

FAIR from the Start: Empowering Data Sharing in Agricultural Research

FAIR Scientific Literature
Working Group Update

AgBioData Workshop

PAG33

1/11/2026

Leonore Reiser



Members



Leyla Cabugos
Librarian, Cal Poly



Jacqueline Campbell
Geneticist, USDA ARS
SoyBase curator



Jenna Daenzer
Peer review Coordinator
Genetics Society of
America



Sook Jung
Asst Research
Professor, GDR



David Molik
Center for Scholarly
Publishing, Kansas
State University
Department head



Daniela Raciti
Exec Editor,
microPublication &
Wormbase Curator



Leonore Reiser
TAIR curator



Adam Wright
Software Engineer
WormBase, Reactome



Karen Yook
Exec Editor,
microPublication &
Wormbase Curator



We gratefully acknowledge support from NSF, grant #2126334



FSL Working Group Goals



Identify bottlenecks in the publication-curation pipeline.

- Identify sets of existing or desired tools or biocuration resources to increase literature curation throughput and accuracy.
- Publish recommendations and a roadmap for authors and publishers to increase the FAIRness of research.



A better literature curation workflow

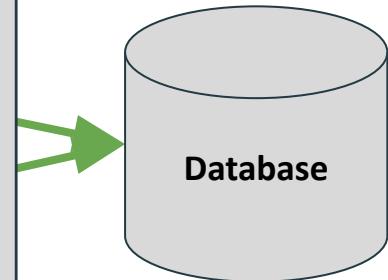
AUTHOR
Data generator



Database
CURATOR



- ✓ Data curation at time of publication
- ✓ Clear guidance on data submission & databases
- ✓ Validators to check for data submission
- ✓ Validators for data formats
- ✓ Nomenclature standards and validators
- ✓ No more data available upon request
- ✓ Better paper metadata, AI based paper classification
- ✓ More funding for curation, AI curation assistants

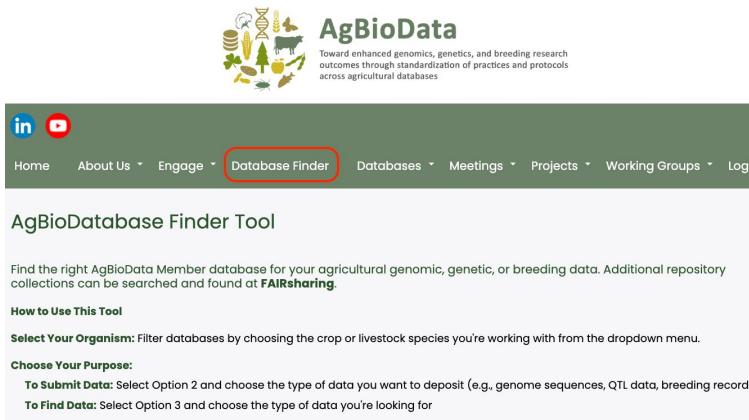


FSL Working Group Goals

- Identify bottlenecks in the publication-curation pipeline.
- ✓ Identify sets of existing or desired tools or biocuration resources to increase literature curation throughput and accuracy.
- Publish recommendations and a roadmap for authors and publishers to increase the FAIRness of research.



Tool: AgBio Database Finder Drupal Module



AgBioDatabase Finder Tool

Find the right AgBioDatabase Member database for your agricultural genomic, genetic, or breeding data. Additional repository collections can be searched and found at [FAIRsharing](#).

How to Use This Tool

Select Your Organism: Filter databases by choosing the crop or livestock species you're working with from the dropdown menu.

Choose Your Purpose:

To Submit Data: Select Option 2 and choose the type of data you want to deposit (e.g., genome sequences, QTL data, breeding records)

To Find Data: Select Option 3 and choose the type of data you're looking for

Option 1: Filter by species

<https://www.agbiodata.org/databasefinder>
<https://doi.org/10.17912/micropub.biology.001896>

Answer these questions to help identify databases that are suitable for your needs

[Clear Answers](#)

1. Organism

- Arabidopsis**
- Bacteria**
- Bovinae**
- Citrus**
- Cotton**
- Fish**
- Forest trees**
- Grasses**
- Insects**
- Legumes (chickpea, soy and lentil)**
- Generalist Animal DB**
- Generalist Plant DB**
- Maize**
- Poaceae (e.g., wheat, rice, barley, etc.)**
- Rosaceae (e.g., apple, pear, almond, etc.)**
- Solanaceae (e.g., tomato, eggplant, potato, etc.)**
- Sorghum**
- Soybean**
- Vaccinium (e.g., blueberry, cranberry, etc.)**

Click one or more tiles to view/compare database details below.

