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

Sharon Wei

ASPB, June 2024

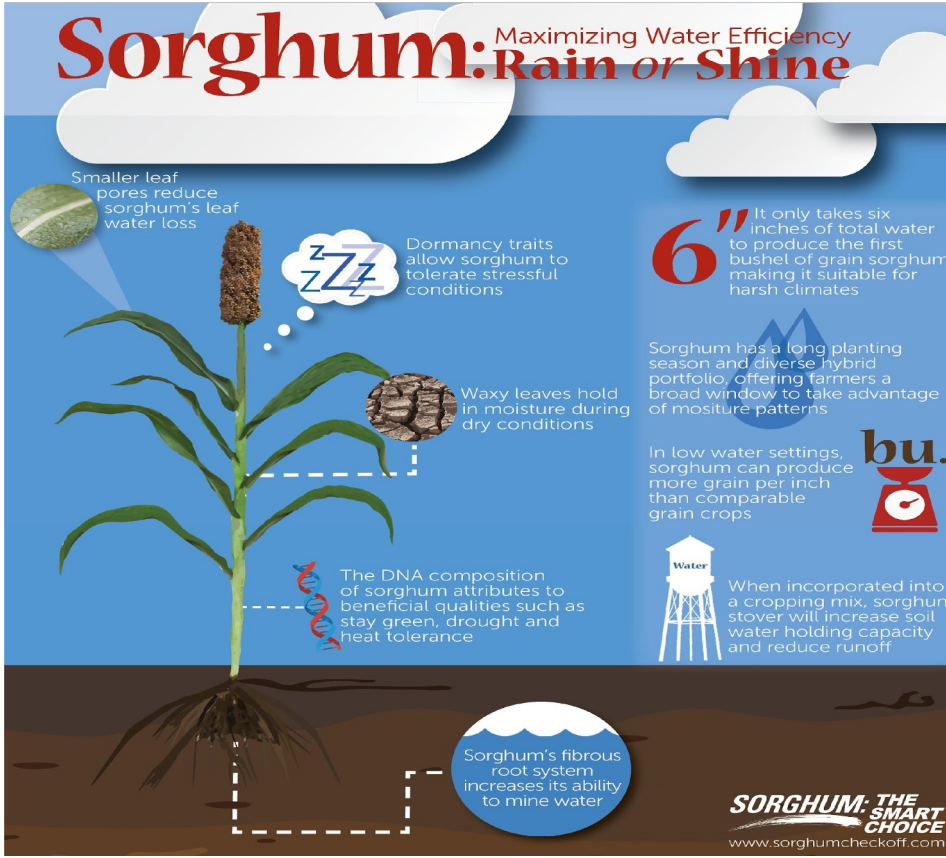
USDA Agricultural Research Service

CSH

Follow us **@SorghumBase**

Why Sorghum?

- **Fifth largest crop worldwide**
  - grain, forage, sugar crop (wine)
- **2nd 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**



