



# SorghumBase 2024: Building Partnerships and Integrating Genetic Knowledge for the Sorghum Community

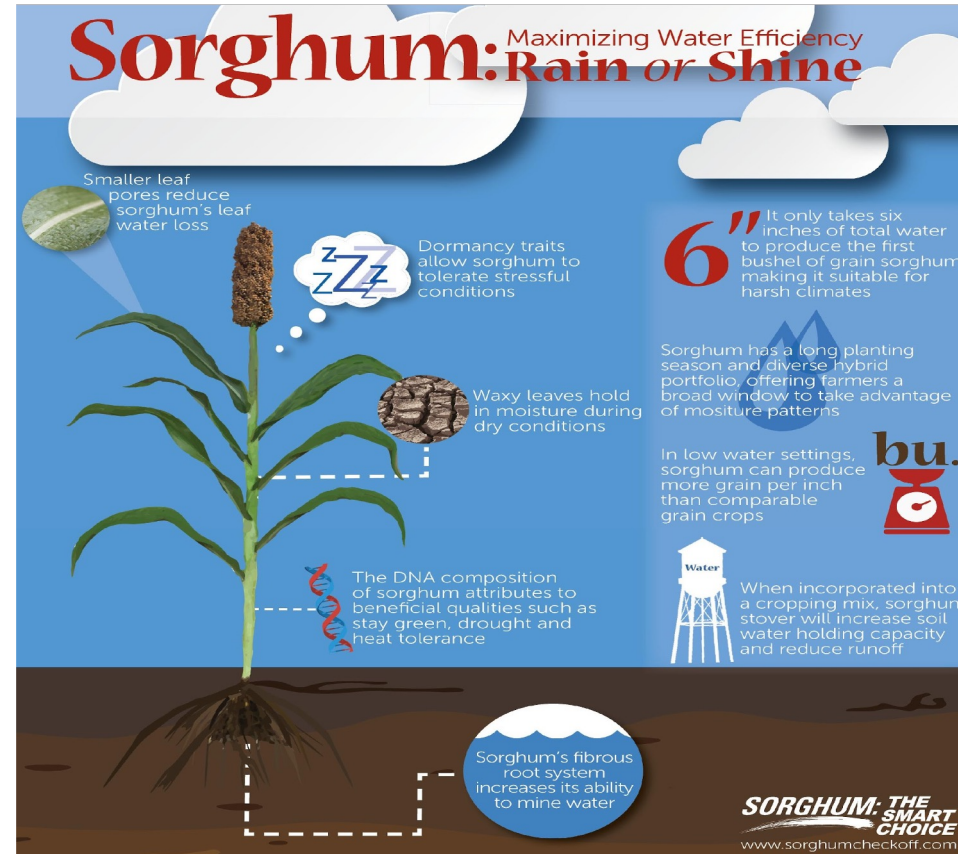
**Sharon Wei**  
**ASPB, June 2024**



Follow us  
[@SorghumBase](https://twitter.com/SorghumBase)

# Why Sorghum?

- ❖ Fifth largest crop worldwide
  - grain, forage, sugar crop (wine)
- ❖ 2<sup>nd</sup> largest source of biofuel in US
- ❖ Climate resilience
  - The camel of the crop
    - Drought resistant
    - Heat resistant
    - Soil adaptability
  - Efficient in conversion of solar energy and use of water
  - Sustainable agriculture
- ❖ Model organism for C4 photosynthesis





# Welcome To Sorghumbase

Genomic Resources for the Sorghum Research Community

## SorghumBase is a USDA funded Community Knowledgebase



[Genetic Variation](#)

[Explore the Impact of Variants c](#)



[News](#)

[Relevant developments in Sorghu](#)

> [Planta](#). 2022 Jan 11;255(2):35. doi: 10.1007/s00425-022-03821-6.

[Genomes](#)

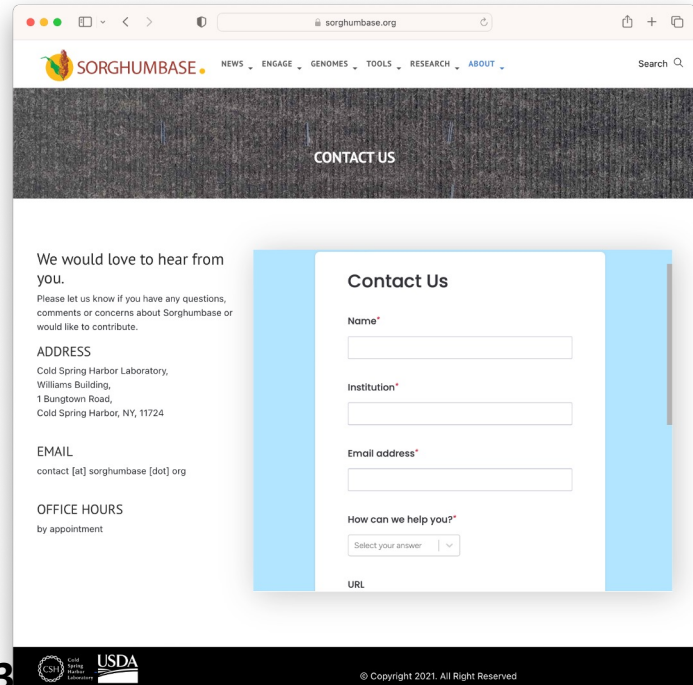
## SorghumBase: a web-based portal for sorghum genetic information and community advancement

Nicholas Gladman<sup>1</sup>, Andrew Olson<sup>1</sup>, Sharon Wei<sup>1</sup>, Kapeel Chougule<sup>1</sup>, Zhenyuan Lu<sup>1</sup>, Marcela Tello-Ruiz<sup>1</sup>, Ivar Meijs<sup>1</sup>, Peter Van Buren<sup>1</sup>, Yinping Jiao<sup>2</sup>, Bo Wang<sup>1</sup>, Vivek Kumar<sup>1</sup>, Sunita Kumari<sup>1</sup>, Lifang Zhang<sup>1</sup>, John Burke<sup>3</sup>, Junping Chen<sup>3</sup>, Gloria Burow<sup>3</sup>, Chad Hayes<sup>3</sup>,

## ❖ Resource interoperability / unification

## ❖ Community outreach

- News - 120 posts in the last 12 months
- Events - 153 events last year + 16 upcoming
- Publications - 953 papers
- Research Highlights - 87 in last 12 months
- Funded Projects - 289
- Training - User guides, YouTube vignettes\*
- CSI seminars
- Workshops - In-person bootcamp December, 2023
- Custom Workflows
- Contact - [sorghumbase.org/contact](https://sorghumbase.org/contact)



