Mapping in Sorghum." *Crop Science* 48 (1): 30–40. doi: [10.2135/cropsci2007.02.0080](#).
- Tao Y, Luo H, Xu J, Cruickshank A, Zhao X, Teng F, et al. Extensive variation within the pan-genome of cultivated and wild sorghum. *Nat Plants*. 2021;7: 766–773. PMID: [34017083](#). doi: [10.1038/s41477-021-00925-x](#).

More information

General information about this species can be found in [Wikipedia](#).

Summary

Assembly:	IS12661, May 2021
Database version:	87.1
Base Pairs:	622,443,737
Golden Path Length:	622,443,737
Genebuild method:	Ware-lab
Genebuild started:	May 2021
Genebuild released:	May 2021
Genebuild last updated/patched:	May 2021
Genebuild version:	2021-05

Gene counts

Gene transcripts:	33,979
--------------------------	--------

Coordinate Systems

chromosome 10 sequences

Sequence	Length (bp)
1	70864131
2	68988801
3	70364006
4	56710613
5	50543705
6	50440342
7	58129253
8	53619283
9	52860237
10	50420780

scaffold 704 sequences

Sequence	Length (bp)
scaffold27	972315
scaffold40	874591
scaffold48	835291
scaffold58	935888
scaffold65	724219
scaffold76	689693
scaffold78	698964
scaffold81	670117
scaffold95	607080
scaffold124	534030

Showing 1 to 10 of 704 entries

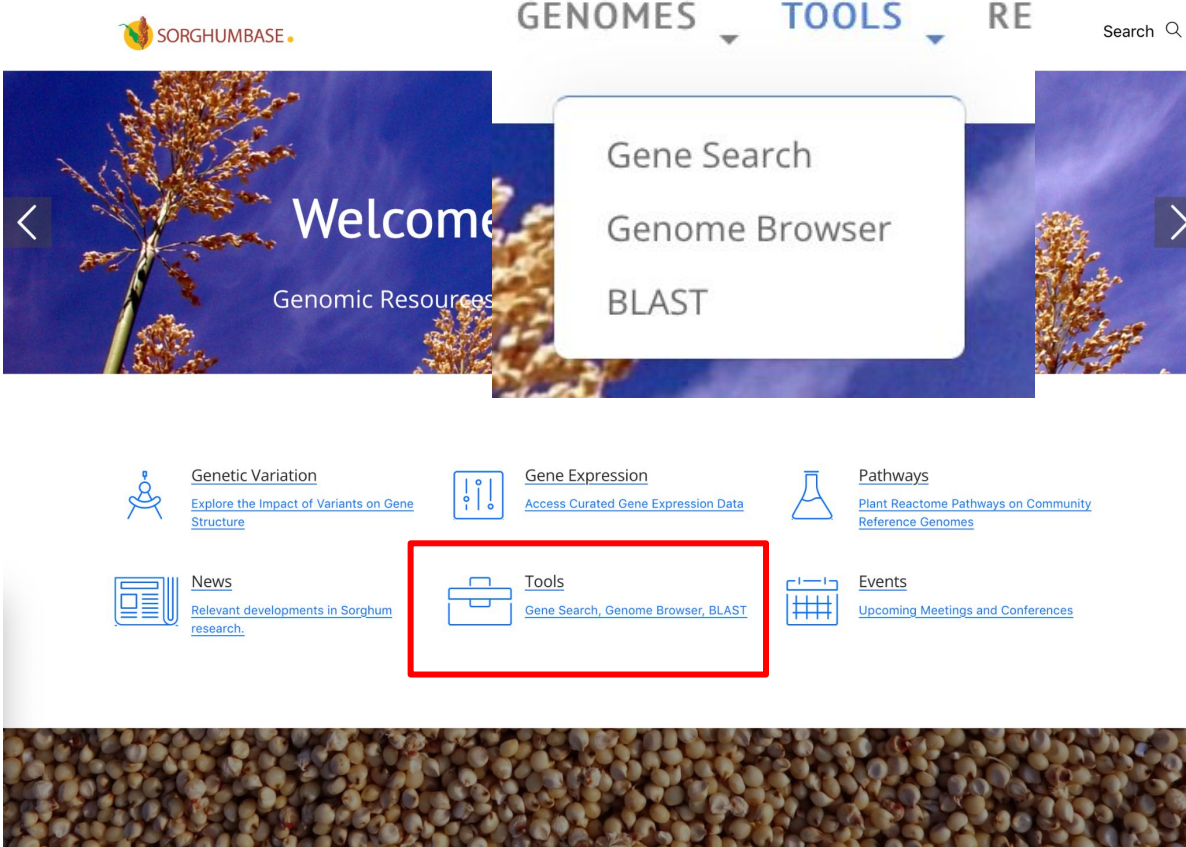
chunk 6805 sequences

GRIN (USDA):
germplasm &
population panels

Passport data
(germplasm
repositories)

References

SB Homepage: Tools



The screenshot shows the SORGHUMBASE homepage. At the top left is the SORGHUMBASE logo. In the center, there are navigation tabs for GENOMES, TOOLS, and RE. A search bar is located to the right of these tabs. Below the navigation is a large banner with a sorghum plant and the text "Welcome Genomic Resources". A dropdown menu is open under the TOOLS tab, listing "Gene Search", "Genome Browser", and "BLAST". Below the banner are several navigation links with icons: Genetic Variation (Explore the Impact of Variants on Gene Structure), Gene Expression (Access Curated Gene Expression Data), Pathways (Plant Reactome Pathways on Community Reference Genomes), News (Relevant developments in Sorghum research), Tools (Gene Search, Genome Browser, BLAST), and Events (Upcoming Meetings and Conferences). The Tools link is highlighted with a red box. At the bottom of the page is a wide image of sorghum grains.

SB Homepage: Tools

Scroll down...

TOOLS



Gene Search

Search for genes by id, name, pathway, domain, or ontology term



Genome Browser

Genome browser powered by Ensembl



BLAST

Search by sequence

SB Homepage: Tutorial or User Guide

QUICK GUIDES

Guides

- Genomic Resources
 - search-interface
 - gene-search
 - gene-page
 - configure-image
 - comparative-genomics
 - regulation
 - gene-expression
 - pathways
 - variation
 - QTLs
 - references
- BLAST
 - overview
 - blast-job

Genomic Resources

search-interface

Using the genes search interface

Watch the [demo video on YouTube](#).

