NRSP10 Progress Report

Database Resources for Crop Genomics, Genetics and Breeding Research

National Plant Germplasm Committee Meeting

Cornell University
June 1, 2017

Jim Moyer
Washington State University

NRSP 10 Progress by Objective

Expand online community databases currently housing high quality genomic, genetic and breeding data for Rosaceae, citrus, cotton, cool season food legumes and *Vaccinium* crops

- Four of the five databases redesigned and implemented in Tripal 2.1 with improved access to data, tools and tutorials - easy upgrade path to Tripal 3
- Presentations at over 50 conferences, workshops, and meetings
- Widespread use of the databases (Oct 1, 2014 – May 30, 2017)

Database	# Visitors	# Visits	# Pages Served	# Countries
Rosaceae	50,458	131,742	605,960	171
Cotton	25,262	60,331	349,731	163
Citrus	12,169	18,756	83,045	146
Legumes	8,027	13,218	60,606	144
Vaccinium	4,049	5,527	18,934	94
Total	**99,965**	**229,574**	**1,118,276**	

CSFL Crops Data Search Tools General Help Login Search

Cool Season Food Legume Crop Database Resources

Genomic, Genetic and Breeding Resources for Pea, Lentil, Chickpea and Faba Bean Crop Improvement

News and Events

- PathwayCyc for C. reticulatum PI489777 added (2/1/17)
- New CSFL Video Tutorials (1/7/17)
- CSFL poster at PAG, January 14-18, 2017
- New CSFL User Manual (11/21/16)

more...

Crops Quick Start

lentil

pea

chickpea

faba bean

Tools Quick Start

genomics
- View Genomes
- Find Sequences
- Search Genes
- BLAST Sequences

genetics
- Browse Maps
- Search Markers
- Find QTLs
- Compare Maps

breeding
- Manage Breeding
- Manage Data
- Analyze Data
- Decision Tool

general
- Submit Data
- Presentations
- Work in Progress
- Contact Us

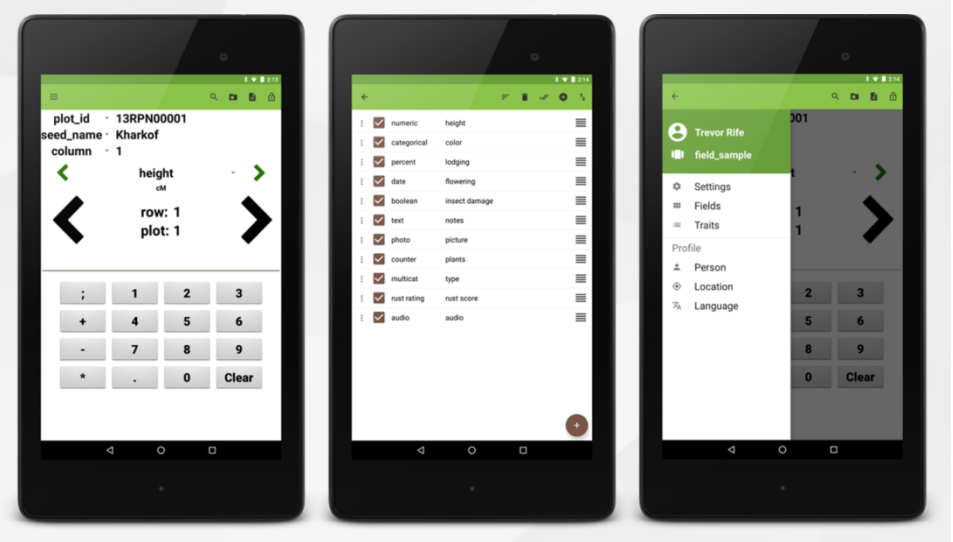
USA Dry Pea & Lentil Council · Northern Pulse Growers Association · NRSP10 Crop Database Resources · USDA ARS Agricultural Research Service · Washington State University

Copyright © 2010-2017. This site is designed to work with IE, Chrome, Firefox, Safari and Opera. Contact us

NRSP 10 Progress by Objective

Develop/Implement a tablet application to improve efficiency and accuracy of phenotypic data collection and ease of upload to private breeding program databases

- Collaborating with Jesse Poland to further develop the widely used FieldBook App for the needs of NRSP10 crops – NRSP10 provides funds for 1/2 developer position
- Provided Samsung tablets with FieldBook to over 50 NRSP10 breeders and scientists – feedback fueling new feature development
- Provide webinars, hands-on training and support with frequent presentations at conferences and meetings – encouraging adoption