## Scientific Advisory Board



LAURA MAYOR  
Sorghum Breeding Lead  
Corteva Agriscience

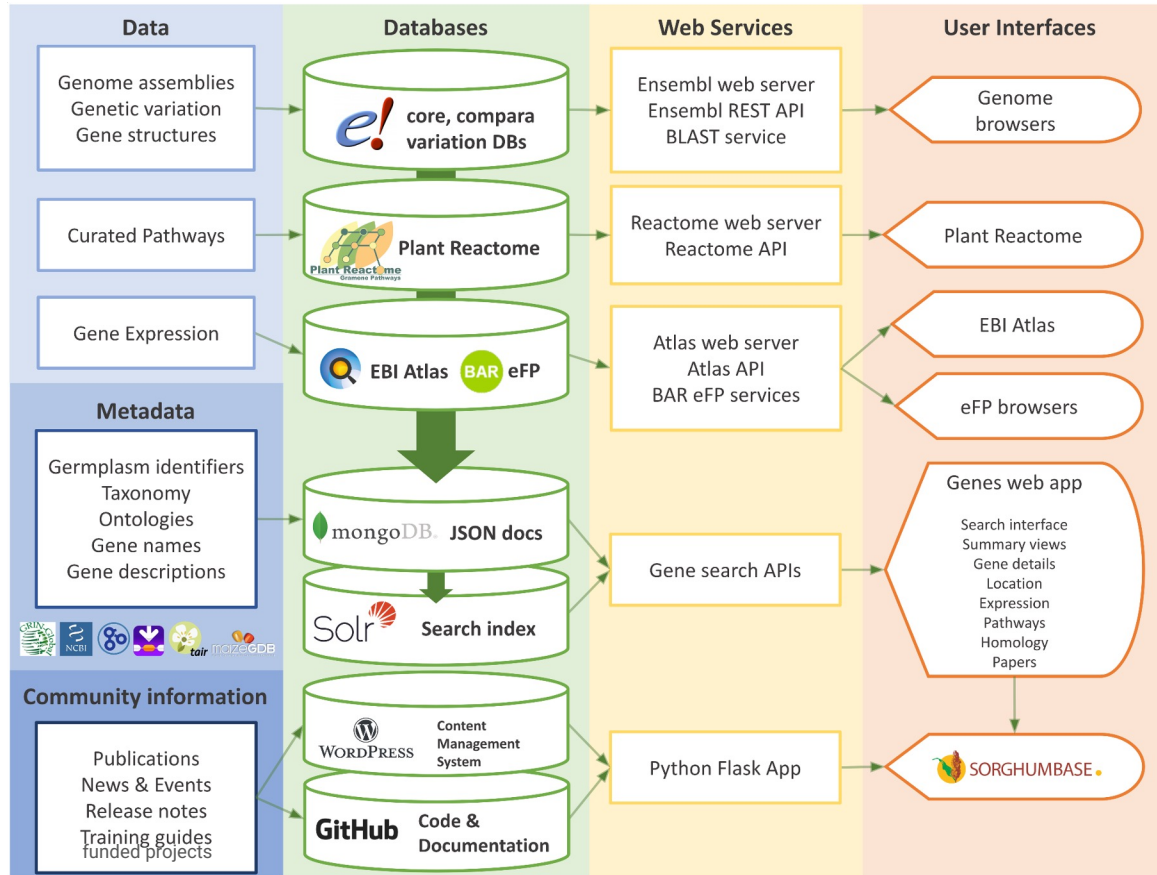


MITCHELL TUINSTRAS  
Purdue University



STEPHEN KRESOVICH  
Clemson University

# Sorghumbase Infrastructure Diagram



# Curated Genomic Input Data

**Data**

- Genome assemblies
- Genetic variation
- Gene structures

**Curated Pathways**

**Gene Expression**

**Metadata**

- Germplasm identifiers
- Taxonomy
- Ontologies
- Gene names
- Gene descriptions

**Community information**

- Publications
- News & Events
- Release notes
- Training guides

	Natural Variation	EMS Populations	All
Publications		Jiao, 2023	All EMS
# Accessions		28	2505
# Studies			59M

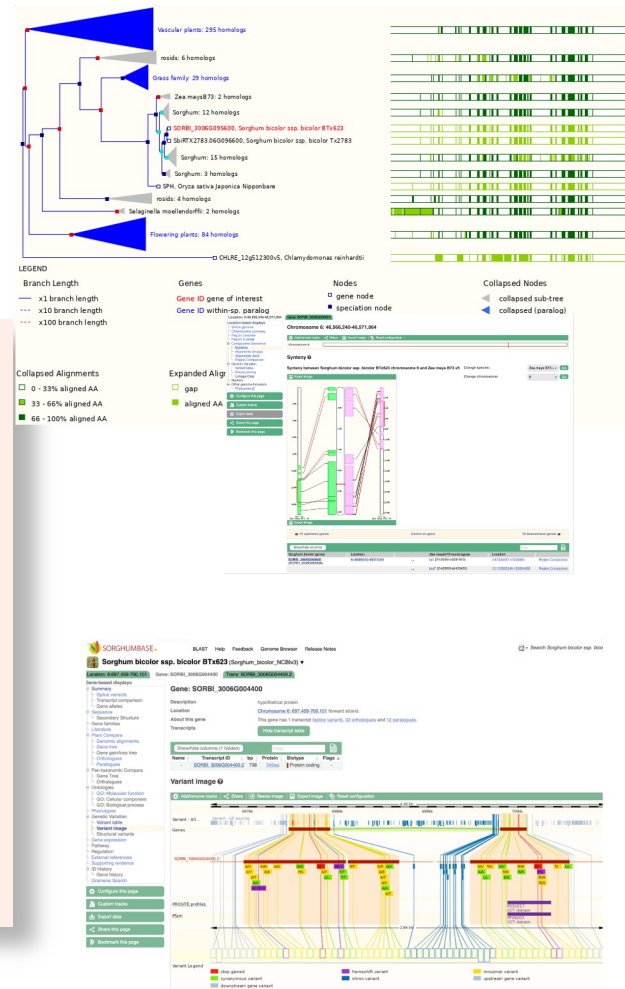
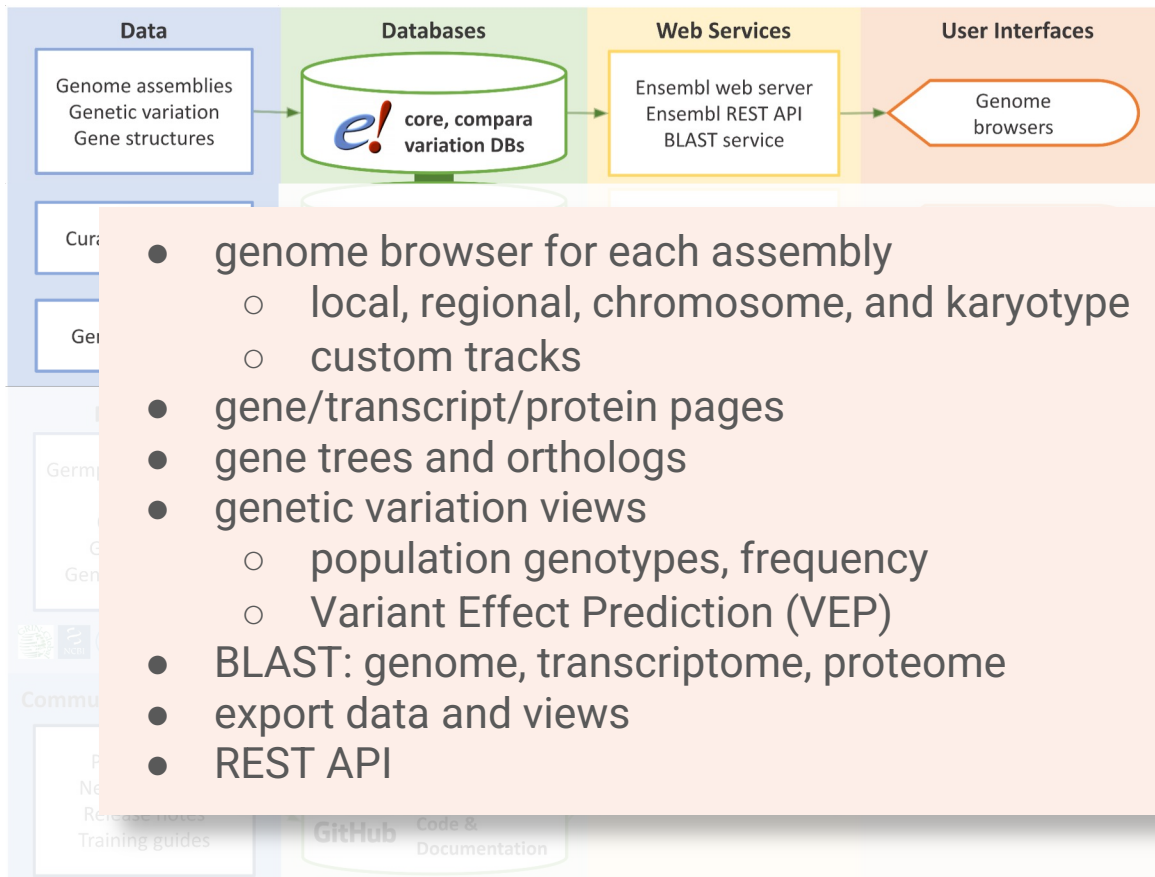
## Gene Expression