SorghumBase is a USDA funded Community Knowledgebase

Genetic Variation

Explore the Impact of Variants c

News

Relevant developments in Sorgh

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

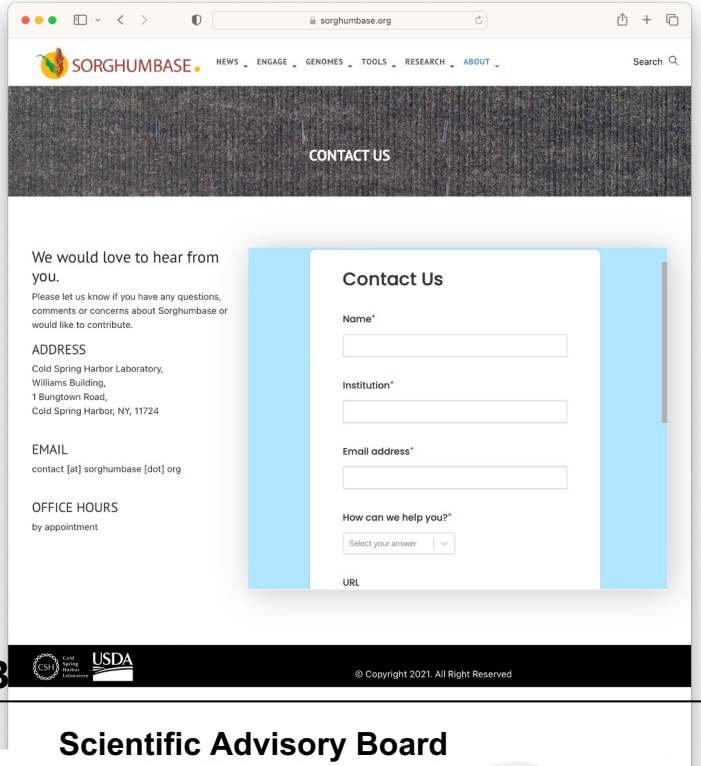
Genomes

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

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

Community Hub

- **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**



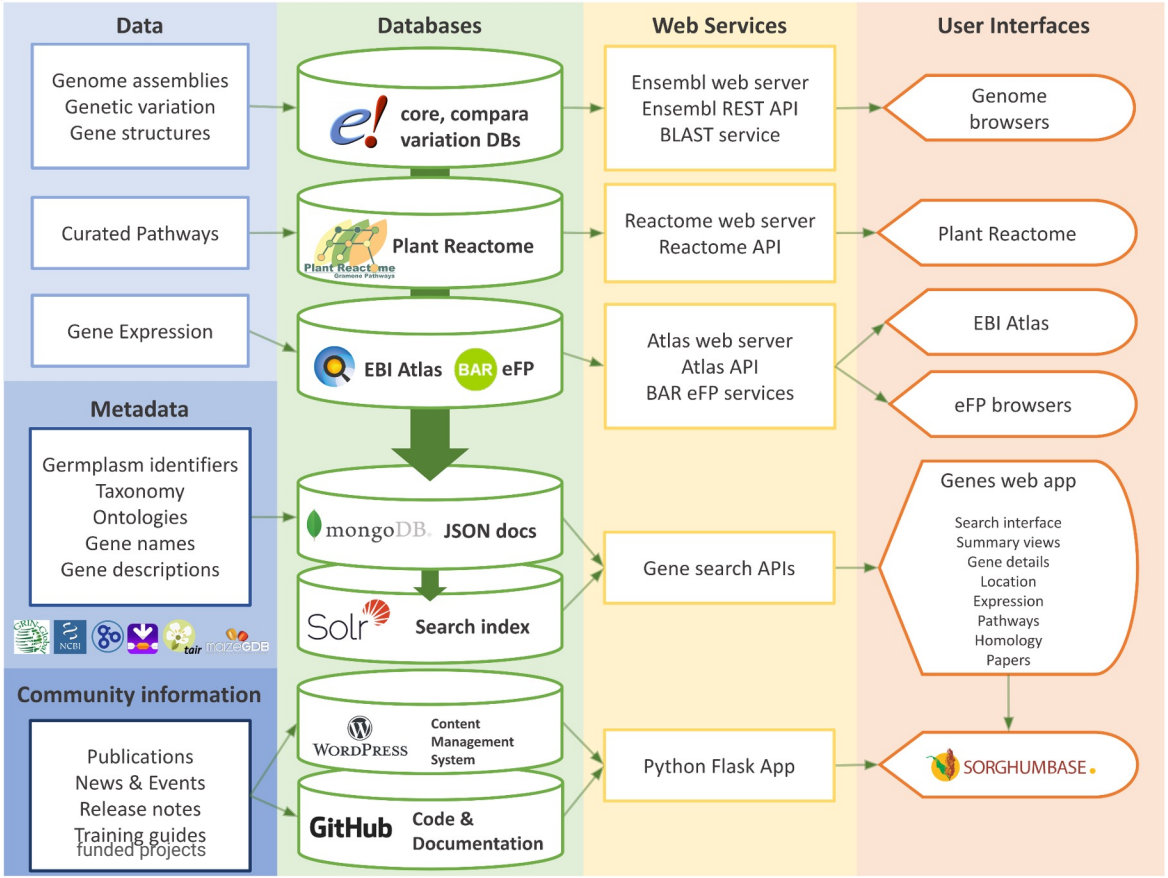
Scientific Advisory Board

LAURA MAYOR
Sorghum Breeding Lead
Corteva Agriscience

MITCHELL TUINSTRA
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

Plant Reactome

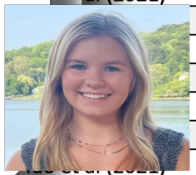
Oregon State University

Gene Expression

BAR

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.

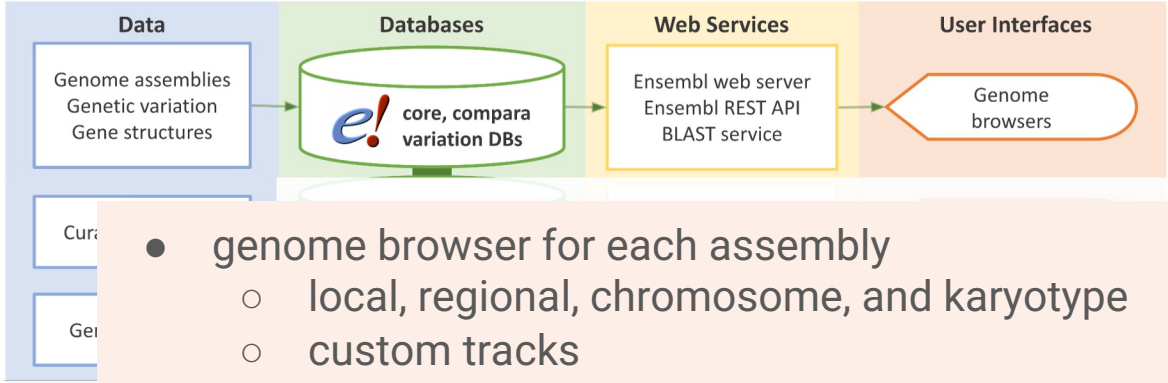
BAR: we curated 6 studies and generated pictographs for eFP browser. 10 new studies to be processed for next release.

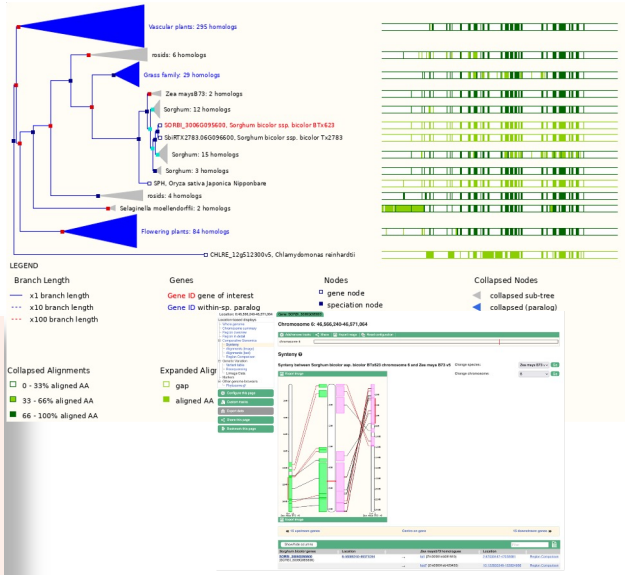


	Natural Variation	EMS Populations	All
Publi		Jiao, 023	All EMS	
# Ac			28	2505
# SI				59M

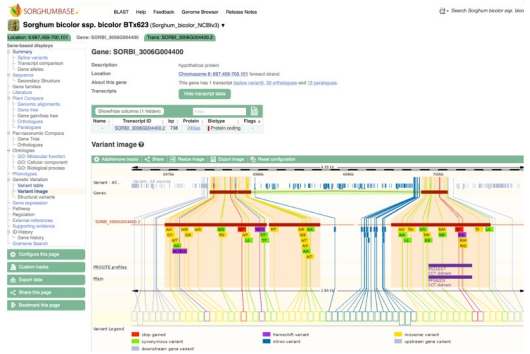
S. bicolor ssp. bic							r et al (2022)
S. bicolor ssp. dru							022)
S. bicolor ssp. verticilliflo							Tao et al (2021)
	AusTRCF317961					CNS0314256	Tao et al (2021)
S. propinquum	S369-1, AGG324436					CNS0314264	Tao et al (2021)