Searching for a gene given a gene id can be done in a couple of ways:

- Click the spyglass icon and type/paste **SORBI_3006G095600** into the search box
- Link directly to http://sorghumbase.org/genes?idList=SORBI_3006G095600

If you don't know the gene id you can search by pathway and taxonomy [example](#):

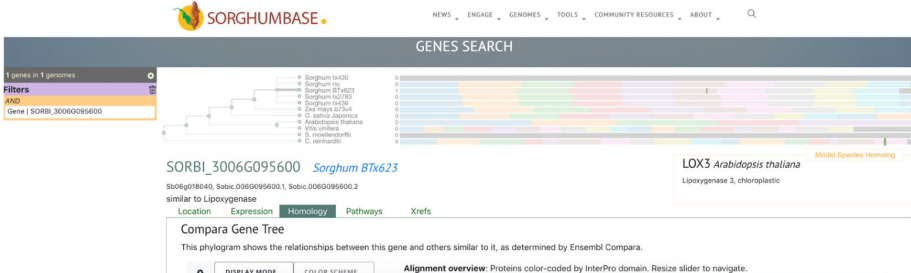
- In the search box, start typing **jasmonic acid biosynthesis** and choose the matching Plant Reactome Pathway term
- To limit the search to genes in *Sorghum bicolor* BTx623, type **sorghum** in the search box and select the matching Taxonomy term

The search results page is organized into three areas:

- A panel on the left side shows the status of the search
- At the top is a visualization of the genomic locations of genes matching the search
- Below this is a paginated list of genes

Each gene has a set of tabs that can be expanded to explore more details:

- Location - Lightweight genome browser showing gene structure
- Expression - Baseline gene expression viewer from [EBI-Expression Atlas](#)
- Homology - Gene family tree viewer (opens by default if only one gene is found)
- Pathways - Associated pathways in [Gramene's Plant Reactome](#)
- Xrefs - Cross-references to other databases



NEWS ENGAGE GE

Training Materials

Mailing List

Contact Us

What can I search?

For genes:

- gene identifiers
- gene names
- gene descriptions
- gene synonyms
- interpro domains
- taxonomy
- ontology terms (interpro2go)
- pathways

For papers:

- author names
- title
- abstract
- key words
- pubmed id
- doi

New in R6:

- QTLs
- trait & plant ontology terms (TO/PO)

How to use SorghumBase?

Search for author, gene name/ID/function, pathway, molecular function, paper

Gene IDs:

- SORBI_3006G095600
- Sobic.006G095600
- Sb06g018040



*Search for traits in
Release 6...*

msd2

Gladman *et al* (2019)

Integrated Search Views



https://sorghumbase.org/



NEWS

ENGAGE

GENOMES

TOOLS

RESEARCH

ABOUT

Search Sorghumbase

gladman

Genes 0 terms

Website 8 matches

Posts 3

Events 1

Papers 4

Sorghum root epigenetic landscape during limiting phosphorus conditions.

Gladman N, Hufnagel B, Regulski M, Liu Z, Wang X, Chougule K, Kochian L, Magalhães J, Ware D

chromatin modification, DNA methylation, histone marks, histone methylation, phosphorus efficiency, phosphorus responsive genes, RNA seq, root system architecture; root system remodeling; phosphorous deficiency

Monday, May 23, 2022

[Read more](#)

Fertility of Pedicellate Spikelets in Sorghum Is Controlled by a Jasmonic Acid Regulatory Module.

Gladman N, Jiao Y, Lee YK, Zhang L, Chopra R, Regulski M, Burow G, Hayes C, Christensen SA, Dampanaboina L, Chen J, Burke J, Ware D, Xin Z

gene expression, jasmonic acid signaling, plant development, transcriptional regulators

Thursday, October 10, 2019

[Read more](#)

SorghumBase: a web-based portal for sorghum genetic information and community advancement.

Gladman N, Olson A, Wei S, Chougule K, Lu Z, Tello-Ruiz M, Meijis I, Van Buren P, Jiao Y, Wang B, Kumar V, Kumari S, Zhang L, Burke J, Chen J, Burow G, Hayes C, Emendack Y, Xin Z, Ware D

Sorghum MSD3 Encodes an ω -3 Fatty Acid Desaturase that Increases Grain Number by Reducing Jasmonic Acid Levels.

Dampanaboina L, Jiao Y, Chen J, Gladman N, Chopra R, Burow G, Hayes C, Christensen SA, Burke J, Ware D, Xin Z
fatty acid desaturase, grain number, jasmonic acid, msd, multiseeded, MutMap,

- Type one or more search terms:
- Author last name (Gladman)
 - Gene name or ID, mutant
 - Protein domain
 - Molecular function (ontology)
 - Pathway

Integrated Search Views

gladman msd2

Genes 1 terms

Website 2 matches

Posts 0

Events 0

Papers 2

Fertility of Pedicellate Spikelets in Sorghum Is Controlled by a Jasmonic Acid Regulatory Module.

Gladman N, Jiao Y, Lee YK, Zhang L, Chopra R, Regulski M, Burow G, Hayes C, Christensen SA, Dampanaboina L.

Thursday, October 10, 2019

[Read more](#)

Sorghum MSD3 Acid Desaturase Number by Red Levels.

Dampanaboina L.

multiseeded, MutMap, sorghum

Wednesday, October 30, 2019

Fertility of Pedicellate Spikelets in Sorghum Is Controlled by a Jasmonic Acid Regulatory Module.