The EBI Expression Atlas: 8 baseline expression studies, 3 differential expression studies available on the site. 10 in the pipeline to be release in next cycle.

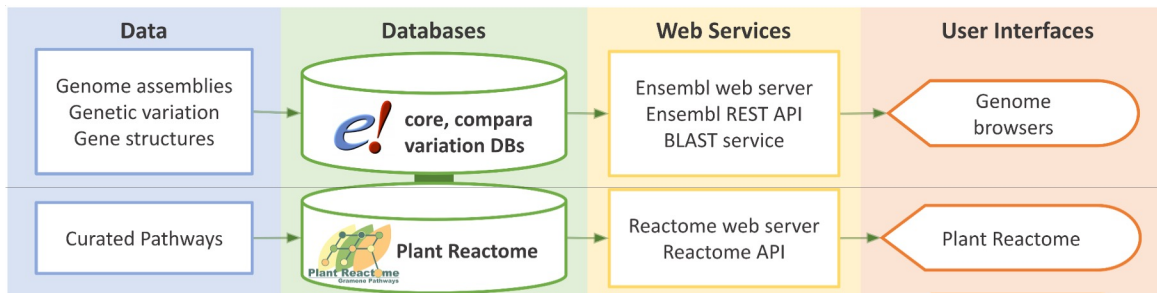
The BAR: we curated 6 studies and generated pictographs for eFP browser. 10 new studies to be processed for next release.

<i>S. bicolor ssp. bicolor</i>	AustrCF317961			CNS0314256	Tao et al (2021)
<i>S. bicolor ssp. drummondii</i>	S369-1, AGG324436			CNS0314264	Tao et al (2021)
<i>S. bicolor ssp. verticillifera</i>					Tao et al (2021)
<i>S. propinquum</i>					Tao et al (2021)

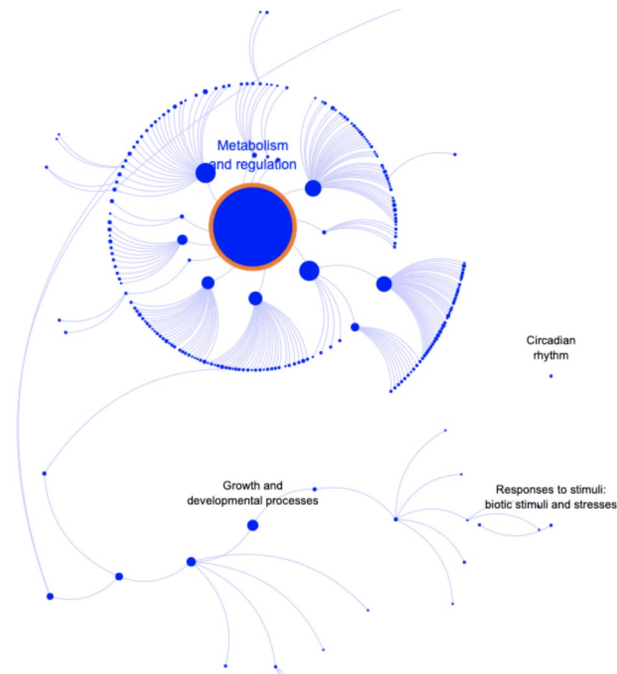
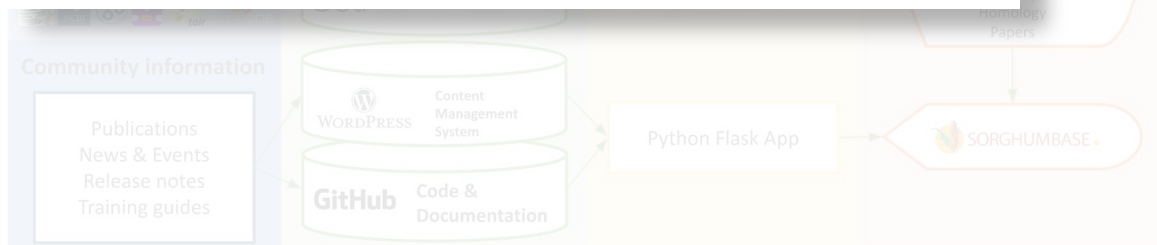
# e!Genome Browser



# Pathway tools



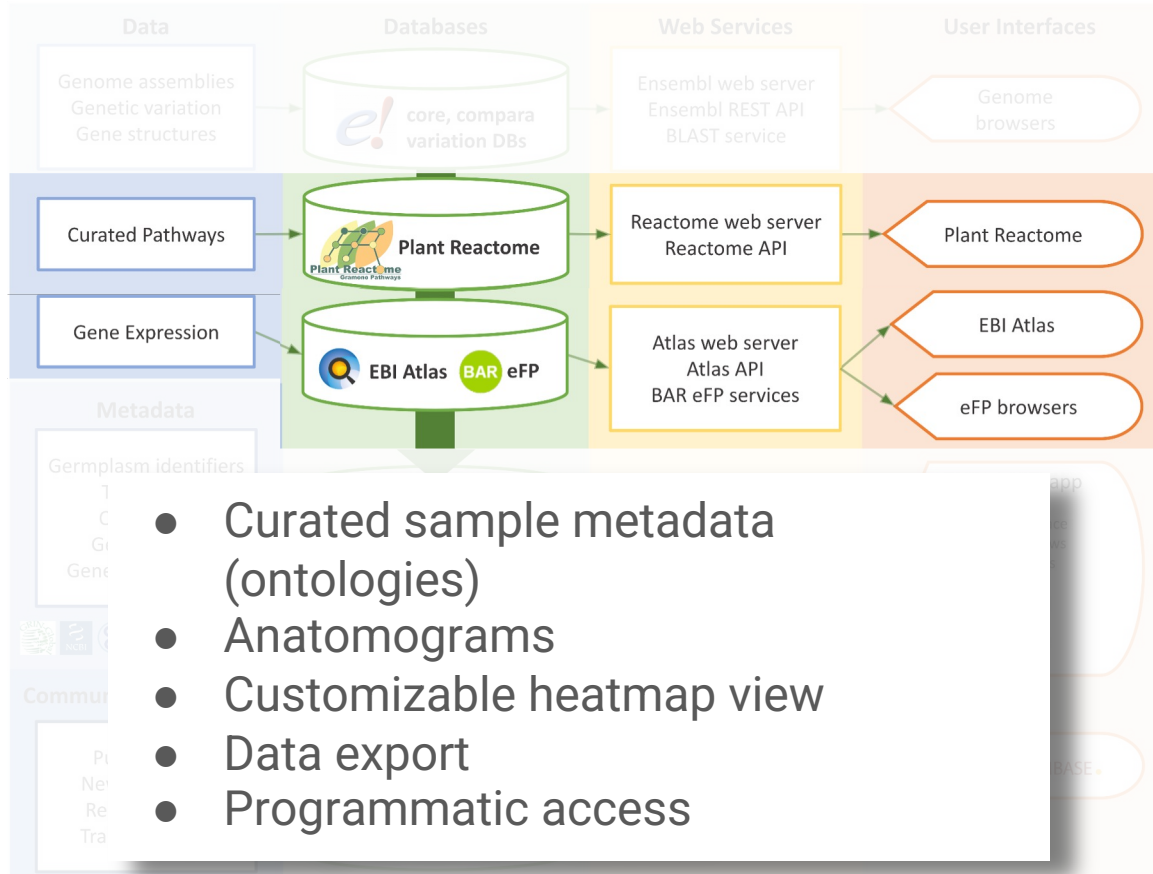
- Fireworks view (hierarchy)
- Pathway diagram
- Access gene expression data
- Integrated interaction data
- Programmatic access via API



