

# NRSP10 Progress in the last 12 months

## (1) Tripal Progress

- Core Tripal releases:
  - Tripal v2.1 released on April 1, 2017
  - Tripal v3.0 beta 2 released February 10, 2017
  - Tripal v2.1-beta 3 released February 10, 2017
- Tripal Extension module releases:
  - Mainlab chado module to load sequence, map, marker, QTL, genotype, phenotype and germplasm released November 11, 2016
  - Mainlab chado module to search interfaces for sequence, map, marker, QTL, genotype, phenotype and germplasm released November 11, 2016
  - Mainlab Tripal module to display sequence, map, marker, QTL, genotype, phenotype and germplasm released November 11, 2016
  - Several extension modules developed and released by the larger Tripal community
  - New Tripal MapViewer v1.0 beta developed and made available on the CSFL and CGD Databases.
  - New SNP marker search, SNP genotype search, and Reference transcriptome search modules developed
- Training and support for Tripal: monthly conference calls, 2 day hackathon and Tripal workshop held at Plant and Animal Genome Conference, 258 correspondences through the Tripal support mailing list, tripal.info website kept up-to-date (Figure 1)

Tripal now being used for more than 90 species/clade databases

The screenshot displays the Tripal website interface. At the top, a navigation bar includes links for Home, Download, Support, Extensions, API, About, Demo Site, and Log In. The main content area is divided into several sections:

- Quick Links:** A list of links for v2.x and v3.x user guides, installation instructions, and documentation.
- Compatible With...:** Logos for CHADO, JBrowse, and the Galaxy Project.
- Tripal Logo and Description:** The Tripal logo is followed by a paragraph explaining that Tripal is a toolkit for constructing online biological databases, integrated with GMOD Chado and Drupal.
- Objectives:** A list of two main goals: providing a framework for publishing biological data and using community standards.
- Sites Using Tripal:** A section titled "Cool Season Food Legume Database" showing a preview of a database interface with search and navigation options.
- News:** A recent announcement: "Tripal v2.1 released 04/01/2017 - 21:35".

Figure 1: Tripal Informational Site

## (2) NRSP10 Databases

- Usage in 2016 (by Google Analytics)
  - [GDR](#) - 19,982 users from 149 countries, 245,787 pages viewed
  - [CottonGen](#) - 10,042 users from 145 countries, 155,207 pages viewed
  - [CSFL](#) - 3,134 users from 116 countries, 24,958 pages viewed
  - [CGD](#) - 4,414 users from 121 countries, 23,943 pages viewed
  - [GDV](#) - 1,560 users from 83 countries, 6,393 pages viewed
- Training and outreach activities include presentations at several international, national and local meetings. These include holding user-taught training workshops and presentations at the 8<sup>th</sup> International Rosaceae Genomics Conference, the Plant and Animal Genome Conference, The 5<sup>th</sup> International Research Conference on Huanglongbing, Cotton Beltwide Conference etc., as well as webinars, brochures, posters. All presentations/webinars available from database websites.
- The Genome Database for Rosaceae (GDR, Figure 2) and the Citrus Genome Databases (CGD, Figure 3) were converted to Tripal2 and Drupal7 with uniform NRSP10 design, making 4 of the 5 databases fully upgraded. The Citrus database has been expanded to include information specifically relevant to HLB research. The Genome Database for Vaccinium (GDV) has been redesigned in dev and will be made available publicly in May 2017. A significant volume of new data has been added to all of the 5 databases.