Genome Browser

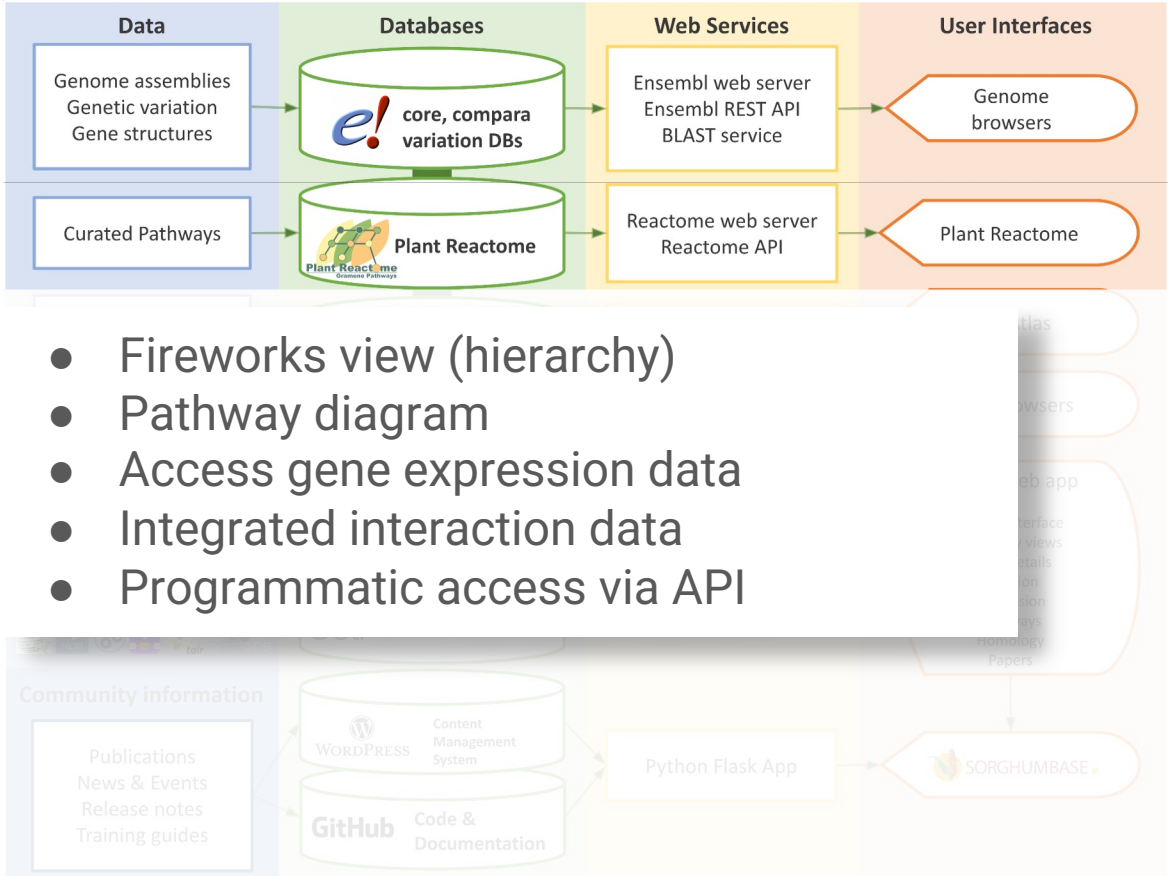




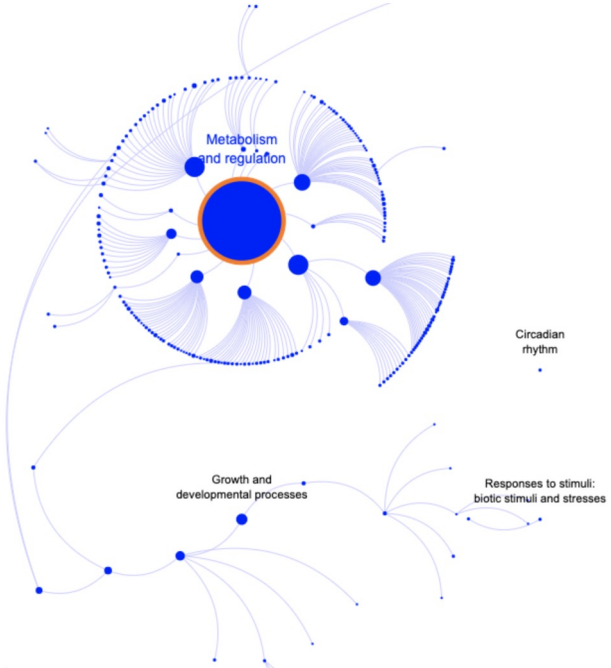
- genome browser for each assembly
  - local, regional, chromosome, and karyotype
  - custom tracks
- gene/transcript/protein pages
- gene trees and orthologs
- genetic variation views
  - population genotypes, frequency
  - Variant Effect Prediction (VEP)
- BLAST: genome, transcriptome, proteome
- export data and views
- REST API