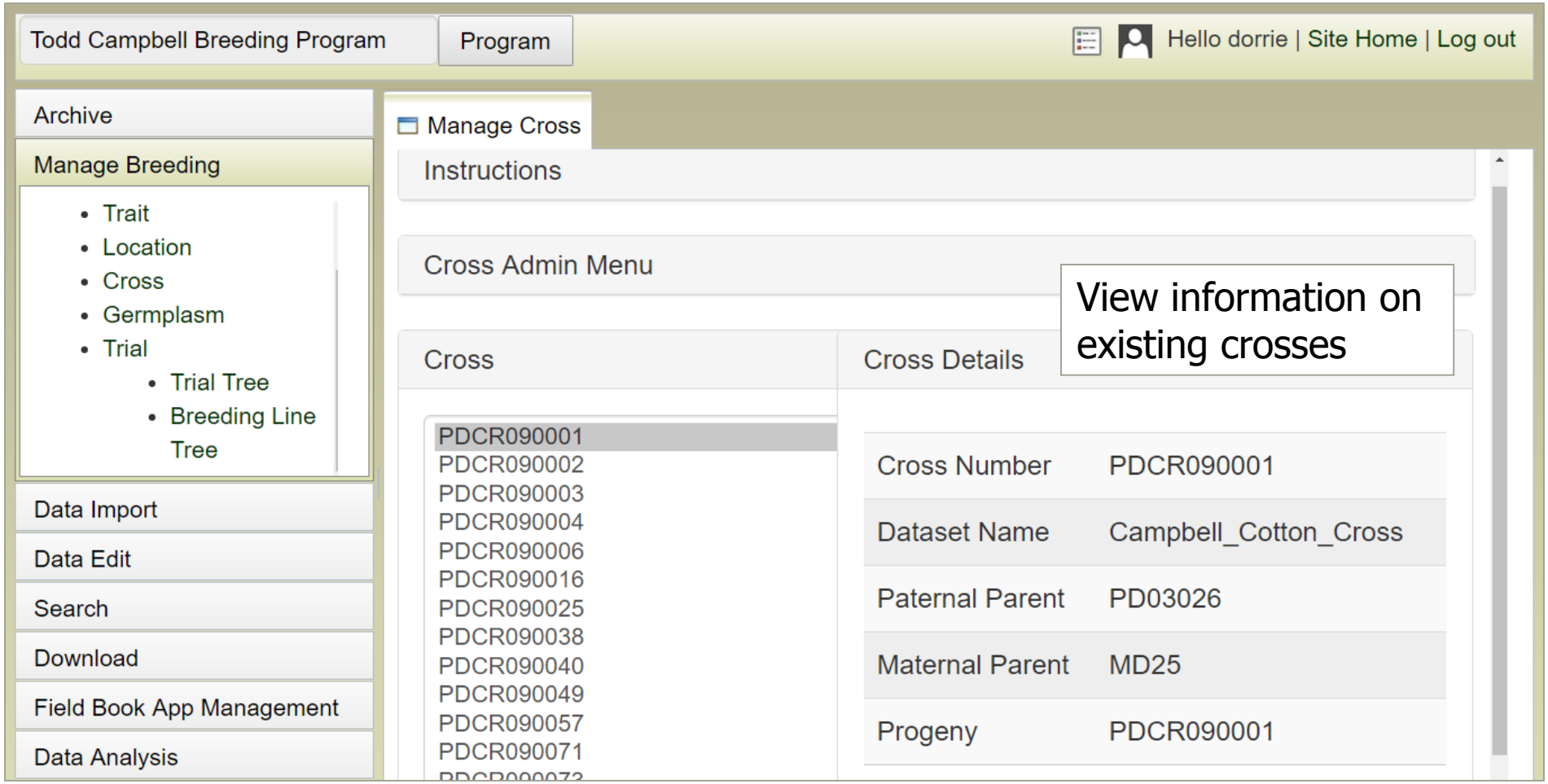
3090 installs by users from 137 countries

NRSP 10 Progress by Objective

Develop a Tripal Application Programming Interface for building breeding databases to connect private breeding databases with public genomics, genetics and breeding data

- Many input sessions held with NRSP10 breeders to identify specific needs for a Tripal Breeding Information Management System (Tripal BIMS)
- Beta testing of development by focus group of 4 NRSP10 breeders
- Presentations at conferences/workshops by NRSP10 breeders Ksenija Gasic (Clemson University Peach Breeding) and Todd Campbell (USDA-ARS, Cotton Breeding) to provide more feedback and advertise for breeding program planning (e.g. start using FieldBook)
- Status: design for V1.0 completed, data loaders in place for breeders to upload their own data, data uploaded for 3 programs, breeders testing and providing feedback. Full release of V1.0 in early 2018.