The screenshot shows the homepage of the Genome Database for Rosaceae (GDR). The header features the GDR logo and navigation menus for Species, Data, Search, Tools, General, Help, and Community, along with a Login button and a search bar. The main content area is divided into two columns. The left column is titled 'GENOME DATABASE FOR ROSACEAE' and includes a row of fruit icons and the tagline 'Resources for Rosaceae Research Discovery and Crop Improvement'. The right column is titled 'News and Events' and lists several recent updates, such as reference transcriptome availability for Malus x domestica, Prunus avium, and Prunus persica, and GDR's participation in the PAG 2017 conference. Below the main content, there are two sections: 'Major Genera Quick Start' and 'Tools Quick Start'. The 'Major Genera Quick Start' section features six icons representing different genera: Fragaria, Malus, Prunus, Pyrus, Rosa, and Rubus. The 'Tools Quick Start' section is organized into three columns: Genomics (View Genomes, Find Sequences, Search Genes, BLAST Sequences), Genetics (Browse Maps, Search Markers, Find QTLs, Compare Maps), and Breeding (Search Phenotype, Search Genotype, Search Haplotype, Decision Tools). The footer contains information about the database's development by Mainlab Bioinformatics at Washington State University, funding from USDA, NSF, and industry, and contact information.

Figure 2: Genome Database for Rosaceae in Tripal 2/Drupal 7

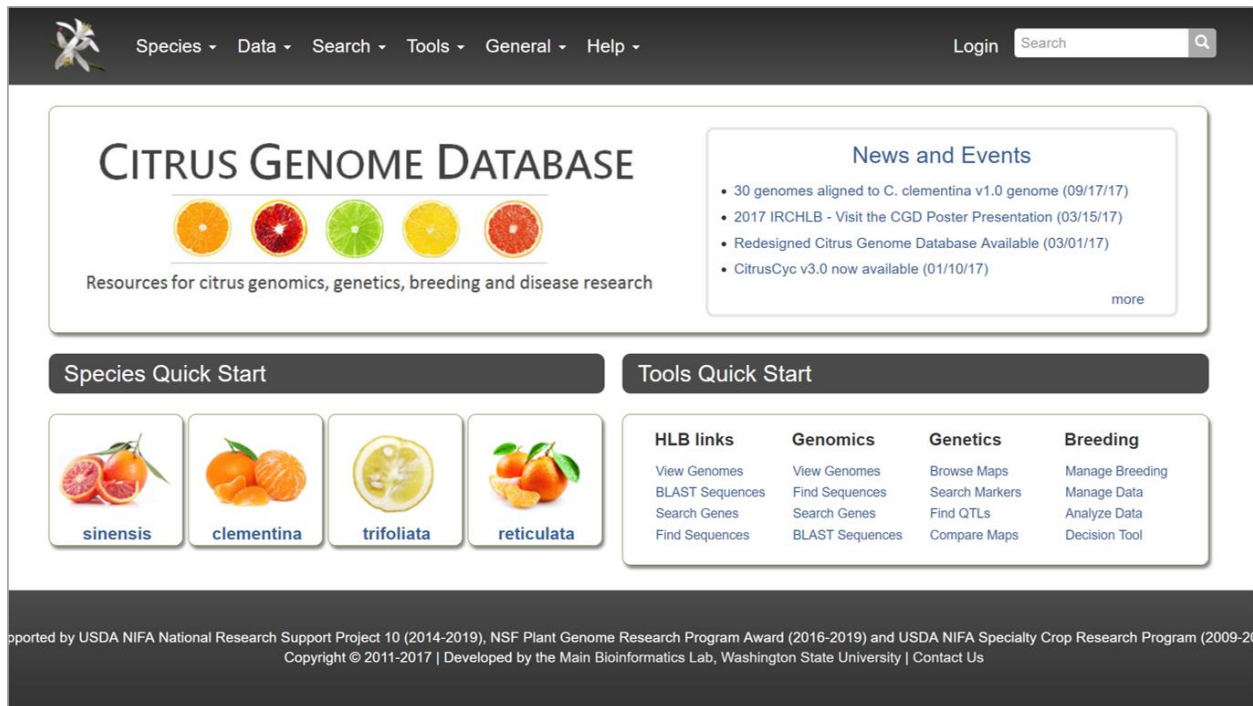


Figure 3: Citrus Genome Database in Tripal2/Drupal 7

### (3) Development of Breeding Tools (TripalBIMS)

- Data in BIMS:
  - 11 years (2004 to 2014) of breeding data from Todd Campbell's cotton program (USDA-ARS South Carolina) has been converted from AgroBASE and loaded to BIMS in CottonGEN.