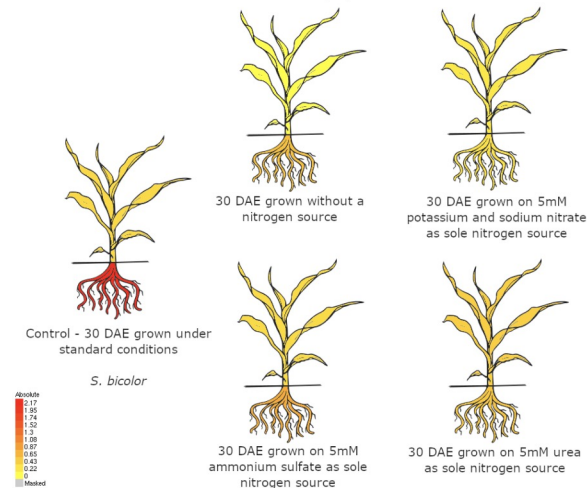
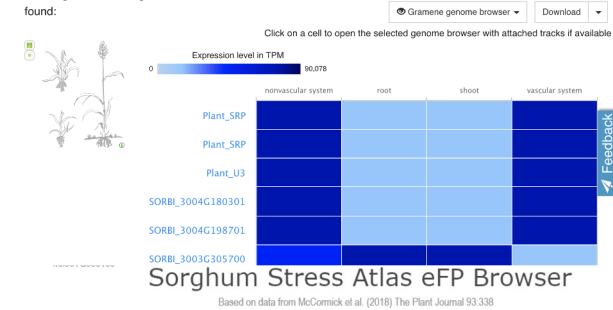


# Gene Expression

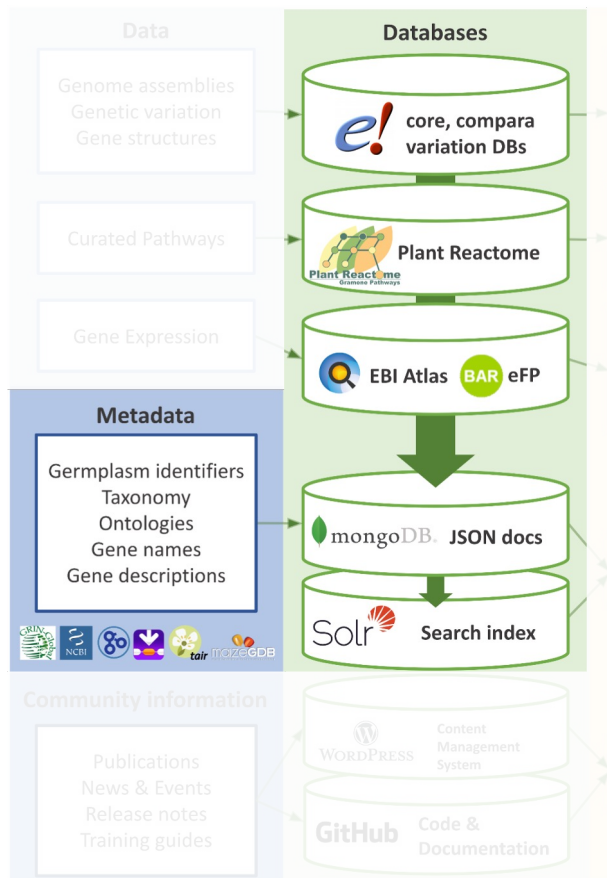
BAR



Showing 50 of 24,388 genes found:



# Building the Search Index

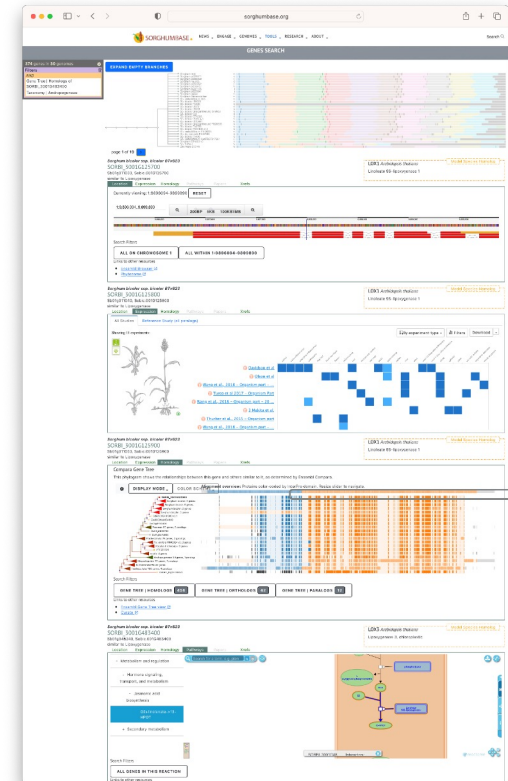
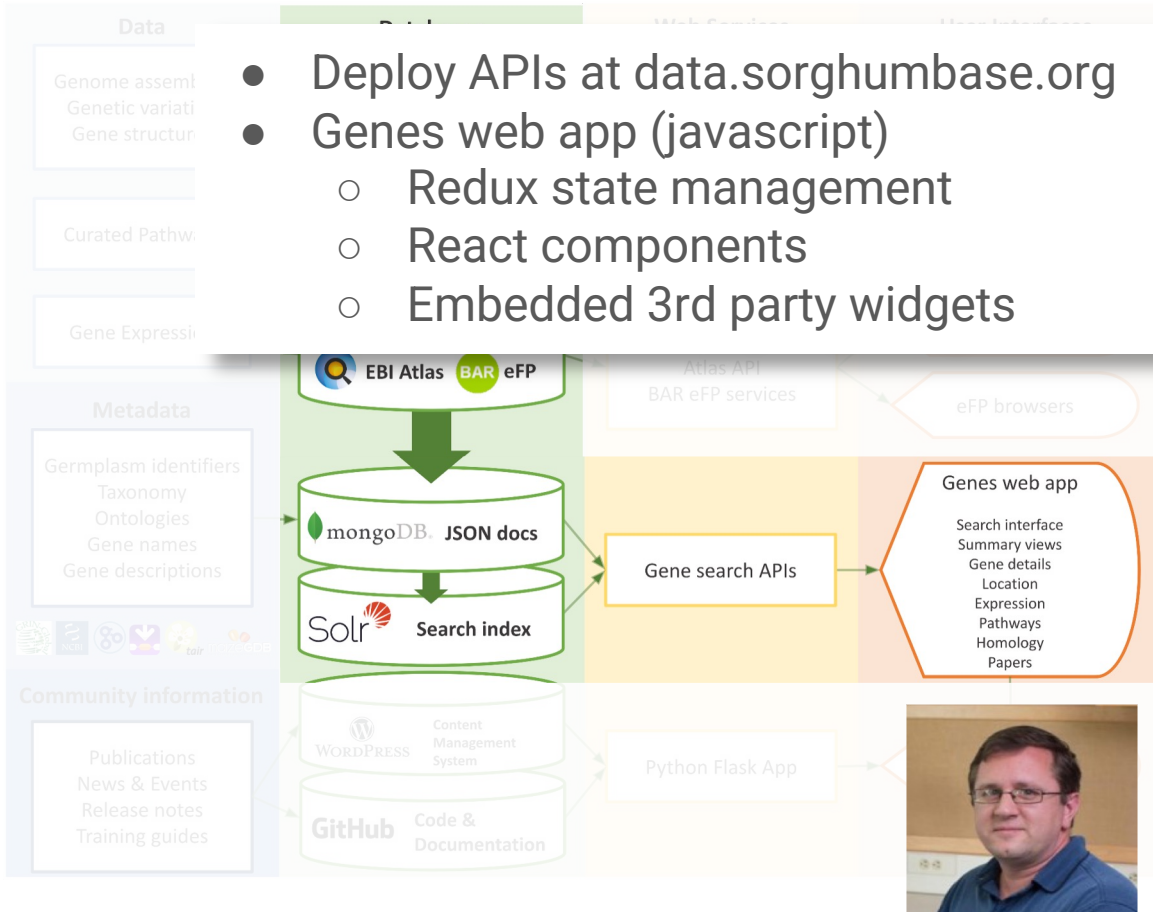


- Extract genes, gene trees, pathways, expression levels and domain annotations from primary databases
- Extend annotations with synonyms, ontologies and hierarchical metadata
  - GO, PO, TO, taxonomy, etc.
  - Interpro domain hierarchy
  - Pathway hierarchy
- Identifiers and common names for genes and germplasm
- Load into mongoDB collections
- Build Solr indexes
  - genes
  - suggestions (search terms)



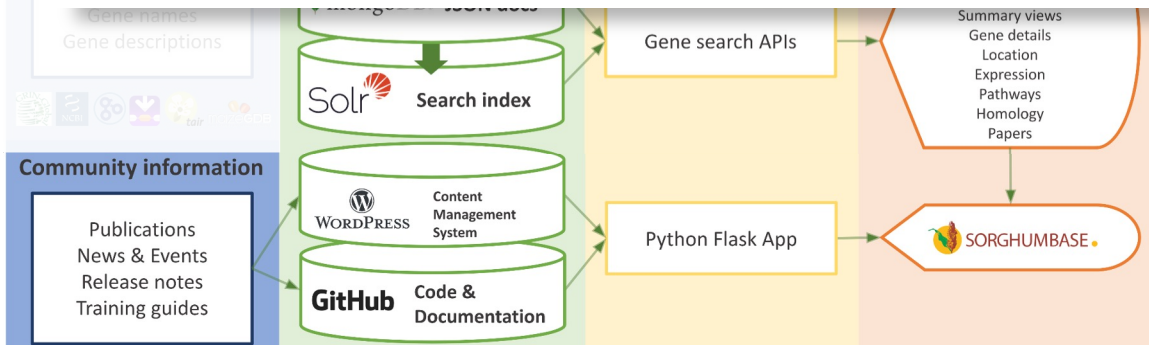
# Integrated Search

- Deploy APIs at data.sorghumbase.org
- Genes web app (javascript)
  - Redux state management
  - React components
  - Embedded 3rd party widgets



# Community info

- WordPress content management system
  - Publications
  - Research Highlights
  - News and Events
  - Funded Projects
- GitHub: code, release notes, and documentation
- Access through sorghumbase.org
- Bulk data access at ftp.sorghumbase.org



# Search Views Are Completely Integrated

- ❖ many ways to find genes of interest
- ❖ not all genes are well annotated - homology helps
- ❖ not all gene models are okay - curate
- ❖ new integrated views based on user feedback
  - Sequences tab - quick access to DNA, RNAs, proteins
  - Germplasm with protein truncating variant (PTV) alleles

Short Root

Genes 404 terms Website 4 matches

Plant Ontology: anatomy

ROOT APICAL MERISTEM (6) LATERAL ROOT (37) LATERAL ROOT PRIMORDIUM (31) ROOT PRIMORDIUM (23) PRIMARY ROOT TIP (3) PRIMARY ROOT (31) LATERAL ROOT TIP (9) ROOT CAP (6) ROOT TIP (64) ROOT ELONGATION ZONE (9)

InterPro: Family

ROOT\_CAP (280) GLV2 (86) RPD1-LIKE (31) NPH3-LIKE (41) NAKR1/2/3 (163) SHORT-ROOT (6) RUS\_FAM (216) NAKR2/3 (5) BPS1\_PLN (976) RGF1/2/3 (24)

Trait ontology

ADVENTITIOUS ROOT LENGTH (8) ROOT WEIGHT (16) ROOT THICKNESS (18) ROOT HAIR LENGTH (18) ADVENTITIOUS ROOT NUMBER (8) ROOT VOLUME (1) ROOT NUMBER (29) ROOT SHAPE (14) ROOT DRY WEIGHT (3) ROOT LENGTH (162)

Gene Ontology: biological process

LATERAL ROOT MORPHOGENESIS (18) POST-EMBRYONIC ROOT MORPHOGENESIS (16) LATERAL ROOT DEVELOPMENT (208) LATERAL ROOT FORMATION (14) TRICHOBLAST DIFFERENTIATION (29) ROOT MERISTEM GROWTH (15) ATRICHOBLAST DIFFERENTIATION (3) ROOT CAP DEVELOPMENT (22) BASIPETAL AUXIN TRANSPORT (16) ACROPETAL AUXIN TRANSPORT (6)

News  
[Relevant developments in Scirpium research.](#)

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

Events  
[Upcoming Meetings and Conferences](#)