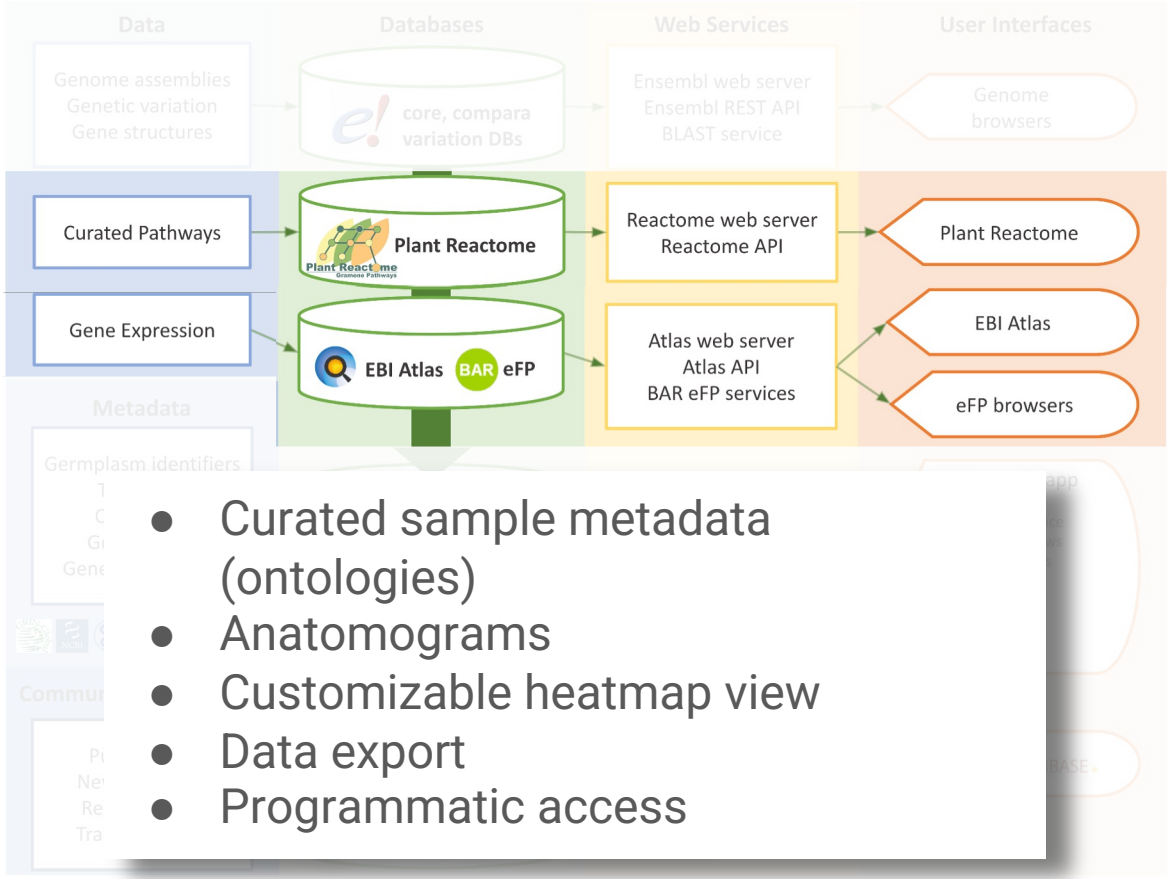
Pathway tools



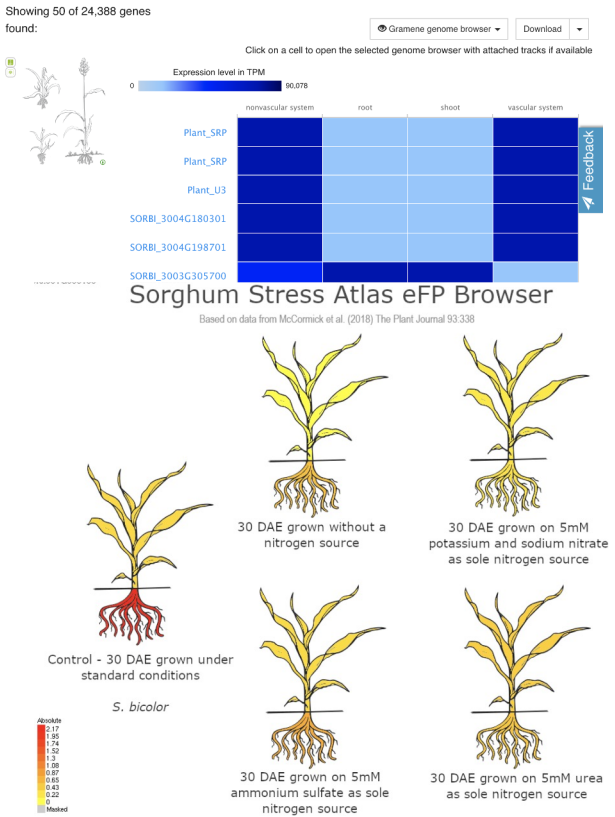
- Fireworks view (hierarchy)
- Pathway diagram
- Access gene expression data
- Integrated interaction data
- Programmatic access via API



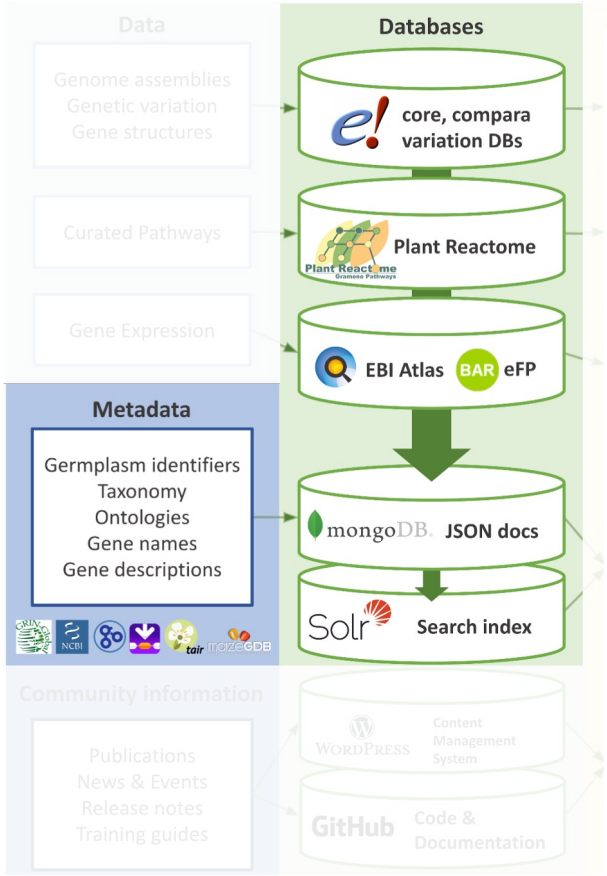
Gene Expression



- Curated sample metadata (ontologies)
- Anatomograms
- Customizable heatmap view
- Data export
- Programmatic access