  - 3 years (2011-2013) of breeding data from Ksenija Gasic's peach breeding program (Clemson University) has been loaded to BIMS in GDR.
- BIMS development (Figure 4)
  - Functionality to view and download breeding data by cross population
  - Functionality to generate the input files for Field Book App
  - Functionality to upload trait descriptor data has been updated to match the trait categories of the Field Book App.
- Field Book App
  - Development on Field Book has primarily focused on adding user-requested features and patching user-reported bugs. A new trait format, 'Location', was added to facilitate collection of location point data. A button was added to the main screen for missing values to help breeders distinguish between missing data and missing entries. Users can now load files directly from Dropbox, eliminating a file transfer step and streamlining the data collection process. Photos now also include the name of the trait to help researchers know better what they're looking at. A dedicated Android programmer was hired in the Poland lab in January who is learning how the apps work and working on

rewriting parts of the apps to fix bugs, increase efficiency, and better-adhere to best programming practices.

- Handheld Samsung tablets with Field Book have been provided to more than 40 NRSP10 associated breeders and researchers to test and use. Feedback critical in informing development through the Poland lab (provided with funds for a ½ time developer through NRSP10 since Sept 1, 2017)
- Field Book and BIMS webinar held in November, 2016. FieldBook and BIMS presented at several conferences.