Example: Todd Campbells BIMS in CottonGen



NRSP 10 Progress by Objective

Convert GenSAS, a comprehensive online community genome annotation and curation tool, to Tripal

- A single online portal that combines numerous genome annotation tools into a single platform
- Easy-to-use interfaces, with integrated instructions allowing researchers at all skill levels to annotate genome sequences
- User accounts keep data private and secure as well as allow users to share data for collaborative annotation
- Initially funded by the 2011-2015 USDA NIFA PineRefSeq Project
- Since v3.0 release it has been accessed by 2,129 visitors from 82 countries, with 5,701 sessions and 24,419 pages viewed

NRSP 10 Progress by Objective

Convert GenSAS, a comprehensive online community genome annotation and curation tool, to Tripal

Why is it GenSAS needed?

Driven by the volume of genome sequences being generated by individual labs, most without access or skills to run and reformat the output of gene prediction programs, perform manual and format the results into files needed for publication.

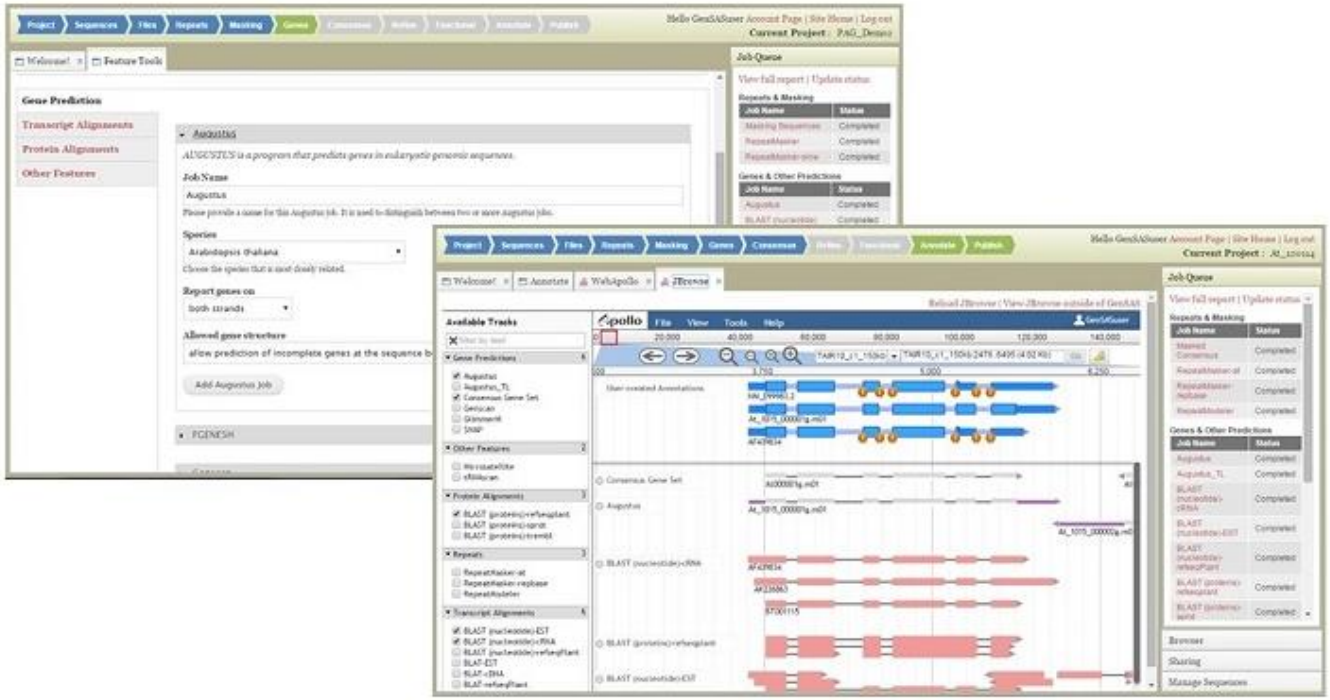
- V5.0 released January 2017, V5.1 in June
- More functionality and improved performance
- Linked with Tripal but not fully integrated yet (GenSAS and GenSASLite)
- Being used now by many communities to do community curation of genomes
- We are exploring commercialization opportunities for the standalone software (largely due to the processor power needed to support the demand for this software and the resources needed to run big jobs)
- Manuscript in preparation for submission in July