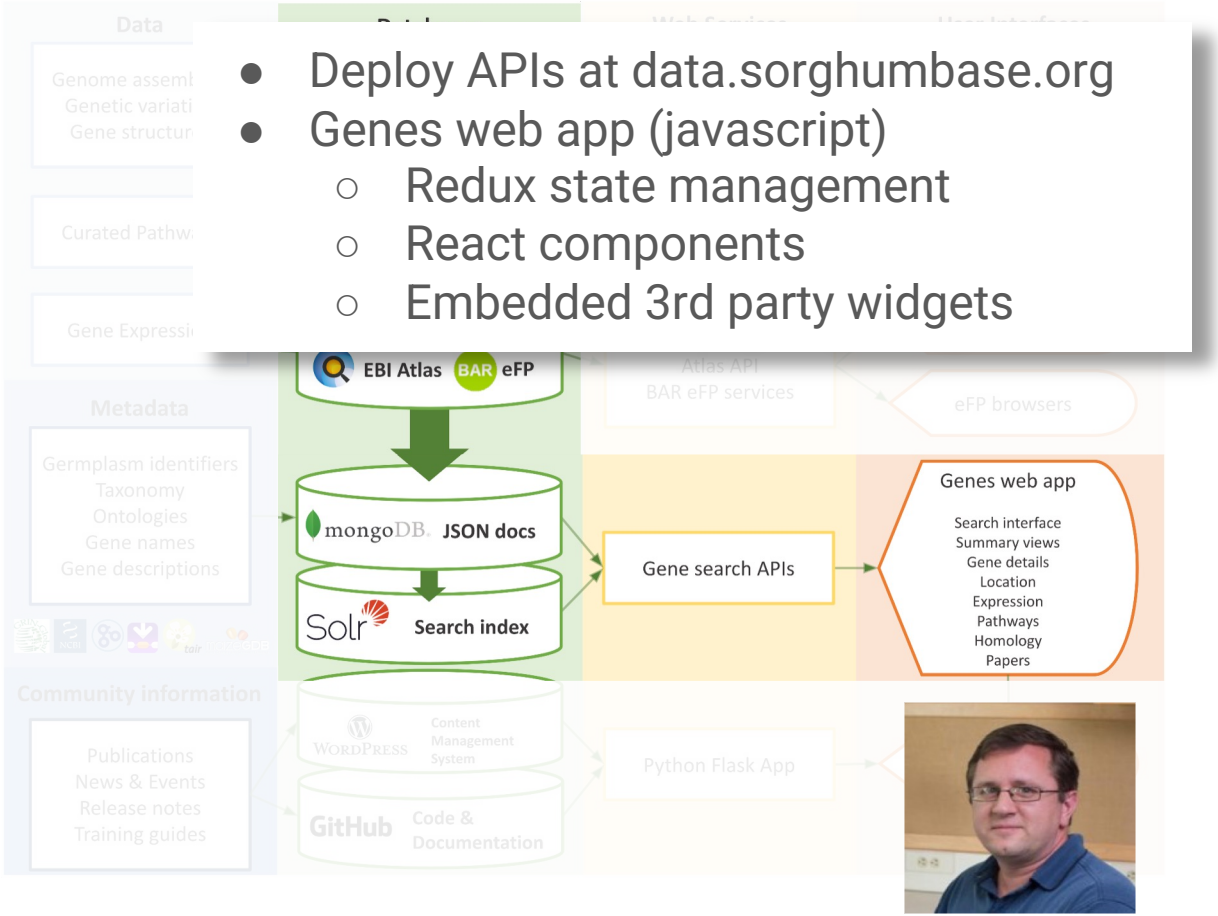
Building the Search Index

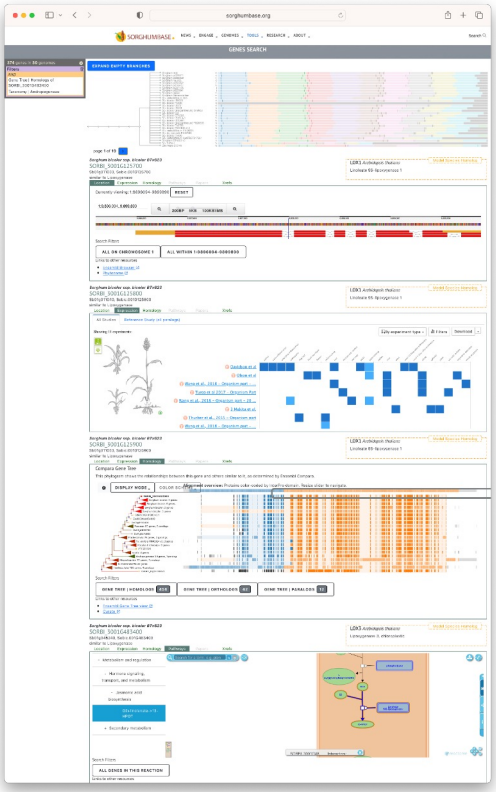


- Extract genes, gene trees, pathways, expression levels and domain annotations from primary databases
- Extend annotations with synonyms, ontologies and hierarchical metadata
  - GO, PO, TO, taxonomy, etc.
  - Interpro domain hierarchy
  - Pathway hierarchy
- Identifiers and common names for genes and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

Integrated Search

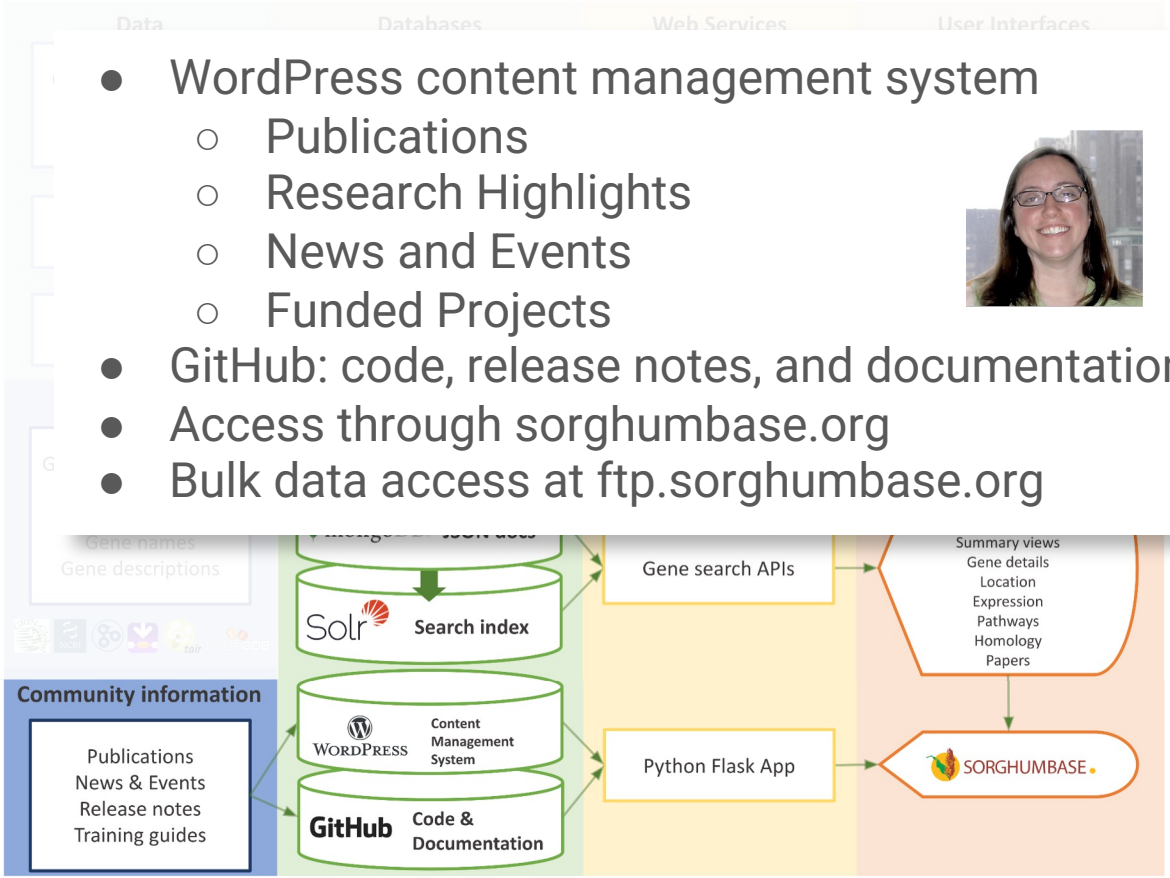
- Deploy APIs at data.sorghumbase.org
- Genes web app (javascript)
  - Redux state management
  - React components
  - Embedded 3rd party widgets