### [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](#)



# Streamlining Data Integration Through Standards



## AgBioData



**Findable**



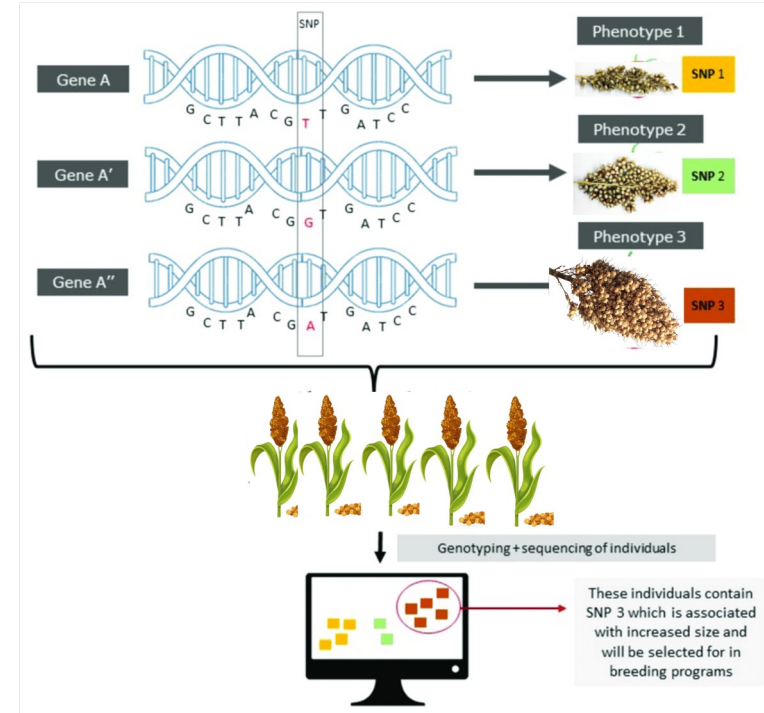
**Accessible**



**Interoperable**



**Reusable**



# Standards for Genetic Variation WG

## Coordinate With EVA For

### Stable Variant IDs



**AgBioData**

Toward enhanced genomics, genetic  
outcomes through standardization

EMBL-EBI



- 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

Identifier type	Example	Description
-----------------	---------	-------------

ssID	ss335	Submitted SNP ID assigned by dbSNP or EVA.
------	-------	--

rsID	rs334	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.
------	-------	--

# Sorghumbase 2023-2024 Highlights

## ❖ Data

- **41M SNPs assigned rsIDs (EVA release5)**
- **234 GWAS traits from meta-analysis of 25 studies (SAP population)**
- Non-coding feature tracks for all genomes (tRNA, miRNA, snoRNA ...)
- Sorghum Literature: 214 papers, 30 highlights, 289 funded projects
- expression data sets from 11 studies were curated


## ❖ Tools and functions

- eFP expression browser
- Sequence tab
- **Germplasm tab serves Protein Truncation mutant lines from GRIN Global And SorbMutDB**
- Papers tab for community curation




## Phenotype data

# Genotype => Phenotypes


**SORGHUMBASE**

[BLAST](#)
[Help](#)
[Feedback](#)
[Genome Browser](#)
[Release Notes](#)


**Sorghum bicolor ssp. bicolor BTx623** (Sorghum\_bicolor\_NCBIV3) ▼

Location: 2:57,664,706-57,664,706 Variant: rs872512714

Variant displays
 

- Explore this variant
  - Genomic context
  - Genes and regulation
  - Flanking sequence
- Genotype frequency
- Phenotype data**
- Sample genotypes
- Linkage disequilibrium
- Phylogenetic context
- Citations
- 3D Protein model

rs872512714 SNP

Most severe consequence  
 Alleles  
 Location  
 HGVS names  
 Synonyms

External Links  
 Original source  
 About this variant

missense variant | [See all predicted consequences](#)

**A/T** | Highest population MAF: **0.63**

**Chromosome 2:57664706** (forward strand) | VCF: 2 57664706 rs872512714 A T

This variant has 3 HGVS names - [Show](#)

SAP\_Boatwright\_study tmp\_2\_57664706\_A\_T

[Short variant data imported from EVA](#) (release 5)

This variant overlaps 1 [transcript](#), has 400 [sample genotypes](#) and is associated with 7 [phenotypes](#)

**Phenotype Data**

**Significant association(s)**

Show All entries

Show/hide columns (1 hidden)


Phenotype, disease and trait	Source(s)	External reference	Reported g
<b>Fat</b>	MURAL_META_GWAS_MLM	<a href="#">Rhodes et al. 2017</a>	SORBI_3002G190300
<b>Protein</b>	MURAL_META_GWAS_MLM	<a href="#">Rhodes et al. 2017</a>	SORBI_3002G190300
<a href="#">Protein+Fat</a>	MURAL_META_GWAS_MLM	<a href="#">Rhodes et al. 2017</a>	SORBI_3002G190300
<a href="#">Starch+Protein+Fat</a>	MURAL_META_GWAS_MLM	<a href="#">Rhodes et al. 2017</a>	SORBI_3002G190300
<a href="#">Starch+(Protein+Fat)</a>	MURAL_META_GWAS_MLM	<a href="#">Rhodes et al. 2017</a>	SORBI_3002G190300
<a href="#">Starch+Fat</a>	MURAL_META_GWAS_MLM	<a href="#">Rhodes et al. 2017</a>	SORBI_3002G190300
<a href="#">Starch+Protein</a>	MURAL_META_GWAS_MLM	<a href="#">Rhodes et al. 2017</a>	SORBI_3002G190300

**Genes in this region**

The following gene(s) in the region of this variant might have associated phenotype data:

Show/hide columns

Gene	HGNC name	Position
<a href="#">SORBI_3002G190300</a>	-	Over


**SorghumBASE**


BLAST

Help

Feedback

Genome Browser

Repeats


**Sorghum bicolor ssp. bicolor BTx623** (Sorghum\_bicolor\_NCB1v3)

Phenotype: Protein

Location: 2:57,684,706-57,664,706

Variant: rs72512714

Phenotype-based displays

Associated loci

Related conditions

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Loci associated with Protein

Filter

Feature type: All

Annotation source: All

Name(s)	Type	Genomic
ns73693523	Variant	1:314297
ns542134780	Variant	2:575692
ns5421528892	Variant	2:575693
ns5421590975	Variant	2:576077
ns162660399	Variant	2:576078
ns5421275688	Variant	2:576282
ns268635906	Variant	2:576282
ns266788758	Variant	2:576334
ns872411270	Variant	2:576335
ns872411270	Variant	2:576335
ns873138595	Variant	2:576335
ns267048224	Variant	2:576335
ns266386605	Variant	2:576346
ns266318394	Variant	2:576346
ns5421475604	Variant	2:576519
ns875367578	Variant	2:576519
ns5421510298	Variant	2:576520
ns5421495708	Variant	2:576520
ns871056070	Variant	2:576520
ns5421581394	Variant	2:576628
ns873057973	Variant	2:576646
ns266985672	Variant	2:576646
ns872512714	Variant	2:576647
ns873209183	Variant	2:576647
ns871203478	Variant	2:576647
ns162413204	Variant	2:576647

## Related conditions for Protein Q0

**Ontology term**

TO:0000490 - seed protein composition trait

**Selected term:**

TO:0000598 - protein content

**Term**

protein content

**Accession**

TO:0000598 (TO)

**Definition**

**Filter**

Mapped ontology term: All

Filter Other Columns

Show/hide columns

Search...