GenSAS v5.0

Home | Use GenSAS | Available Tools | Help

Welcome to GenSAS

The Genome Sequence Annotation Server (GenSAS) is an online tool that provides a pipeline for whole genome structural and functional annotation. Users can upload genome sequences and select from a variety of tools for repeat masking, prediction of gene models and other structural features as well as functional annotation tools. GenSAS integrates with JBrowse and Apollo to provide visualization and editing. Please see our video tutorial under the "Help" tab.



User login

Username *

Password *

- Create new account
- Request new password

Log in

NRSP 10 Progress by Objective

Home | Download | Support | Extensions | API | About | Demo Site | Log In

Quick Links

- v2.x User's Guide
  - Quick Docker install
  - Download and install
  - Upgrade instructions
  - Developer's handbook
- Other documentation
- Tripal @ gitHub.com
- Tripal @ drupal.org

Tweets by @TripalProject

The Tripal Project
@TripalProject
Tripal provides an out-of-the-box setup for a genomics website easily put a new genome assemblies & annotations online #PAGXXV
16h

The Tripal Project
@TripalProject
Tripal provides an integration with the @gmodproject Chado database schema and Drupal, a popular Content Management Systems #PAGXXV
18h

Tripal

Tripal is a toolkit for construction of online biological (genetics, genomics, breeding, etc), community database, and is a member of the GMOD family of tools. Tripal provides by default integration with the GMOD Chado database schema and Drupal, a popular Content Management Systems (CMS). Tripal is licensed under the GNU General Public License version 2.

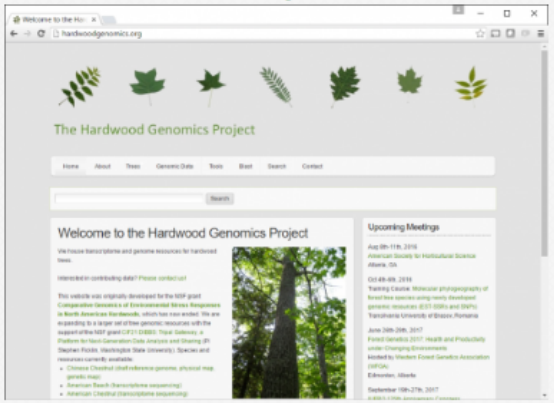
Objectives

- Provide a framework for those with genomic, genetic and breeding data for creation of an online site for display, search and visualization.
- To use community-derived standards and ontologies to facility continuity between sites which in turn fosters collaboration and sharing
- Provide an out-of-the-box setup for a genomics site for those who simply want to put new genome assemblies and annotations online.
- Provide Application Programming Interfaces (APIs) for complete customization such that more advanced

Search

Sites Using Tripal

Hardwood Genomics Project



[View more]

News

TRIPAL: Meeting & Hackathon

01/03/2017 - 23:42

Where: Plant & Animal Genome, San Diego, CA
Town & Country in Royal Palm Salon 4
When: Thursday, January 12 and Friday, January 13 from 8am - 5pm

Support

- Highly engaged and supportive user communities ☺
- Leveraged funding
  - Further Development of CottonGEN: A Genomics, Genetics and Breeding Database for Cotton Improvement (PI Main). Cotton Incorporated and USDA-ARS, Jan 1, 2017- Dec 31, 2020, **$741,889**.
  - Standards and CyberInfrastructure that Enable Big-Data Driven Discovery for Tree Crop Research (PI Main) NSF Plant Genome Research Program, July 15 2016 - July 142019**, $2,983,307**.
  - Genome Database for Rosaceae: Empowering Specialty Crop Research through Big-Data Driven Discovery and Application in Breeding (PI Main) USDA NIFA Specialty Crop Research Initiative, Sept 1 2014 – Aug 31 2019, **$2,700,000**.
  - CIF21 DIBBS: Tripal Gateway, a platform for next-generation data analysis and sharing" (PI Ficklin, CoPI Main) NSF CIF21 Program, Jan 2015 – Dec 31 2017, **$1,498,066**
  - US Land Grand Universities – participating scientist salaries
- Community of developers working on Tripal Database modules