Gladman N, Jiao Y, Lee YK, Zhang L, Chopra R, Regulski M, Burow G, Hayes C, Christensen SA, Dampanaboina L, Chen J, Burke J, Ware D, Xin Z

Published: 11 October 2019 in *International journal of molecular sciences*

Keywords: gene expression, jasmonic acid signaling, plant development, transcriptional regulators

Pubmed ID: 31597271

DOI: 10.3390/ijms20194951

Abstract

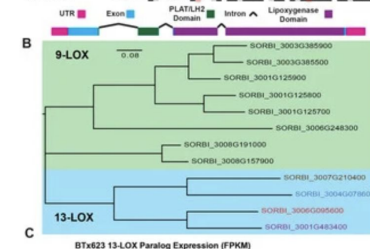
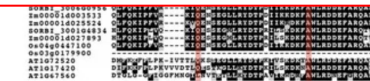


As in other cereal crops, t sessile spikelets (SSs) are eventually abort. Consequ underlying this differentia multiseeded (msd) mutan branched/Cycloidea/PCF) show that MSD2 encodes binds to the promoters of development and spikelet increasing GNP and grain

3.1. MSD2 Encodes a Lipoygenase in the Jasmonic Acid Biosynthetic Pathway

Sorghum bicolor (L.) Moench plants manifesting the *multiseeded* phenotype were identified from a collection of EMS-induced single nucleotide polymorphisms (SNP) [24]. *MSD1*, which encodes a TCP (*Teosinte branched 1* (*TB1*), *Cycloidea* (*Cyc*), and *Proliferating Cell nuclear antigen binding Factor* (*PCF*)) [26,27,28] transcription factor, was the first to be characterized, revealing a role in controlling bioactive JA levels in developing floral meristems [7]. To identify additional causative alleles, we subjected ~~seventeen different msd2 mutants to whole genome sequencing followed by comparative variant calling analysis.~~ Three of these independent alleles, *msd2-1*, *-2*, and *-3*, localized to SORBI_3006G095600 (Sb06g018040) [7], which encodes a class II 13-lipoygenase that shares >95% amino acid identity with the maize *tasselseed 1* (*TS1*) gene [16]. SORBI_3006G095600 is syntenic to *TS1* and is the closest related maize orthologue based on a maximum likelihood phylogenetic analysis (Supplemental Figure S1). The *msd2-1* mutant harbors a nonsense mutation (peptide residue Q402*) and *msd2-2* a missense mutation (peptide residue A423V), respectively, within the lipoygenase (LOX) domain (Figure 1A). The *msd2-3* allele contains the same mutation as *msd2-1*, but the lines are not siblings, as evidenced by the lack of these shared SNPs.

***msd2-1*, *-2*, and *-3*, localized to SORBI_3006G095600 (Sb06g018040)**



GENES SEARCH

1 genes in 1 genomes

Filters

AND

Gene | SORBI_3006G095600

EXPAND EMPTY BRANCHES

Sb. bicolor BTx623

SORBI_3006G095600 *Sorghum bicolor ssp. bicolor BTx623*

Sb06g018040, Sobic.006G095600.1, Sobic.006G095600.2

similar to Lipoxigenase

Location Expression Homology Pathways Xrefs

Model Species Homolog

LOX3 *Arabidopsis thaliana*

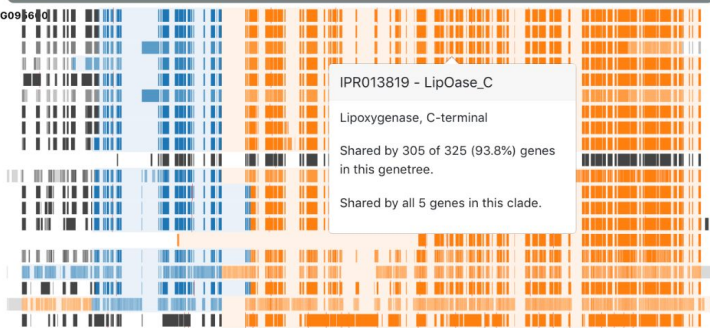
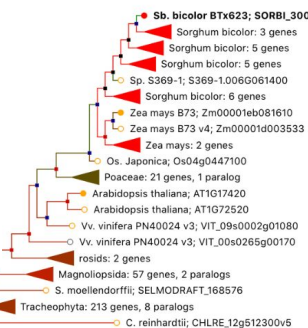
Lipoxygenase 3, chloroplastic

Compara Gene Tree

This phylogram shows the relationships between this gene and others similar to it, as determined by Ensembl Compara.

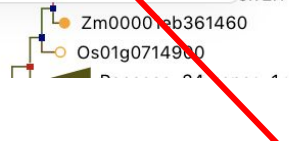
DISPLAY MODE COLOR SCHEME

Alignment overview: Proteins color-coded by InterPro domain. Resize slider to navigate.



DISPLAY MODE COLOR SCHEME

- Alignment Overview
- Multiple Sequence Alignment
- Neighborhood Conservation



Search Filters

GENE TREE | HOMOLOGS 325 | GENE TREE | ORTHOLOGS 81 | GENE TREE | PARALOGS 13

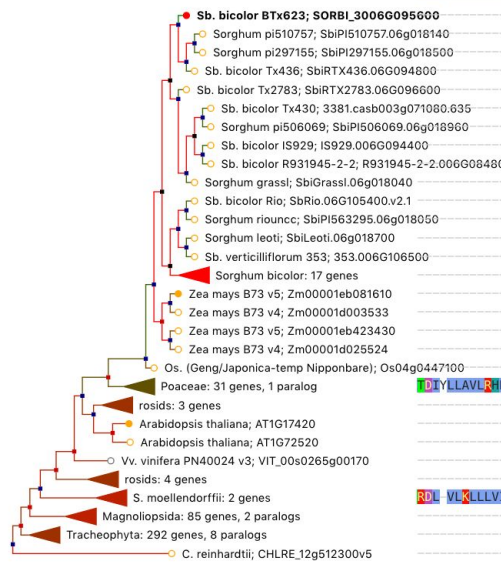
Links to other resources

- Ensembl Gene Tree view

Compara Gene Tree

This phylogram shows the relationships between this gene and others similar to it, as determined by Compara.

DISPLAY MODE COLOR SCHEME

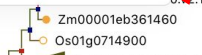


Multiple Sequence Alignment: Amino acid MSA. Drag slider to reposition.