Phenotype/Disease/Trait description	Mapped ontology Term	Relationship with TO:0000598	Variant	QTL
Protein	protein content	■ ■ ■	32	
Protein content	protein content	■ ■ ■	-	51
Leaf crude protein	protein content	■ ■ ■	-	1
Protein (% dry basis)	protein content	■ ■ ■	7	-
Protein digestibility	protein content	■ ■ ■	-	2

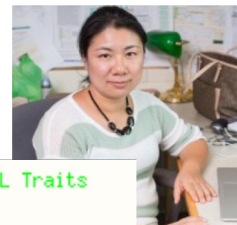
■ ■ ■ : Equivalent to the ontology term "TO:0000598"  
■ ■ : Child term of "TO:0000598"  
■ : Term related to "TO:0000598"

73 (+)	-	MURAL_META_GWAS_MLM	Rhodes et al. 2017
31 (+)	-	MURAL_META_GWAS_MashR	Rhodes et al. 2017
31 (+)	-	MURAL_META_GWAS_MLM	Rhodes et al. 2017
47 (+)	-	MURAL_META_GWAS_MLM	Rhodes et al. 2017
60 (+)	-	MURAL_META_GWAS_MLM	Rhodes et al. 2017
65 (+)	-	MURAL_META_GWAS_MLM	Rhodes et al. 2017
68 (+)	-	MURAL_META_GWAS_MLM	Rhodes et al. 2017
22-57651958 (+)	-	MURAL_META_GWAS_MLM	Rhodes et al. 2017
48 (+)	-	MURAL_META_GWAS_MLM	Rhodes et al. 2017
108 (+)	-	MURAL_META_GWAS_MLM	Rhodes et al. 2017
150 (+)	-	MURAL_META_GWAS_MLM	Rhodes et al. 2017
169 (+)	-	MURAL_META_GWAS_MLM	Rhodes et al. 2017
126 (+)	-	MURAL_META_GWAS_MLM	Rhodes et al. 2017
173 (+)	-	MURAL_META_GWAS_MLM	Rhodes et al. 2017
87 (+)	SORBI_3002G190300	MURAL_META_GWAS_MLM	Rhodes et al. 2017
106 (+)	SORBI_3002G190300	MURAL_META_GWAS_MLM	Rhodes et al. 2017
29 (+)	SORBI_3002G190300	MURAL_META_GWAS_MLM	Rhodes et al. 2017
33 (+)	SORBI_3002G190300	MURAL_META_GWAS_MLM	Rhodes et al. 2017
186 (+)	SORBI_3002G190300	MURAL_META_GWAS_MLM	Rhodes et al. 2017

## GWAS Data from Mural *et al* (2021)

# Phenotype => Genotypes

# Associated Phenotypes in Karyotype View



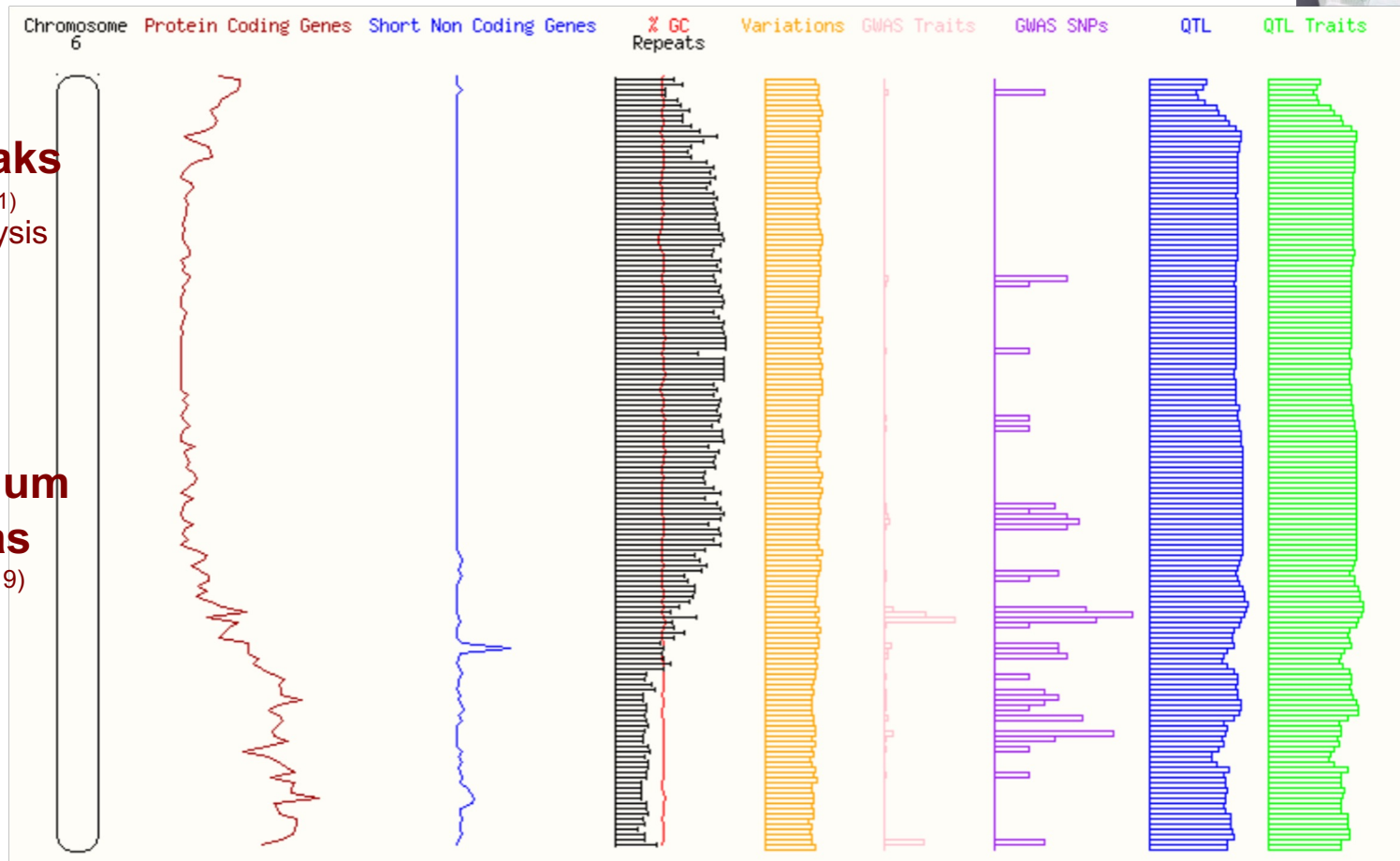
## GWAs peaks

(Mural *et al*, 2021)  
Meta-analysis  
234 traits  
400 SAP

## The Sorghum

### QTL Atlas

(Mace *et al*, 2019)



# Germplasm Tab

## NEW Search Feature

- Integrated LOF alleles
- Links to obtain germplasm
- Find other impacted genes

Use Case: Find EMS lines disrupting my gene

*Sorghum bicolor ssp. bicolor BTx623*

SORBI\_3006G095600

Sb06g018040, Sobic.006G095600

similar to Lipoxygenase

Closest Annotated Homolog  
*Os. Japonica Nipponbare*  
SPH  
86% identity

Germplasm Sequences Location Expression Homology Pathways Add Papers Xrefs

Predicted loss-of-function alleles were detected in these germplasm.

Explore other variants within this gene in the [Variant image](#) page in the Ensembl genome browser.