[Select All](#) [Clear Selections](#)

Animal QLdb	AquaMine	BAR	BGD
Animal Quantitative Trait Loci Database	Aquatic eumetazoan species genome data mine resource	Bio-Analytic Resource for Plant Biology	Bovine Genome Database
CGD	CorrDB	CottonGen	FAANGMine
Database for Citrus Genomics, Genetics and Breeding Research	Animal Trait Correlation Database	Cotton Database Resources	Integrates genome assemblies and currently available annotation data for FAANG species
GDR	GDV	GrainGenes	Gramene
Database for Rosaceae Genomics, Genetics and Breeding Research	Database for Vaccinium Genomics, Genetics and Breeding Research	A Database for Triticeae and Avena	Comparative plant genomics for crops and model organisms
HGD	LIS	MaizeGDB	MaizeMine
Hymenoptera Genome Database	the Legume Information System	Maize Genetics and Genomics Database	Integrated queryable database for Zea mays genetics and genomics
PCD	Plant Reactome	Planteome	SGN
Database for Pulse Genomics, Genetics and Breeding Research	Gramine pathways	Ontologies and annotations of plant phenotypes and traits	Solgenomics
SorghumBase	SoyBase	TAIR	TreeGenes
Genomic resources for the sorghum research community	Integrating Genetics and Genomics to Advance Soybean Research	The Arabidopsis Information Resource	Forest tree genome database

<https://github.com/CU-CommunityApps/CD-finder>

Option 2: Filter by ...

Data to submit
OR

Data to find

2. Data to submit

- database cross-mapping
- geographic location
- gene expression
- gene functional annotation
- gene regulation
- gene report
- genetic map
- genetic variation
- genome analysis

3. Data to find

- database cross-mapping
- gene expression
- gene functional annotation
- gene regulation
- gene report
- genetic map
- genetic variation
- genome analysis
- genome annotation
- genomics
- genotype data
- geographic location
- homology-based gene prediction
- image

Click one or more tiles to view/compare database details below.

Select All

Clear Selections

Animal QTLdb	AquaMine	BAR	BGD	CGD
Animal Quantitative Trait Loci Database	Aquatic eumetazoan species genome data mine resource	Bio-Analytic Resource for Plant Biology	Bovine Genome Database	Database for Citrus Genomics, Genetics and Breeding Research
CorrDB	CottonGen	FAANGMine	GDR	GDV
Animal Trait Correlation Database	Cotton Database Resources	Integrates genome assemblies and currently available annotation data for FAANG species	Database for Rosaceae Genomics, Genetics and Breeding Research	Database for Vaccinium Genomics, Genetics and Breeding Research
GrainGenes	Gramene	HGD	LIS	MaizeGDB
A Database for Triticeae and Avena	Comparative plant genomics for crops and model organisms	Hymenoptera Genome Database	the Legume Information System	Maize Genetics and Genomics Database
MaizeMine	PCD	Plant Reactome	Planteome	SGN
Integrated queryable database for Zea mays genetics and genomics	Database for Pulse Genomics, Genetics and Breeding Research	Gramene pathways	Ontologies and annotations of plant phenotypes and traits	Solgenomics
SorghumBase	SoyBase	TAIR	TreeGenes	i5k
Genomic resources for the sorghum research community	Integrating Genetics and Genomics to Advance Soybean Research	The Arabidopsis Information Resource	Forest tree genome database	i5k Workspace@NAL

Apply filters and compare resources

2. Data to submit

- database cross-mapping
- geographic location
- gene expression
- gene functional annotation
- gene regulation
- gene report
- genetic map
- genetic variation

CottonGen Cotton Database Resources	FAANGMine Integrates genome assemblies and currently available annotation data for FAANG species	GDR Database for Rosaceae Genomics, Genetics and Breeding Research
Gramene Comparative plant genomics for crops and model organisms	HGD Hymenoptera Genome Database	LIS the Legume Information System
PCD Database for Pulse Genomics, Genetics and Breeding Research	Plant Reactome Grapene pathways	Planteome Ontologies and annotations of plant phenotypes and traits
SoyBase Integrating Genetics and Genomics to Advance Soybean Research	TAIR The Arabidopsis Information Resource	TreeGenes Forest tree genome database