Gene	Sequence
Sb. bicolor BTx623; SORBI_3006G095600	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWDAITRWGDAYVAMYPSPD AVIGDTELOSNYREA VDTGHAQK RGAP WWPRLSTF
Sorghum pi510757; SbiPI510757.06g018140	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWDAITRWGDAYVAMYPSPD AVIGDTELOSNYREA VDTGHAQK RGAP WWPRLSTF
Sorghum pi297155; SbiPI297155.06g018500	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWDAITRWGDAYVAMYPSPD AVIGDTELOSNYREA VDTGHAQK RGAP WWPRLSTF
Sb. bicolor Tx436; SbiRTX436.06G094800	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWDAITRWGDAYVAMYPSPD AVIGDTELOSNYREA VDTGHAQK RGAP WWPRLSTF
Sb. bicolor Tx2783; SbiRTX2783.06G096600	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWDAITRWGDAYVAMYPSPD AVIGDTELOSNYREA VDTGHAQK RGAP WWPRLSTF
Sb. bicolor Tx430; 3381.casb003g071080.635	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWDAITRWGDAYVAMYPSPD AVIGDTELOSNYREA VDTGHAQK RGAP WWPRLSTF
Sorghum pi506069; SbiPI506069.06g018960	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWDAITRWGDAYVAMYPSPD AVIGDTELOSNYREA VDTGHAQK RGAP WWPRLSTF
Sb. bicolor IS929; IS929.006G094400	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWDAITRWGDAYVAMYPSPD AVIGDTELOSNYREA VDTGHAQK RGAP WWPRLSTF
Sb. bicolor R931945-2-2; R931945-2-2.006G084800	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWDAITRWGDAYVAMYPSPD AVIGDTELOSNYREA VDTGHAQK RGAP WWPRLSTF
Sorghum grassl; SbiGrassl.06g018040	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWDAITRWGDAYVAMYPSPD AVIGDTELOSNYREA VDTGHAQK RGAP WWPRLSTF
Sb. bicolor Rio; SbRio.06G105400.v2.1	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWDAITRWGDAYVAMYPSPD AVIGDTELOSNYREA VDTGHAQK RGAP WWPRLSTF
Sorghum riouccc; SbiPI563295.06g018050	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWDAITRWGDAYVAMYPSPD AVIGDTELOSNYREA VDTGHAQK RGAP WWPRLSTF
Sorghum leoti; SbiLeoti.06g018700	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWDAITRWGDAYVAMYPSPD AVIGDTELOSNYREA VDTGHAQK RGAP WWPRLSTF
Sb. verticilliflorum 353; 353.006G106500	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWDAITRWGDAYVAMYPSPD AVIGDTELOSNYREA VDTGHAQK RGAP WWPRLSTF
Sorghum bicolor: 17 genes	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWDAITRWGDAYVAMYPSPD AVIGDTELOSNYREA VDTGHAQK RGAP WWPRLSTF
Zea mays B73 v5; Zm00001eb081610	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWSAIRWGDYVAMYPSPDE AVIGDTELOSNYREA VHTGHAQK RDAP WWPRLSTF
Zea mays B73 v4; Zm00001d003533	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWSAIRWGDYVAMYPSPDE AVIGDTELOSNYREA VHTGHAQK RDAP WWPRLSTF
Zea mays B73 v5; Zm00001eb423430	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWSAIRWGDYVAMYPSPDE AVIGDTELOSNYREA VHTGHAQK RDAP WWPRLSTF
Zea mays B73 v4; Zm00001d025524	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWSAIRWGDYVAMYPSPDE AVIGDTELOSNYREA VHTGHAQK RDAP WWPRLSTF
Os. (Geng/Japonica-temp Nipponbare); Os04g0447100	RGMAVEDPSPRP HG LRLLTTE DYPYAADGLLWSAITSRWGEYVAAYVSPDE AVIGDTELOSNYREA VDSGHAQK RGAP WWPRLSTF
Poaceae: 31 genes, 1 paralog	LDIYLLAVRHLVLRGVAVEDASQP HG LRLLTTE DYPYAADGLLWSAITSRWGEYVADYDYPDAG AVIGDTELOSNYREA VHTGHAQK RHAP WWPRLSTF
rosids: 3 genes	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWSAITSRWGEYVAMYPSPDE AVIGDTELOSNYREA VHTGHAQK RDAP WWPRLSTF
Arabidopsis thaliana; AT1G17420	RGMAIPDITQP HG LRLLTTE DYPYADGLLWSAITSRWGEYVAMYPSPDE AVIGDTELOSNYREA VHTGHAQK RDAP WWPRLSTF
Arabidopsis thaliana; AT1G72520	RGMAIPDITQP HG LRLLTTE DYPYADGLLWSAITSRWGEYVAMYPSPDE AVIGDTELOSNYREA VHTGHAQK RDAP WWPRLSTF
Vv. vinifera PN40024 v3; VIT_005o0265g00170	RGVAVPDTITQP HG LRLLTTE DYPYADGLLWSAITSRWGEYVAMYPSPDE AVIGDTELOSNYREA VHTGHAQK RDAP WWPRLSTF
rosids: 4 genes	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWSAITSRWGEYVAMYPSPDE AVIGDTELOSNYREA VHTGHAQK RDAP WWPRLSTF
S. moellendorffii: 2 genes	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWSAITSRWGEYVAMYPSPDE AVIGDTELOSNYREA VHTGHAQK RDAP WWPRLSTF
Magnoliopsida: 85 genes, 2 paralogs	RDL VLKLLLVLSRGMAVEDKISAP DG LRLLVLE DYPYAADGLLWSAITSRWGEYVAMYPSPD AVIGDTELOSNYREA VHTGHAQK RDAP WWPRLSTF
Tracheophyta: 292 genes, 8 paralogs	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWSAITSRWGEYVAMYPSPDE AVIGDTELOSNYREA VHTGHAQK RDAP WWPRLSTF
C. reinhardtii; CHLRE_12g512300v5	RGMAVEDPITQP HG LRLLTTE DYPYADGLLWSAITSRWGEYVAMYPSPDE AVIGDTELOSNYREA VHTGHAQK RDAP WWPRLSTF

DISPLAY MODE COLOR SCHEME

Alignment Overview 6276100 bicolor: 6 genes
Multiple Sequence Alignment bicolor: 6 genes
Neighborhood Conservation bicolor: 4 genes