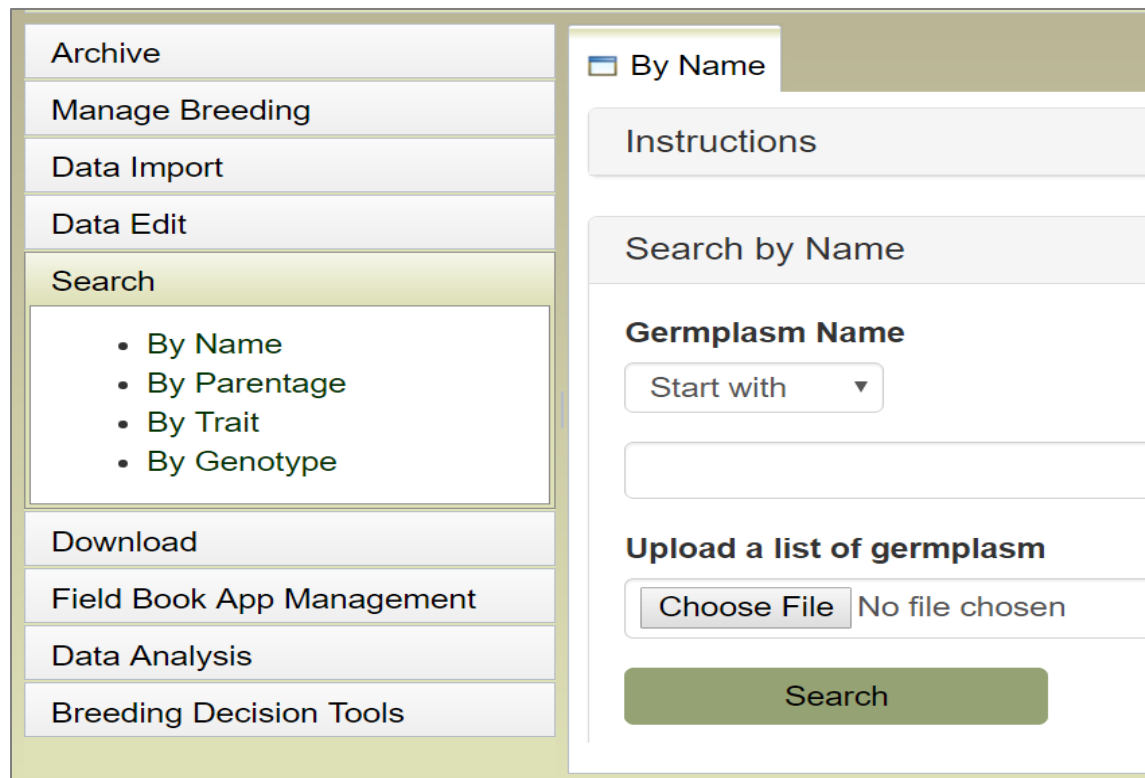


Figure 4: Tripal BIMS

#### (4) Genome Sequence Annotation Server (GenSAS)

- GenSAS v5.0 released in January 2017 (Figure 5). Major improvements completed include; (1) ability to upload sequences before project creation so sequence subsets can be created from multiple-sequence fasta files. Sequence subsets can also be filtered by sequence name or minimum size (2) ability to upload RNA-seq reads use them to train gene model predictors (Augustus, during structural step) (3) addition of the tool Tophat to enable alignment specifically for the RNA-Seq reads. Alignment tools (blast, blat, PASA, tophat) are now included as a step before structural annotation (4) structural annotation (previously labelled as the “genes” step) now also has GeneMark for prokaryotic and eukaryotic gene prediction and (5) an official genes step (OGS) was added where users can either use EvidenceModeler to create a genes consensus and use it as the OGS, or can select another data track (from gene prediction program or

alignment) as the OGS. Manual curation is merged with the OGS at the end of the GenSAS protocol.

- GenSAS was demonstrated/presented at several conferences. In Year 2 GenSAS was been accessed by 1,135 visitors from 73 countries, with 2,333 sessions.

