SorghumBase: Public Genetic and Genomic Database for the Sorghum Community

Nicholas Gladman
ASPB, August 2023
SorghumBase is a Community Knowledge-base

- New database resource!
- Genetic, Genomic, Variation, Expression, and Metabolic Data and Visualizations
- Our stakeholders are geneticists, genomicists, molecular biologists, and breeders
- Working toward breeding resource development
- FAIR
- Release 6 (June 2023)

Scientific Advisory Board

Laura Mayor  
Mitchell Tuinstra  
Stephen Kresovich

Cite SorghumBase: Gladman et al. (2022)  
10.1007/s00425-022-03821-6
Search Views Are Completely Integrated
SbPho2, Phosphorus Starvation Signaling Gene, Identified in Sorghum and Shown to be Involved in Regulating the Uptake of Multiple Nutrients

Sorghum mutant, sbpho2, exhibits leaf necrosis and short roots, and is likely involved in phosphate accumulation and root architecture alteration.

Monday, June 5, 2023

Read more
Search for Publications, Gene, Germplasm & Ontologies
SorghumBase Contains EMS-induced and Natural Variant Resources
Variant Image: Find Disruptive SNPs in Gene Model
Variant Table: Filterable View for SNP

Gene: SORBI_3003G385900

Description: hypothetical protein
Location: Chromosome 3: 69,827,192-69,830,766 reverse strand.

About this gene: This gene has 1 transcript (splice variant), 39 orthologues and 11 paralogues.

Variant table

This table shows known variants for this gene. Use the 'Consequence Type' filter to view a subset of these.

Filter

Consequence: Splice acceptor variation (5/20)
Filter Other Columns

<table>
<thead>
<tr>
<th>Variant ID</th>
<th>Chr: bp</th>
<th>Alleles</th>
<th>Class</th>
<th>Source</th>
<th>Conseq. Type</th>
<th>AA</th>
<th>AA coord</th>
<th>Transcript</th>
</tr>
</thead>
<tbody>
<tr>
<td>tnp_3_69827957_C_T</td>
<td>3,698,279,57</td>
<td>C/T</td>
<td>SNP</td>
<td>USDA_Lubbock_EMS</td>
<td>Stop gained</td>
<td>W*</td>
<td>677</td>
<td>SORBI_3003G385900.1</td>
</tr>
<tr>
<td>tnp_3_69827964_G_A</td>
<td>3,698,279,64</td>
<td>G/A</td>
<td>SNP</td>
<td>USDA_Lubbock_EMS</td>
<td>Stop gained</td>
<td>Q*</td>
<td>675</td>
<td>SORBI_3003G385900.1</td>
</tr>
<tr>
<td>tnp_3_69828296_G_C</td>
<td>3,698,282,96</td>
<td>G/C</td>
<td>SNP</td>
<td>SAP_Bowarrow_t_study</td>
<td>Stop gained</td>
<td>Y*</td>
<td>591</td>
<td>SORBI_3003G385900.1</td>
</tr>
<tr>
<td>tnp_3_69828503_C_T</td>
<td>3,698,285,03</td>
<td>C/T</td>
<td>SNP</td>
<td>USDA_Lubbock_EMS</td>
<td>Stop gained</td>
<td>W*</td>
<td>522</td>
<td>SORBI_3003G385900.1</td>
</tr>
<tr>
<td>tnp_3_69829703_G_A</td>
<td>3,698,297,03</td>
<td>G/A</td>
<td>SNP</td>
<td>USDA_Lubbock_EMS</td>
<td>Stop gained</td>
<td>Q*</td>
<td>196</td>
<td>SORBI_3003G385900.1</td>
</tr>
<tr>
<td>tnp_3_69830063_C_T</td>
<td>3,698,300,63</td>
<td>C/T</td>
<td>SNP</td>
<td>USDA_Lubbock_EMS</td>
<td>Stop gained</td>
<td>W*</td>
<td>107</td>
<td>SORBI_3003G385900.1</td>
</tr>
<tr>
<td>tnp_3_69830123_C_T</td>
<td>3,698,301,23</td>
<td>C/T</td>
<td>SNP</td>
<td>USDA_Lubbock_EMS</td>
<td>Stop gained</td>
<td>W*</td>
<td>87</td>
<td>SORBI_3003G385900.1</td>
</tr>
<tr>
<td>tnp_3_69827422_G_C</td>
<td>3,698,274,22</td>
<td>G/C</td>
<td>SNP</td>
<td>Lozano_study</td>
<td>Missense variant</td>
<td>K/N</td>
<td>855</td>
<td>SORBI_3003G385900.1</td>
</tr>
<tr>
<td>tnp_3_69827469_C_T</td>
<td>3,698,274,69</td>
<td>C/T</td>
<td>SNP</td>
<td>USDA_Lubbock_EMS</td>
<td>Missense variant</td>
<td>E/K</td>
<td>840</td>
<td>SORBI_3003G385900.1</td>
</tr>
<tr>
<td>tnp_3_69827478_G_A</td>
<td>3,698,274,78</td>
<td>G/A</td>
<td>SNP</td>
<td>Addo-Quay's_study</td>
<td>Missense variant</td>
<td>R/C</td>
<td>837</td>
<td>SORBI_3003G385900.1</td>
</tr>
<tr>
<td>tnp_3_69827492_G_A</td>
<td>3,698,274,92</td>
<td>G/A</td>
<td>SNP</td>
<td>Addo-Quay's_study</td>
<td>Missense variant</td>
<td>P/L</td>
<td>832</td>
<td>SORBI_3003G385900.1</td>
</tr>
</tbody>
</table>
Coordinating With EVA For Stable Variant IDs