Search Filters

GENE TREE | HOMOLOGS 457 GENE TREE | ORTHOLOGS 44 GENE TREE | PARALOGS 12

Links to other resources

- Ensembl Gene Tree view

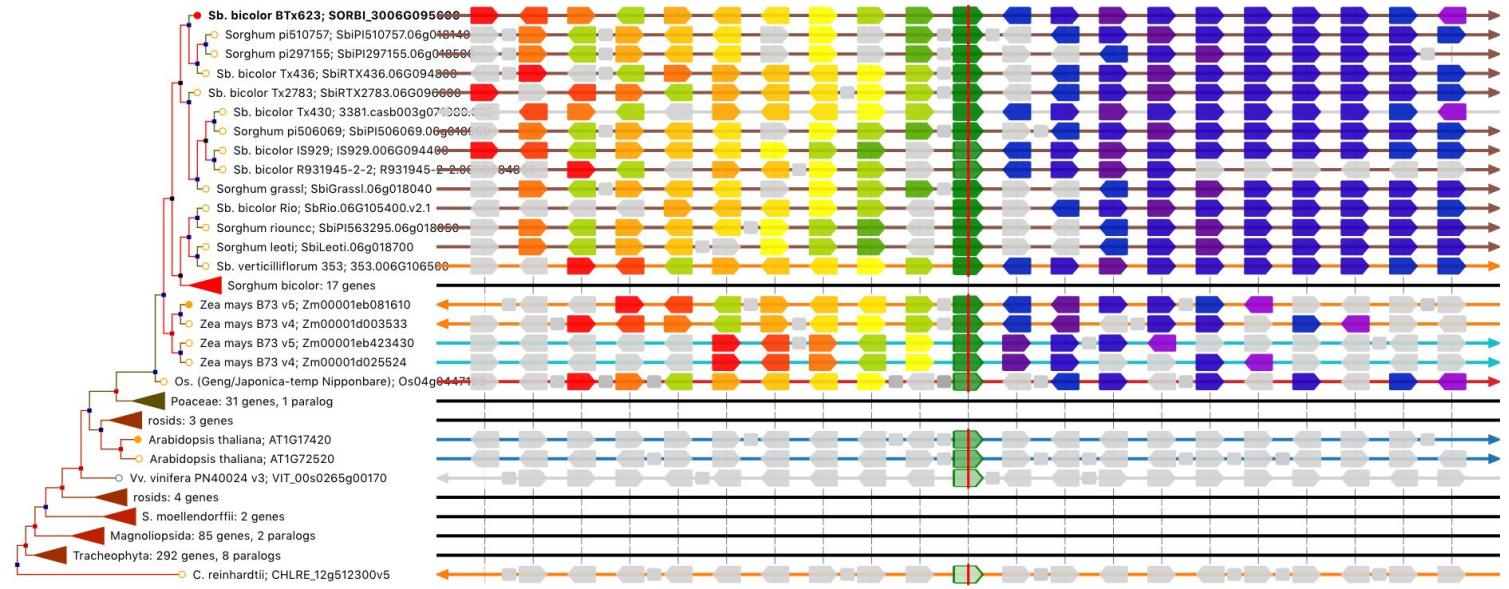
Location Expression Homology Pathways Xrefs

Compara Gene Tree

This phylogram shows the relationships between this gene and others similar to it, as determined by Ensembl Compara.

⚙ DISPLAY MODE COLOR SCHEME

Neighborhood conservation: +/- 10 genes flanking the gene of interest (top row) are color-coded by gene family. Gray genes are from other families. Central genes (green) are shaded based on similarity with gene of interest.



Search Filters

GENE TREE | HOMOLOGS 457 GENE TREE | ORTHOLOGS 44 GENE TREE | PARALOGS 12

Links to other resources

- [Ensembl Gene Tree view](#)

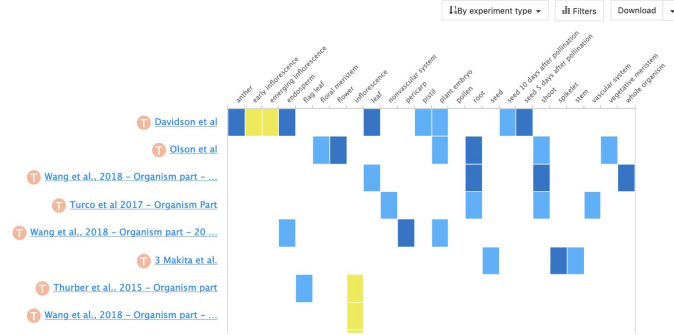
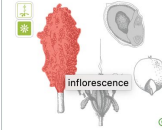
SORBI_3006G095600 *Sorghum bicolor ssp. bicolor BTx623*

Sb06g018040, Sobic.006G095600.1, Sobic.006G095600.2

similar to Lipoxygenase

Location Expression Homology Pathways Xrefs

Showing 11 experiments:



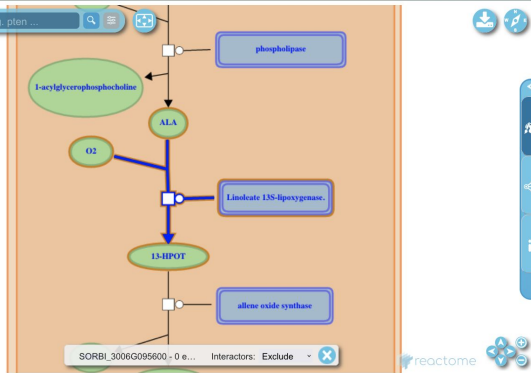
SORBI_3006G095600 *Sorghum bicolor ssp. bicolor BTx623*

Sb06g018040, Sobic.006G095600.1, Sobic.006G095600.2

similar to Lipoxygenase

Location Expression Homology Pathways Xrefs

- Metabolism and regulation
 - Hormone signaling, transport, and metabolism
 - Jasmonic acid biosynthesis
 - O2+linolenate->13-HPOT
 - Secondary metabolism
 - 13-LOX and 13-HPL pathway
 - O2+linolenate->13-HPOT