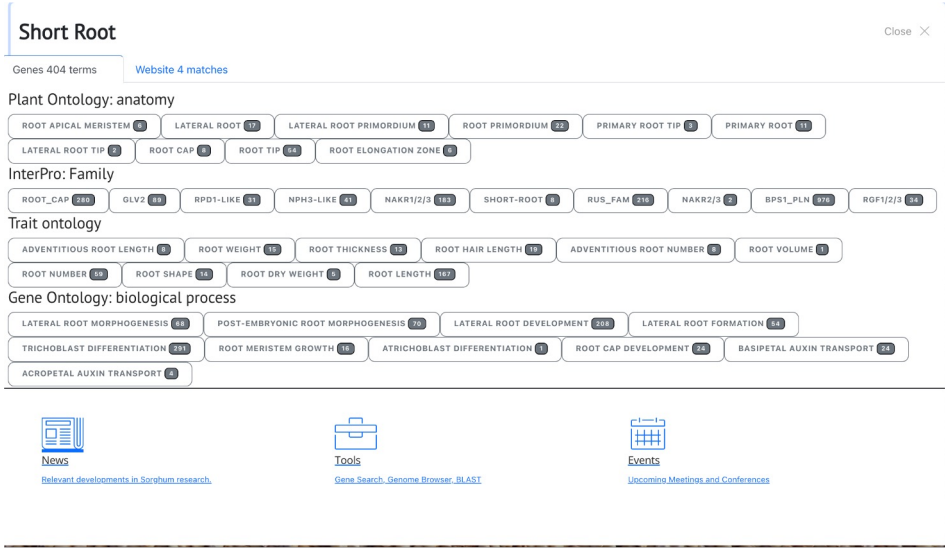
Community info

- 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

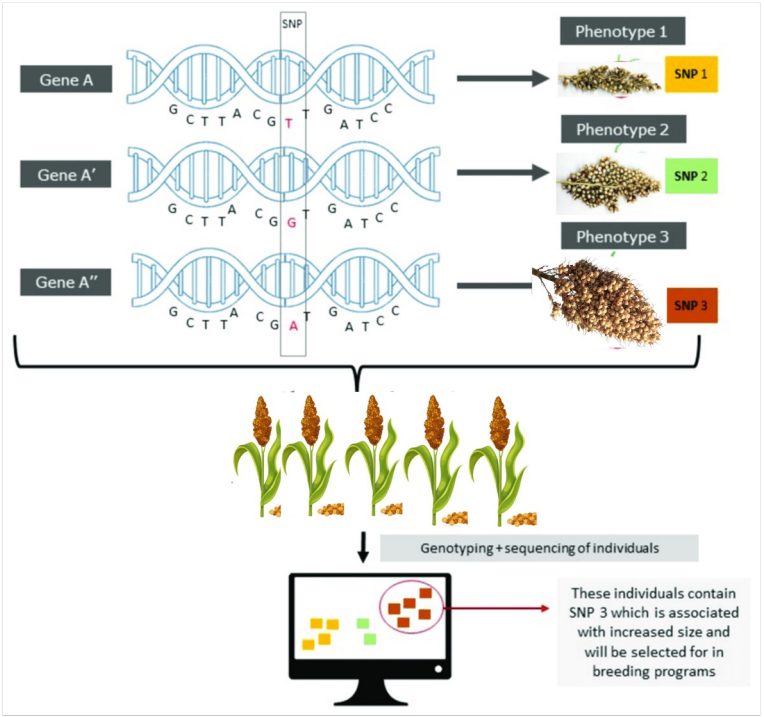
- many ways to find genes of interest
- not all genes are well annotated - homology helps
- not all gene models are okay - curate
- new integrated views based on user feedback
  - Sequences tab - quick access to DNA, RNAs, proteins
  - Germplasm with protein truncating variant (PTV) alleles



SORGHUMBASE
NEWS
ENGAGE
GENOMES
TOOLS
RESEARCH
ABOUT
Search
Welcome To Sorghumbase
Genomic Resources for the Sorghum Research Community
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



Standards for Genetic Variation WG Coordinate With EVA For Stable Variant IDs

AgBioData

Toward enhanced genomics, genet
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

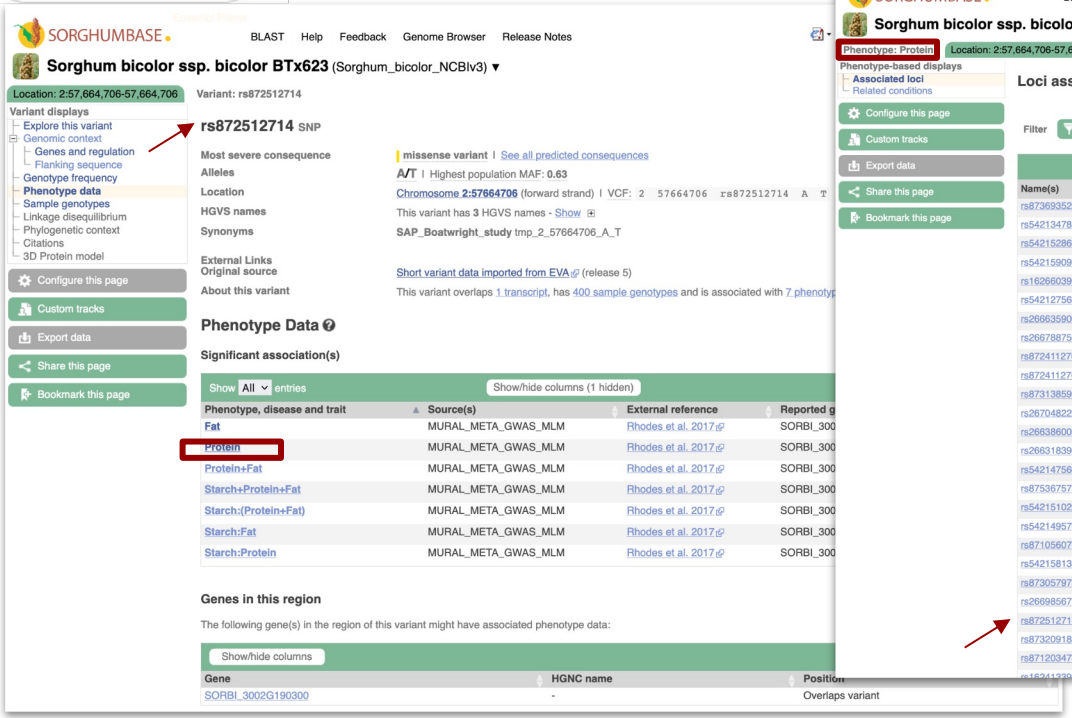
- Data
  - **41M SNPs assigned rsIDs (EVA release5)**
  - **234 GWAS traits from meta-analysis of 25 studies (SAP population)**
  - Non-coding feature tracks for all genomes (tRNA, miRNA, snoRNA …)
  - Sorghum Literature: 214 papers, 30 highlights, 289 funded projects
  - expression data sets from 11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