Order Germplasm	Synonyms	Study/Population	VEP Consequence	Allele Status	All LOF Genes
<a href="#">PI 678262</a>	PI678262	Purdue EMS	stop gained	heterozygous	<a href="#">SEARCH</a>
<a href="#">PI 701663</a>	ARS105	USDA Lubbock EMS	splice acceptor variant	heterozygous	<a href="#">SEARCH</a>
<a href="#">PI 701655</a>	ARS97	USDA Lubbock EMS	splice acceptor variant	heterozygous	<a href="#">SEARCH</a>
<a href="#">SorbMutDB</a>	LBK_15M2-1535	USDA Lubbock EMS	stop gained	heterozygous	<a href="#">SEARCH</a>

Click here to  
order seeds



Search for all genes with potential LOF alleles in this accession

# Community Engagement Is Our Priority!

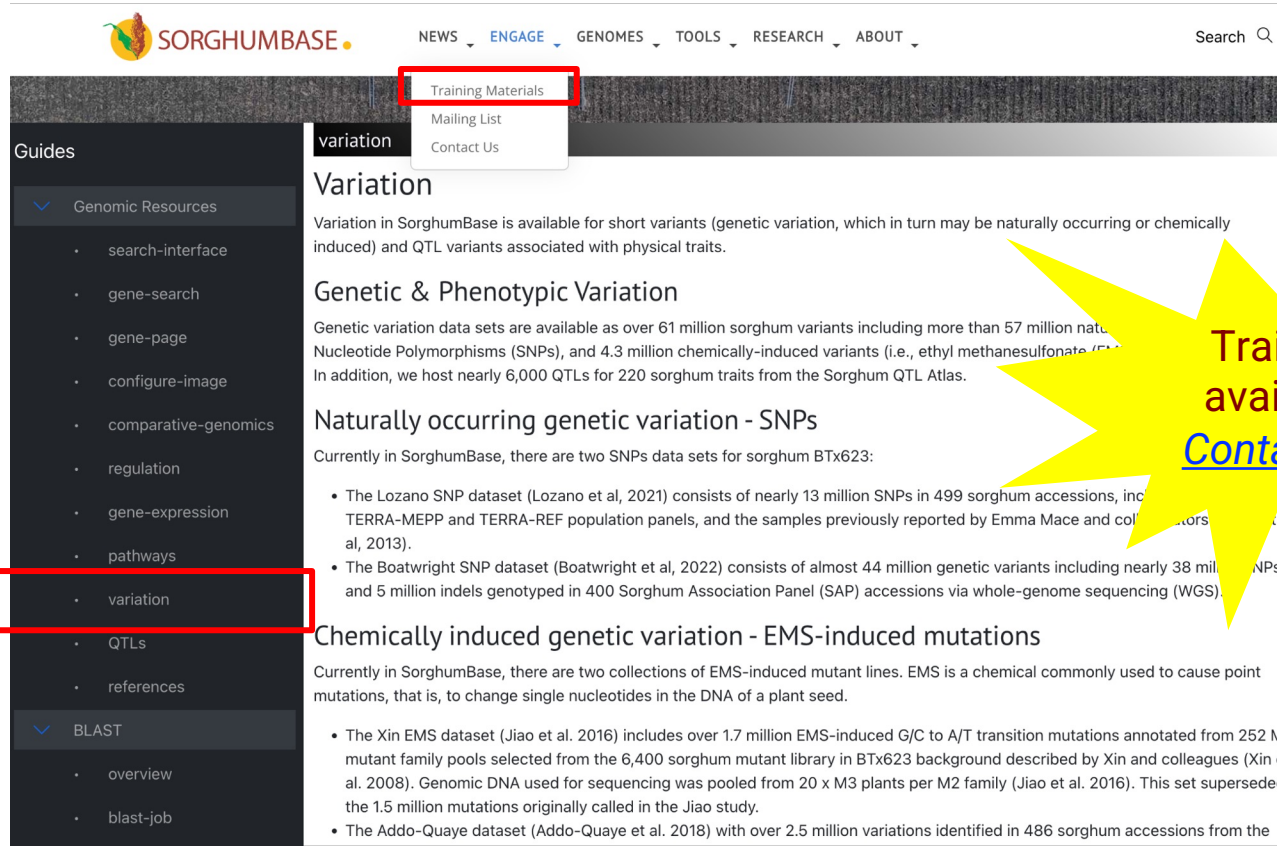
[contact@sorghumbase.org](mailto:contact@sorghumbase.org)

- Mailing list
  - Visit [www.sorghumbase.org](http://www.sorghumbase.org) to join!
- Events (<https://www.sorghumbase.org/events>)
- Training guides and videos (<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)



## Quick Guide

# Becoming an expert at SB



**SORGHUMBASE** NEWS ENGAGE GENOMES TOOLS RESEARCH ABOUT Search

variation  
Training Materials  
Mailing List  
Contact Us

**Guides**

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

**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 & Phenotypic Variation**

Genetic variation data sets are available as over 61 million sorghum variants including more than 57 million naturally occurring Nucleotide Polymorphisms (SNPs), and 4.3 million chemically-induced variants (i.e., ethyl methanesulfonate (EMS)-induced mutations). In addition, we host nearly 6,000 QTLs for 220 sorghum traits from the Sorghum QTL Atlas.

**Naturally occurring genetic variation - SNPs**

Currently in SorghumBase, there are two SNPs data sets for sorghum BTx623:

- The Lozano SNP dataset (Lozano et al, 2021) consists of nearly 13 million SNPs in 499 sorghum accessions, including the TERRA-MEPP and TERRA-REF population panels, and the samples previously reported by Emma Mace and colleagues (Mace et al, 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 genetic variation - 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 Xin EMS dataset (Jiao et al. 2016) includes over 1.7 million EMS-induced G/C to A/T transition mutations annotated from 252 M3 mutant family pools 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 x M3 plants per M2 family (Jiao et al. 2016). This set superseded the 1.5 million mutations originally called in the Jiao study.
- The Addo-Quaye dataset (Addo-Quaye et al. 2018) with over 2.5 million variations identified in 486 sorghum accessions from the

# Acknowledgements

## Ware Lab 2024



## Posters:

- **#700-52:** rsID work
- **#700-59:** GRAS Family Transcription Factors
- **#900-22:** Nitrogen Use Efficiency & Comparative GRN
- **#900-52:** Gramene
- **#900-61:** SorghumBase 2024

We gratefully acknowledge support from  
USDA-ARS-8062-21000-051-00D



Shukراً جزیلاً  
ngiyabonga  
teşekkür ederim  
danke 謝謝  
спасибо  
Баярлалаа  
faafetai lava  
nanni nandri kiitos dankie  
dhanyavad  
hvala maururu  
koszonom  
vinaka  
spasiobi  
blagodaram  
kia ora barka  
welalin tack  
dank je  
misaoatra matondo  
paldies grazzi  
mahalo  
tapadh leat  
xhala  
asante manana  
obrigada  
tenki  
murakoze  
chokrone  
mamnun  
dijere dieuf tau  
дякую  
mochchakkeram  
sulpay  
taiku  
go raibh maith agat  
arigato takk  
dakujem  
trugarez  
merci  
dhanyavadagalu  
shukriya  
merce  
merci  
xiexie  
감사합니다  
rahmah  
najis tuka  
kam sah hamnida  
didi madloba  
mesa  
sagolun  
sukriya  
kop khun krap  
kop khun krap  
gracias  
gratias ago  
chnorakaloutioun  
dankon aciu  
akun  
dankon aciu  
bedankt  
enkosi  
bayarlalaa  
gracie  
dziękuje  
sobodi  
dekuji  
obrigado  
o