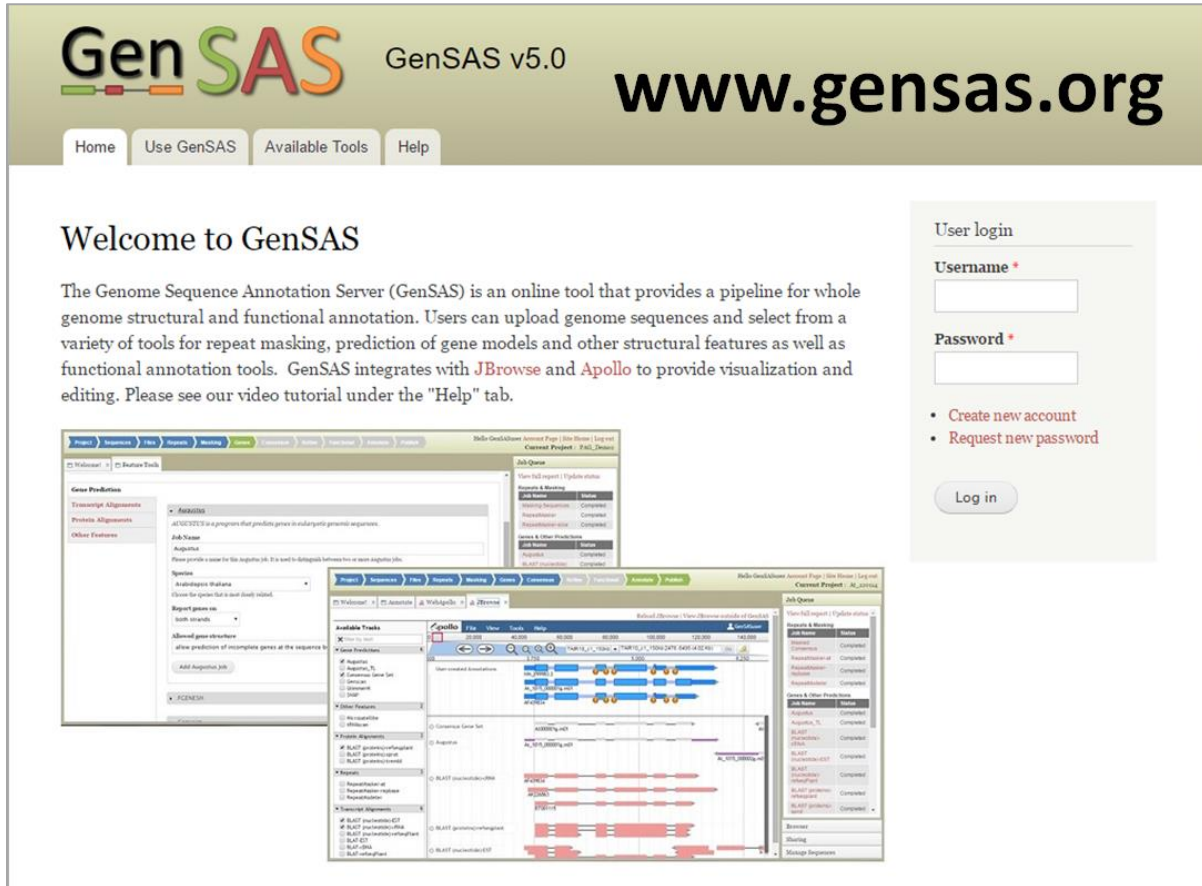


Figure 5: GenSAS Home Page

(5) New Supporting Funding (~ \$4M)

- “Standards and CyberInfrastructure that Enable Big-Data Driven Discovery for Tree Crop Research” (Role – PI) NSF Plant Genome Research Program, July 2016-July 2018, \$2,983,307.
- “Further Development of CottonGEN: A Genomics, Genetics and Breeding Database for Cotton Improvement” (Role - PI). Cotton Incorporated and USDA-ARS, January 1, 2017 – December 31, 2020, \$741,889.
- “Enhancing the sustainability of lentil production in Washington” (Role – CoPI), Washington Specialty Crop Block, \$199,406. October 1 2016 – September 31 2019.
- “Harvesting the Lentil Genome” (Role – CoPI). Northern Pulse Growers Association, July 1 2016 – June 30 2017, \$35,467.

(6) Major Outreach Activities for the remainder of 2017

- Peer-review publications for GenSAS, GDR, CottonGen, CSFL and CGD



- Provide complete copy of CottonGen database (less data) to the SCRI funded CarrotREP project for use as a project database
- Organization of 2 day AgBioData workshop (April 19-20, 2017) (Figure 6)



Figure 6: AgBioData website and workshop

- Holding NRSP10 Workshop at ASHS 2017 (Table 1) – focused on user presentations by users and covering application for specialty crops

<a href="#">(id=25937) What Is NRSP10 and What Can It Do for Specialty Crop Research</a> . <b>D. Main</b>
<a href="#">(id=25935) NRSP10 Resources for Translational Tree Fruit Research</a> . <b>C. P. Peace</b>
<a href="#">(id=25936) Using Tripal As a Project Database</a> . <b>M. Staton</b>
<a href="#">(id=25900) NRSP10 Resources for Small Fruit Research</a> . <b>N. Bassil</b>
<a href="#">(id=25901) Using Gensas for Specialty Crop Community Genome Annotation</a> . <b>J. Humann</b>
<a href="#">(id=25899) Using Fieldbook and Other Free Apps for Specialty Crop Field and Lab Data Collection</a> . <b>T. Rife</b>
<a href="#">(id=25898) A Breeders Perspective on Using the Breeding Information Management System for Fruit Breeding</a> . <b>K. Gasic</b>
<a href="#">(id=25939) Open Discussion on Direction of NRSP10</a> . <b>M. Kahn</b>

Table 1: NRSP10 Workshop Presentations scheduled for ASHS 2017