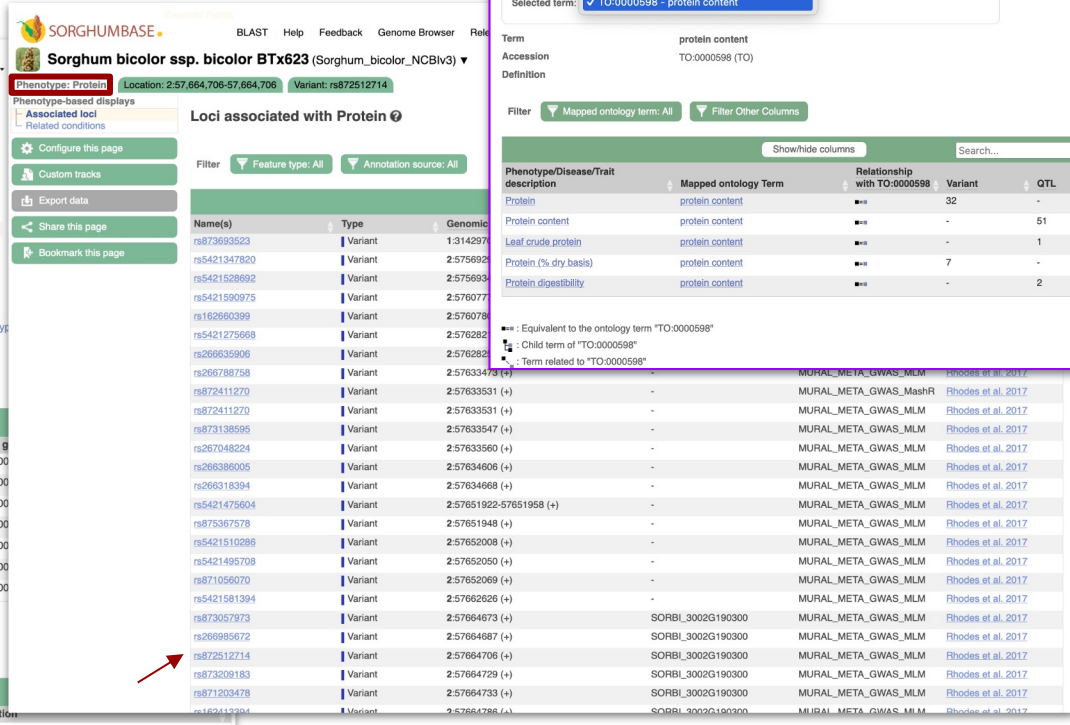
- Tools and functions
  - eFP expression browser
  - Sequence tab
  - **Germplasm tab serves Protein Truncation mutant lines from GRIN Global And SorbMutDB**
  - Papers tab for community curation

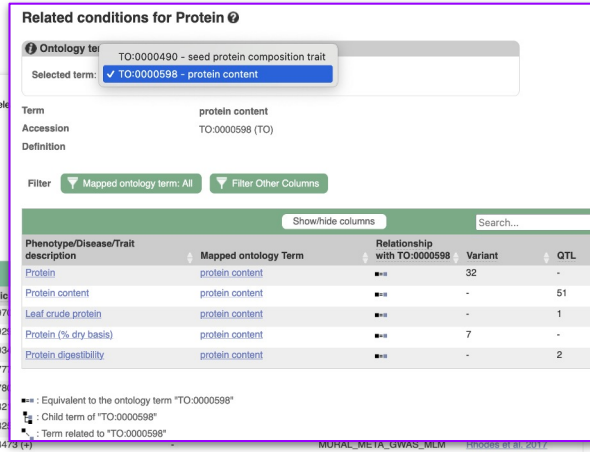
Phenotype data

Variants with associated phenotype (GWAS meta-analysis)

Genotype => Phenotypes



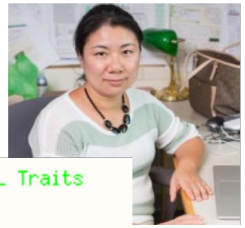




GWAS Data from Mural *et al* (2021)