AgBioData Outreach

- AgBioData consortium workshop (funded by NSP to PI Main)

- Monthly conference calls for the last 2 years
- In person yearly meeting at PAG
- AgBioData communication website developed

- Funded by NSF award to PI Main
- 43 scientists participated representing over 20 databases
- Whitepaper on Agricultural Biological Databases by August
- Opinion papers following the whitepaper

NRSP10 Workshop at ASHS

- An NRSP10 Workshop has been organized for September, 2017 at annual ASHS meeting
- Focuses on presentations by users of NRSP10 resources
- Discussion on users needs and future of NRSP10

(id=25937) What Is NRSP10 and What Can It Do for Specialty Crop Research . D. Main

(id=25935) NRSP10 Resources for Translational Tree Fruit Research . C. P. Peace

(id=25936) Using Tripal As a Project Database . M. Staton

(id=25900) NRSP10 Resources for Small Fruit Research . N. Bassil

(id=25901) Using Gensas for Specialty Crop Commmunity Genome Annotation . J. Humann

(id=25899) Using Fieldbook and Other Free Apps for Specialty Crop Field and Lab Data Collection . T. Rife

(id=25898) A Breeders Perspective on Using the Breeding Information Management System for Fruit Breeding . K. Gasic

(id=25939) Open Discussion on Direction of NRSP10 . M. Kahn

NRSP 10 Future

- We are investigating further avenues for sustainability of the NRSP10 resource - **Stakeholder driven and supported, including the possibility of commercialization (e.g. GenSAS)**
- **However** given (1) the uncertainty of current US funding and (2) the fact that our major grants all end in July-Oct 2019, we would like to explore the possibility of renewing NRSP10 for another 5 years from 2018 instead of 2019 – **will allow us retain some of our highly skilled core team**