Search Filters

ALL GENES IN THIS REACTION

Links to other resources

- Plant Reactome Reaction
- Plant Reactome Gene

Sobic.006G095600 SORBI_3006G095600 *Sorghum bicolor ssp. bicolor BTx623*

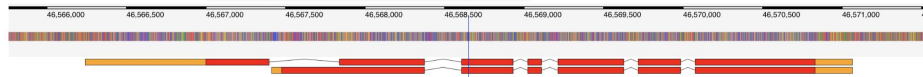
Sb06g018040

similar to Lipoxygenase

Location Expression Homology Pathways Xrefs

Currently viewing: 6:46565758-46571546

6:46,565,758..46,571,546



Search Filters

ALL ON CHROMOSOME 6

ALL WITHIN 6:46565758-46571546

Links to other resources

- Ensembl Browser
- PhytoMine

LOX3 *Arabidopsis thaliana*

Lipoxygenase 3, chloroplastic

Model Species Homology

SORBI_3006G095600



Sorghum bicolor ssp. bicolor BTx623 (Sorghum_bicolor_NCBIv3) Location: 6:46,566,240-46,571,064 Gene: Sobic.006G095600

Gene-based displays

- Summary
 - Splice variants
 - Transcript comparison
 - Gene alleles
- Sequence
 - Secondary Structure
 - Gene families
 - Literature
- Plant Compara
 - Genomic alignments
 - Gene tree
 - Gene gain/loss tree
 - Orthologues
 - Paralogues
- Pan-taxonomic Compara
 - Gene Tree
 - Orthologues
- Ontologies
 - GO: Biological process
 - GO: Molecular function
 - GO: Cellular component
- Phenotypes
 - Genetic Variation
 - Variant table
 - Variant image
 - Structural variants
- Gene expression
- Regulation
- External references
- Supporting evidence
- ID History
 - Gene history

- Configure this page
- Custom tracks
- Export data
- Share this page

Gene: Sobic.006G095600 SORBI_3006G095600

Description: hypothetical protein
Location: Chromosome 6: 46,566,240-46,571,064 forward strand.

About this gene: This gene has 2 transcripts (splice variants), 34 orthologues and 11 paralogues.

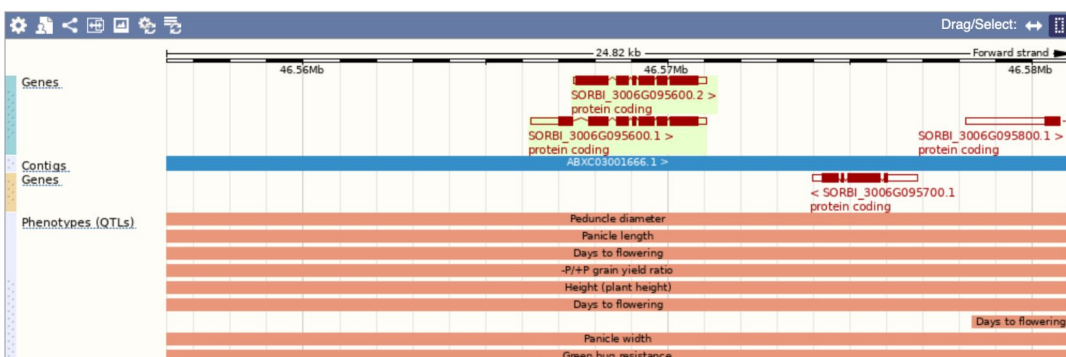
Transcripts: Hide transcript table

Show/hide columns (1 hidden)		Filter			
Name	Transcript ID	bp	Protein	Biotype	Flags
Novel	SORBI_3006G095600.1	3769	924aa	Protein coding	
Novel	SORBI_3006G095600.2	3040	913aa	Protein coding	

Summary

Name: Sobic.006G095600 (Phytozome)
Gene type: Protein coding
Annotation method: Protein coding genes annotated in ENA

Go to Region in Detail for more tracks and navigation options (e.g. zooming)