Phenotype => Genotypes

Associated Phenotypes in Karyotype View



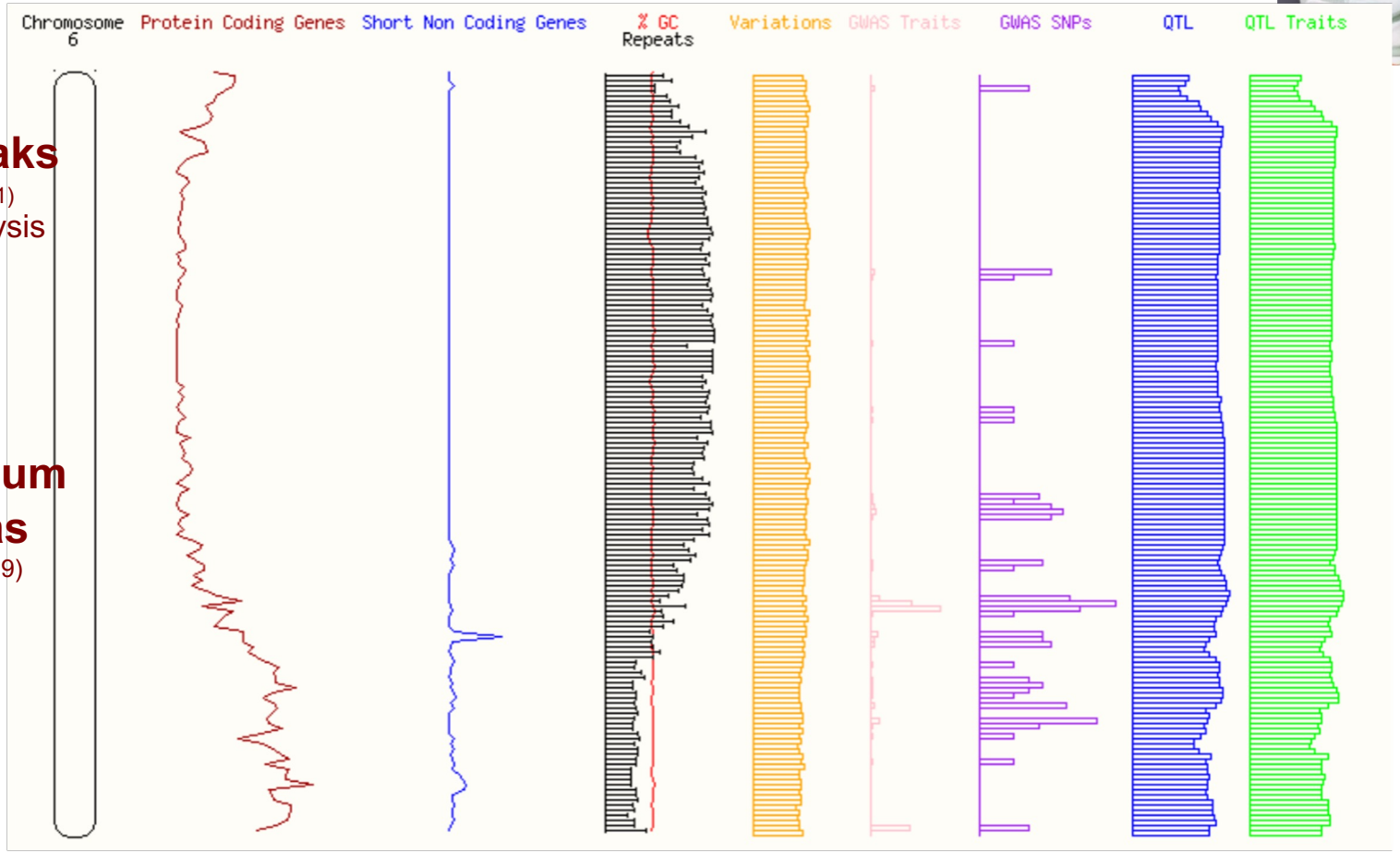
GWAs peaks

(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
PI 678262	PI678262	Purdue EMS	stop gained	heterozygous	SEARCH
PI 701663	ARS105	USDA Lubbock EMS	splice acceptor variant	heterozygous	SEARCH
PI 701655	ARS97	USDA Lubbock EMS	splice acceptor variant	heterozygous	SEARCH
SorbMutDB	LBK_15M2-1535	USDA Lubbock EMS	stop gained	heterozygous	SEARCH

Click here to order seeds

Search for all genes with potential LOF alleles in this accession

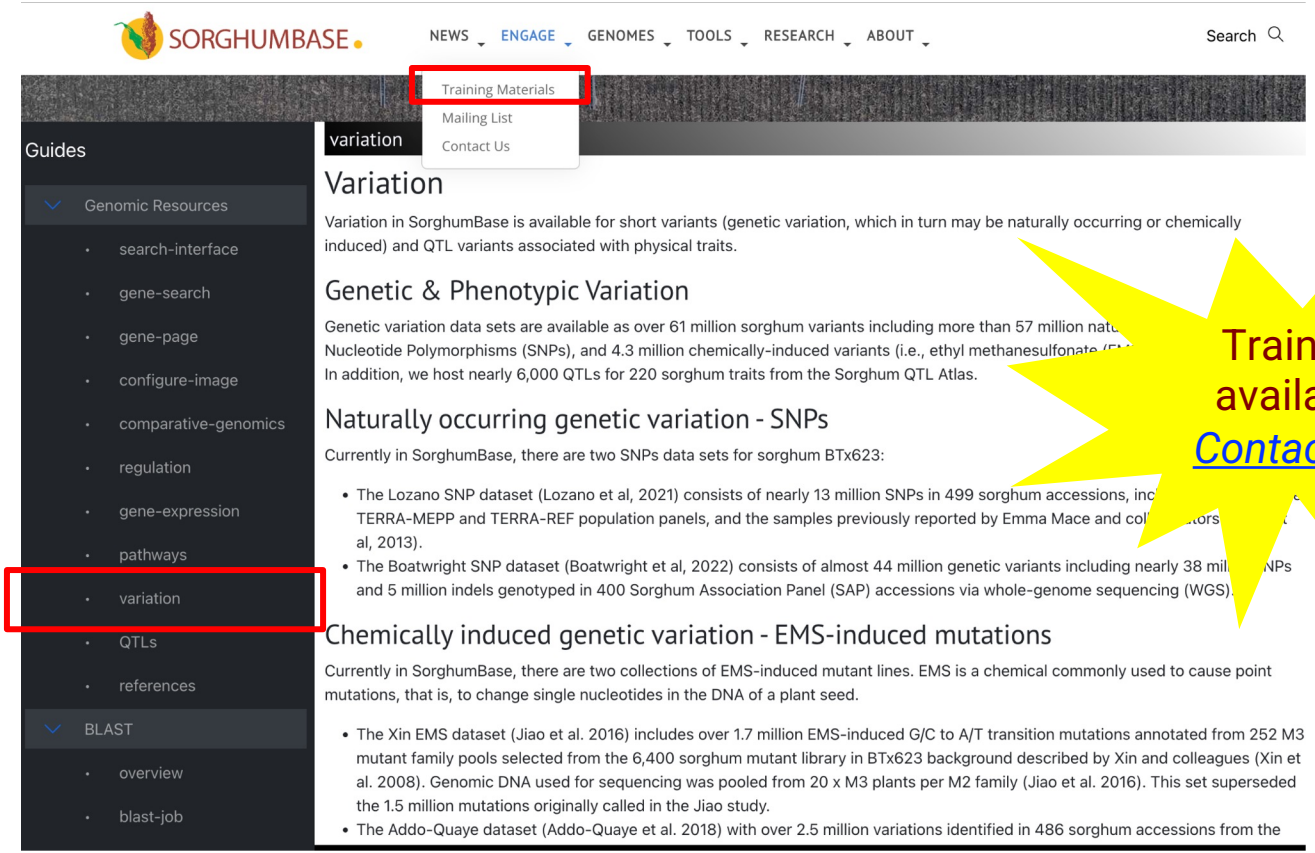
Community Engagement Is Our Priority!

contact@sorghumbase.org

- Mailing list
  - Visit 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)

Becoming an expert at SB

Quick Guide



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

CSH

USDA ARS Agricultural Research Service

شكراً جزيلاً
ngiyabonga
рахмат
danke
謝謝
teşekkür ederim
Баярлалаа
спасибо
faafetai lava
mersi
kia ora
barka
welalin
tack
spas
mahalo
tapadh leat
vinaka
спасибі
blagodaram
dank je
misaotra
matondo
paldies
grazzi
kiitos
dankie
nanni
nandri
dhanyavad
thank you
gracias
хвала
asante
manana
obrigada
tenki
bedankt
bayarlalaa
gràçie
hvala
mauruuru
köszönöm
aciu
dankon
akun
enkosi
djiere dieuf
tau
mochchakkeram
murakoze
дякую
mamnun
chokrane
dziękuję
chnorakaloutioun
gratias ago
gràcies
sulpáy
go raibh maith agat
sobodi
dēkuji
sagolun
sukriya
kop khun krap
taiku
obrigado
mèsi
didi madloba
najis tuke
ありがとう
tanemirt
grazie
arigatô
takk
dakujem
trugarez
kam sah hamnida
rahmat
terima kasih
rahmet
diolch
dhanyavadagalu
shukriya
merce
мерси
তোমাকে ধন্যবাদ
감사합니다
xièxie
ευχαριστώ
merci