- Collect and centralize data, easy to find, share, organize, maintain (FAIR)
- Reduce redundancy, stable unique identifier for each unique genetic locus
- Crosslink to other data archival resources
- NCBI dbSNP & EVA provided ‘ss’ & ‘rs’ identifiers for human and non-human species

<table>
<thead>
<tr>
<th>Identifier type</th>
<th>Example</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ssID</td>
<td>ss335</td>
<td>Submitted SNP ID assigned by dbSNP or EVA.</td>
</tr>
<tr>
<td>rsID</td>
<td>rs334</td>
<td>Reference SNP ID assigned by dbSNP or EVA. ssIDs of the same variant type that colocalise are combined to give an rsID for that locus.</td>
</tr>
</tbody>
</table>
Sorghum Community Mapping Panel Working Group

- Develop a domestic provider for genotyping US sorghum germplasm
- We have finished probe design
- Currently in validation phase (breeding material, populations, and landraces)
- Anticipate creation of PlexSeq array in ~6 months (AgriPlex)
- Final array marker selection based on feedback from breeders with most useful markers for US breeding germplasm
SorghumBase Has Community Curation Functions

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

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

Search Filters
- GENE TREE | HOMOLOGS
- GENE TREE | ORTHOLOGS
- GENE TREE | PARALOGS

Links to other resources:
- Ensembl Gene Tree view
- Curate
This is the gene tree containing SORBI_3001G394400. Description: similar to Phytochrome B.

Curation Instructions

Mark genes as okay or flag genes that you think might have issues and choose a reason from the drop down menu. When finished, enter your email address and submit.

The reason for flagging a gene model is optional, but you may indicate whether there is a gain (G), loss (L) or no change (.) in the beginning, middle, and end of the gene model.

For example L,.,, means there is a loss at the 5’ end.
Supporting Breeding and Phenotype Data

- Evaluating use of BreedBase to host sorghum data
- Web-based portal for sorghum breeders, researchers, and stakeholders
- Integrates phenotypic data management and breeding tools

**Phenotype and Breeding datasets:**
Pedigrees, Field Trials, Phenotypes, Genotyping Trials, Germplasm Accessions and Lots, Image/Spectral Datasets, etc.

**Analysis Workflows:** GWAS, Genomic Prediction, Linear Mixed Modeling, Population Structure (PCA), and Kinship and Clustering
Future Plans for 2023-2024

- 3-4 releases per year
  - Extend current data types while adding new functionality
- Stable identifiers for genetic variation (rsIDs)
- Direct access to GRIN germplasm genotypes
- Continued evaluation of BreedBase to support management of phenotype and breeding data
- Finish community mapping panel and coordinate with working group towards phenotype data standards
Community Engagement Is Our Priority!

contact@sorghumbase.org

- Mailing list
  - Visit www.sorghumbase.org to join!
- Events ([https://www.sorghumbase.org/events](https://www.sorghumbase.org/events))
- Training guides and videos ([https://www.sorghumbase.org/guides](https://www.sorghumbase.org/guides))
- **1-on-1 training is always available**
- Training opportunities
  - Navigating and using website
  - Community curation of gene structure
  - Metadata and sample submission
  - Content management system
- Working Groups
  - Community Mapping Panel, Reference Genomes
  - Phenotype (future working group)
SorghumBase Team

Nick Gladman  
Leadership

Doreen Ware  
Leadership

Kapeel Chougule  
Genome Annotation & Data Resources

Marcela K. Tello-Ruiz  
Communication & Data Resources

Andrew Olson  
Architecture & Infrastructure

Sunita Kumari  
Communication & Data Resources

Zhenyuan Lu  
Workflows & Data Resources

Audra Olson  
Communication

Vivek Kumar  
Infrastructure & Data Resources

Peter Van Buren  
Systems Engineer

Sharon Wei  
Infrastructure & Data Resources
Acknowledgements

Ware Lab 2023

Presence @ ASPB 2023:

- Bioinformatics Workshop - Organizer: Sunita Kumari
  - Gramene by Sunita Kumari
  - SorghumBase by Nick Gladman
- Posters:
  - #500-08: Gramene
  - #700-09: SorghumBase
  - #800-56: DAP-seq - Audrey Fahey
  - #500-24: NUE GRN - Janeen Braynen
  - #1100-19: QPSI - Sunita Kumari
  - #800-64: GRAS TFs - Nick Gladman

We gratefully acknowledge support from USDA-ARS-8062-21000-041-00D