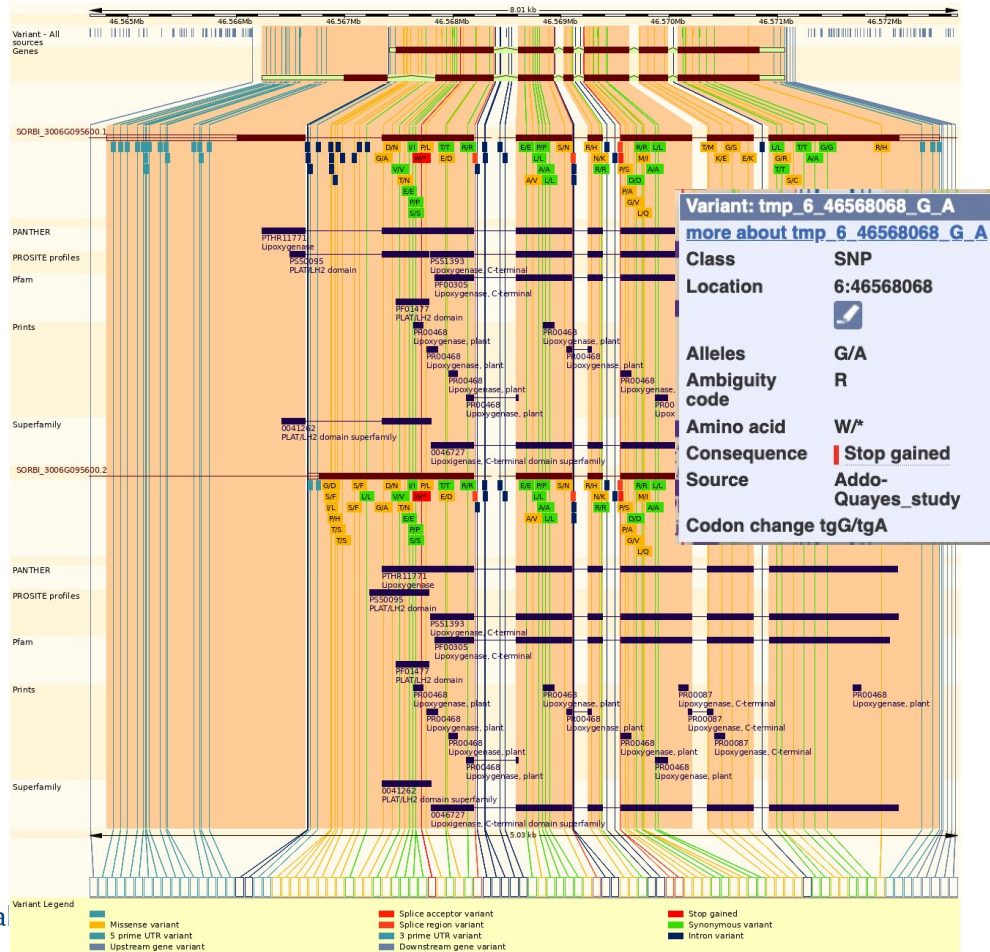


msd2 genetic variants along protein domains



<https://sorghumbase.org/>

Protein domains



Gene-based display

msd2 PTVs = Protein Truncating Variants



SORGHUM
BASE

q. umbase.org/

SORGHUM BLAST

Sorghum bicolor ssp. bicolor BTx6

Variant

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Gene alleles

Sequence

- Secondary Structure
- Gene families
- Literature

Plant Compara

- Genomic alignments
- Gene tree
- Gene gain/loss tree
- Orthologues
- Paralogues

Pan-taxonomic Compara

- Gene Tree
- Orthologues

Ontologies

- GO: Biological process
- GO: Molecular function
- GO: Cellular component

Phenotypes

- Genetic Variation
- Variant table**
- Variant image
- Structural variants
- Gene expression
- Regulation
- External references
- Supporting evidence
- ID History
- Gene history

Configure this page

Custom tracks

Export data

Share this page

Filter Filter Other Columns

Consequences (4/26 on)

Turn All Off PTV PTV & Missense Only Exonic Turn All On

PTV = Protein Truncating Variant

Transcript ablation (0) Off	Inframe deletion (0) Off	5 prime UTR variant (139) Off
Splice donor variant (0) On	Missense variant (81) Off	3 prime UTR variant (48) Off
Splice acceptor variant (3) On	protein altering variant (0) Off	Non coding transcript exon variant (0) Off
Stop gained (3) On	Splice region variant (9) Off	Intron variant (138) Off
Frameshift variant (4) On	Incomplete terminal codon variant (0) Off	NMD transcript variant (0) Off
Stop lost (0) Off	Stop retained variant (0) Off	Non coding transcript variant (0) Off
Start lost (0) Off	Synonymous variant (88) Off	Upstream gene variant (1196) Off
Transcript amplification (0) Off	Coding sequence variant (0) Off	Downstream gene variant (777) Off
Inframe insertion (0) Off	Mature miRNA variant (0) Off	

Apply » Cancel

Search Sorghumbase...

SIFT scores coming in R6

Filter Filter Other Columns

Show/hide columns Search...

Variant ID	Chr: bp	Alleles	Class	Source	Conseq. Type	AA	AA coord	Transcript
tmp_6_46568068_G_A	6:46568068	G/A	SNP	Addo-Quayes_study	Stop gained	W/*	209	SORBI_3006G095600.1
tmp_6_46569212_G_A	6:46569212	G/A	SNP	Yinping-Jiao-2016-Study	Splice acceptor variant	-	-	SORBI_3006G095600.1
tmp_6_46567780_AAT_A	6:46567781-46567782	AT/-	deletion	SAP_Boatwright_study	Frameshift variant	I/X	103	SORBI_3006G095600.2
tmp_6_46567783_T_TCG AAA	6: between 46567783 & 46567784	-/CGAAA	insertion	SAP_Boatwright_study	Frameshift variant	-/RX	104	SORBI_3006G095600.2
tmp_6_46567787_C_CGCG AA	6: between 46567787 & 46567788	-/GCAA	insertion	SAP_Boatwright_study	Frameshift variant	L/RNX	105	SORBI_3006G095600.2
tmp_6_46567803_A_AG	6: between 46567803 & 46567804	-/G	insertion	SAP_Boatwright_study	Frameshift variant	E/EX	110	SORBI_3006G095600.2
tmp_6_46568068_G_A	6:46568068	G/A	SNP	Addo-Quayes_study	Stop gained	W/*	198	SORBI_3006G095600.2
tmp_6_46569212_G_A	6:46569212	G/A	SNP	Yinping-Jiao-2016-Study	Splice acceptor variant	-	-	SORBI_3006G095600.2

EMS Variant: Alleles, Genotypes & Population Frequencies



https://sorghumbase.org/

Sorghum bicolor ssp. bicolor BTx623 (Sorghum_bicolor_NCB1v3) Location: 6:46,566,240-46,571,064 Gene: Sobic.006G095600 Variant: tmp_6_46568068_G_A

- Variant displays
 - Explore this variant
 - Genomic context
 - Genes and regulation
 - Flanking sequence
 - Genotype frequency**
 - Phenotype data
 - Sample genotypes
 - Linkage disequilibrium
 - Phylogenetic context
 - Citations
- Configure this page
- Custom tracks
- Export data
- Share this page

tmp_6_46568068_G_A SNP

Most severe consequence | Stop gained | [See all predicted consequences](#)
Alleles | G/A | Ambiguity code: R
Location | Chromosome 6:46568068 (forward strand) | [View in location tab](#)
HGVSNAMES | This variant has 5 HGVSNAMES - [Show](#)
Original source | Whole-Genome Sequence Accuracy Is Improved by Replication in a Population of EMS Mutagenized Sorghum
About this variant | This variant overlaps [2 transcripts](#) and has [486 sample genotypes](#).