LIS PCD SoyBase

	LIS	PCD	SOYBASE
DB name	LIS - the Legume Information System	Pulse Crop Database (PCD)	SoyBase
DB short description	LegumeInfo is a federally funded informatics service assisting in basic research in the legumes by relating data from multiple species, and traverse among various data types.	The Pulse Crop Database (PCD), formerly the Cool Season Food Legume Database (CSFL), has been developed by the Main Bioinformatics Laboratory at Washington State University in collaboration with the USDA-ARS Grain Legume Genetics and Physiology Research Unit, the USDA-ARS Plant Germplasm Introduction and Testing Unit, the USA Dry Pea and Lentil Council, Northern Pulse Growers and allied scientists in the US and across the world, to serve as a resource for Genomics-Assisted Breeding (GAB). GAB offers tools to identify genes related to traits of interest, among other methods to optimize plant breeding efficiency and research, by providing relevant genomic, genetic, and breeding information and analysis. Therefore, tools such as JBrowse and MapViewer can be found in this database, as well as key resources to provide access to the annotation of available transcriptome data, helping pulse breeders and researchers to succeed in their programs.	SoyBase is a long-term, federally funded genetics and breeding database for soybean breeders and researchers focused on soybeans.
DB URL	https://www.legumeinfo.org/	https://www.puledb.org/	https://www.soybase.org/
DB data types	<ul style="list-style-type: none"> • database cross-mapping* • gene expression* • gene functional annotation* • gene regulation* • gene report* • genetic map* • genetic variation* • genome analysis* • genome annotation* • genomics • genotype data* • geographic location • homology-based gene prediction* • metadata • molecular sequence annotation* • phylogenetic data* • QTL map* • sequence features* • whole genome association study* 	<ul style="list-style-type: none"> • database cross-mapping • gene expression • gene functional annotation • gene regulation • gene report • genetic map • genetic variation • genome analysis • genome annotation • genotype data • homology-based gene prediction • image • metabolic pathway prediction • ontology and terminology • phenotype • QTL • whole genome association study <p>* = Direct submission accepted</p>	<ul style="list-style-type: none"> • database cross-mapping* • gene expression* • gene functional annotation* • genetic map* • gene regulation* • gene report* • genetic variation* • genome analysis • genome annotation • genomics • genotype data • geographic location • homology-based gene prediction • metabolic pathway prediction • metadata • nucleic acid sites, features, and motifs • ontology and terminology • pedigree* • phenotype*

What about source/data type repositories?

- Add more tiles/databases to the tool?
- Create a collection in FAIRsharing?
- Maintain a separate list ordered by data type?
- *AI/Chatbot?*



(How) can AI (tools) help make published data more FAIR?

- What is the future of publishing and co-ordinated data curation in the age of AI?
- Can we make the data and metadata in papers more machine readable/FAIR?
- A freely accessible, universal AI tool that:
 - Scans papers and identifies types of data
 - Extracts experiment metadata
 - Validates data and metadata formats
 - Checks against nomenclature and other standards
 - Readies data for submission to appropriate repository



How to engage with publishers and journals?

How to introduce data curation into the publishing pipeline

How to connect journals to community databases



Community Workshop April 2025

Incentives



Apicuron Credits for authors



FAIRness badges for papers/journals

Workflow changes



Pre-submission or pre-print curatorial review



Advanced vetting to speed peer review

Start with society journals

Journals need assurance that recommended repos are FAIR



Increasing FAIRness of AgBioData Member Databases

- FAIR sharing record completion
- Assess FAIRness of databases
 - What minimum expectation and standards for member databases
 - How can AgBioData support meeting those expectation?



FSL Working Group Goals

- Identify bottlenecks in the publication-curation pipeline.
- Identify sets of existing or desired tools or biocuration resources to increase literature curation throughput and accuracy.
- Publish recommendations and a roadmap for authors and publishers to increase the FAIRness of research.



Thoughts for Discussion

- What if the roadmap was an *interactive tool* not a static white paper?
- How could an integrated curation- publication data pipeline be made sustainable?