The Team

and years working in the Mainlab

10

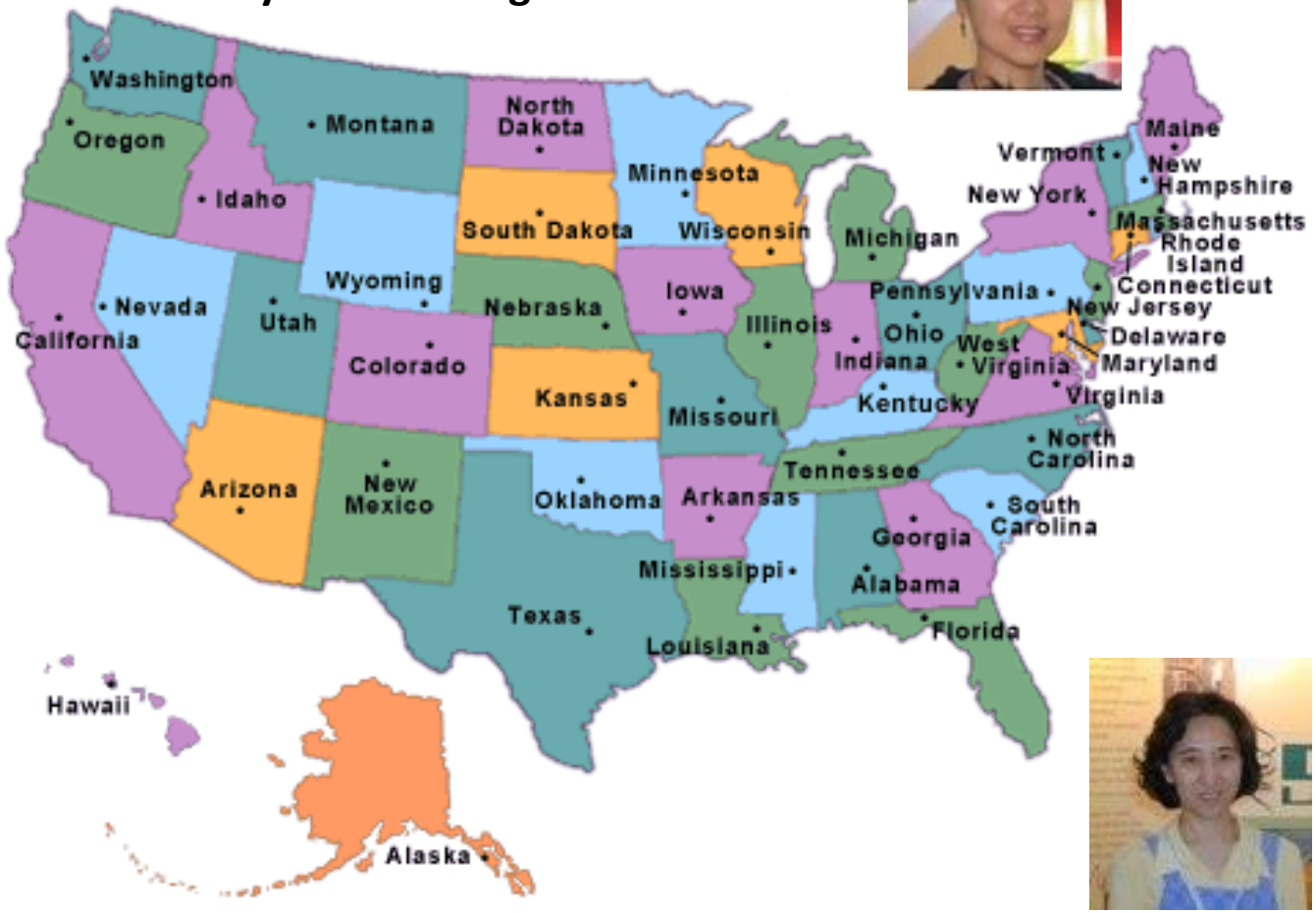
7

11

6

1

15



7

2

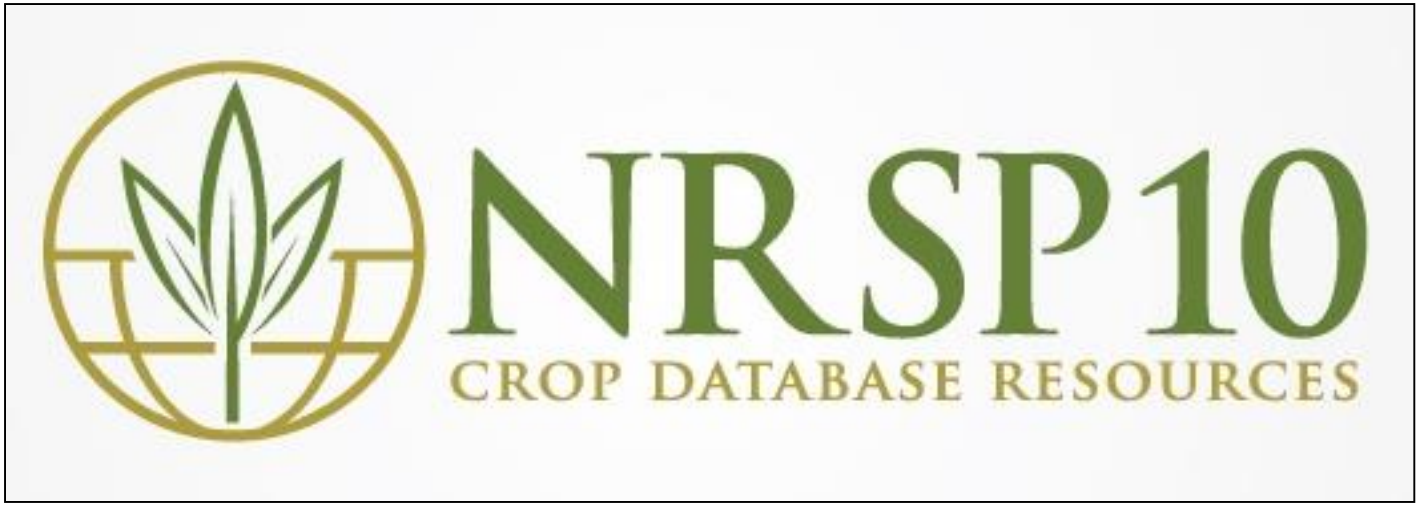
2

2

2

2

Thanks for your attention and support



Acknowledged with Thanks

- Mainlab Bioinformatics Team
- Project coPIs and associated project Pis/CoPIs
- Rosaceae, Citrus, Cacao, Blueberry, Legume, Cotton and Bioinformatics Communities
- USDA NIFA SCRI, NSF Plant Genome Program, USDA-ARS, SAAEDS, Mars Inc, Washington Tree Fruit Research Commission, Cotton Incorporated, USA Dry Pea and Lentil Commission, Northern Pulse Growers
- US Land Grant University researchers and extension agents