Genotype frequency

Frequency data (1)

Population	Allele: frequency (count)	Genotype: frequency (count)	Genotype detail
purdueEMS	G: 0.999 (971) A: 0.001 (1)	GIG: 0.998 (485) AIG: 0.002 (1)	Show

Genotypes for purdueEMS

Sample	Genotype (forward strand)	Description
PI678262	GIA	-
PI677886	GIG	-
PI677884	GIG	-
PI677885	GIG	-
PI677886	GIG	-

Variation Summary with EMS3 + SAP



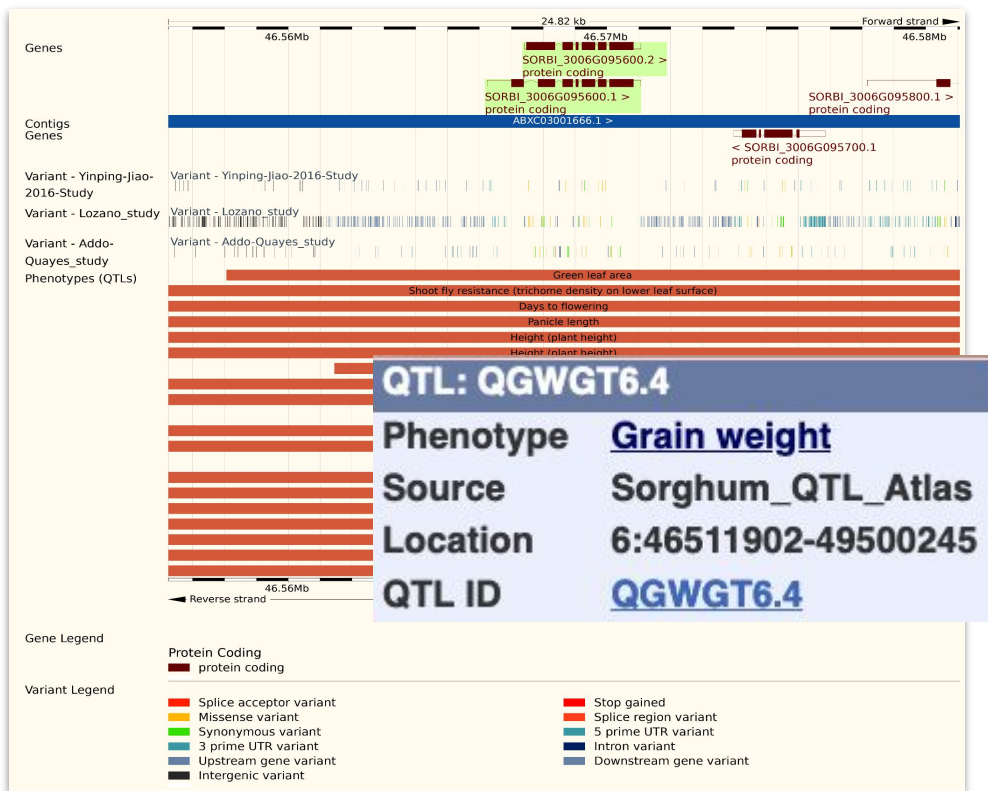
<https://sorghumbase.org/>

	EMS1	EMS2	EMS3	All EMS	NP		All
Publication	Addo-Quaye, 2018	Jiao, 2016	USDA, 2022		Lozano, 2021	Boatwright, 2022 (SAP)	
# Germplasm	486	252	890	1628	499	378	2505
# SNPs	2.55M	1.74M	8.6M	12.89M	13.1M	33.3M	59M
# Exonic SNPs	295,609	230,908	907,780	1,434,297	1,057,476	1,644,881	4,101,175
# PTV¹ SNPs	7,331	5,783	21,209	34,322	6,278	106,811	147,409
# Genes with exonic SNPs	33,334	31,376	34,487	34,686	33,463	34,539	34,997
# Genes with PTV percent²	6,177 17%	4,864 14%	14,010 40%	18,323 52%	4,641 13%	18,835 53%	27,201 77%

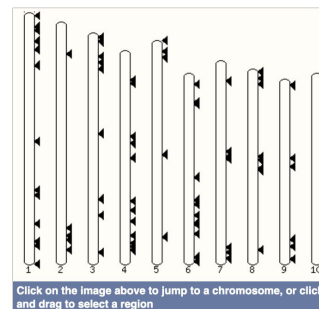
¹ PTV = Protein Truncating Variant

² Total 35,480 gene models

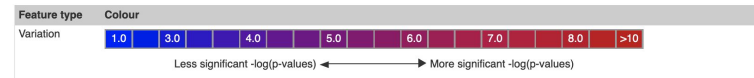
QTLs available as tracks on browser



Locations of features associated with Grain weight



Key



Features associated with phenotype Grain weight

Name(s)	Genomic location (strand)	Feature type	Reported gene(s)	Annotation source(s)	Study	P value (negative log)
QGWGT1.28	1:1-1975781(1)	QTL	-	Sorghum QTL Atlas	-	-
QGWGT1.27	1:1-2224195(1)	QTL	-	Sorghum QTL Atlas	-	-
QGWGT1.2	1:1899396-7281594(1)	QTL	-	Sorghum QTL Atlas	-	-
QGWGT1.1	1:2208844-6613450(1)	QTL	-	Sorghum QTL Atlas	-	-
QGWGT1.14	1:2219058-7378436(1)	QTL	-	Sorghum QTL Atlas	-	-
QGWGT1.18	1:2235340-16183968(1)	QTL	-	Sorghum QTL Atlas	-	-
QGWGT1.3	1:2238425-5107979(1)	QTL	-	Sorghum QTL Atlas	-	-
QGWGT1.17	1:4178848-6878476(1)	QTL	-	Sorghum QTL Atlas	-	-

Contact us

How can we help?... We would love to hear from you!

sorghumbase.org/contact

Join our [mailing list](#)



@sorghumbase

SorghumBase Exercise

Live Demo

SorghumBase Exercises

- Your Favorite Gene (YFG)
- Orthologs & paralogs => SbYFG
- Protein structure
- Gene expression
- Genetic variants
- Germplasm with deleterious alleles

Useful links

- Live SB site R5: <https://www.sorghumbase.org>
- Test SB site R6: <https://dev.sorghumbase.org>
- Examples of [agriculturally important rice & maize genes](#)
- Screenshot slides [LINK provided during the workshop]
- [Survey](#)

Template Findings

- YFG:
- SbYFG:
- Germplasm with PTV:
- Interesting findings:
- Challenges:
- Suggestions:

Template Screenshots

Protein structure URL:

Gene expression (URL):

Genetic variants (URL):