# Database Resources for Crop Genomics, Genetics and Breeding Research

**2014 SAAESD Spring Meeting**  
Savannah, GA

## Administrative Advisors
- Susan Brown (NE)
- Steven Lommel (S)
- Jim Moyer (W – Main)
- Karen Plaut (NC)

## Writing Team
- Dorrie Main (WSU)
- Sook Jung (WSU)
- Mike Kahn (WSU)
- Cameron Peace (WSU)
- Jim McFerson (WTRC)
- Reviewers (US Wide)
The Team
Presentation Outline

• What is a database?
• Types of genomic databases
• Community databases
  • Importance
  • Challenges
  • Proposed Solution (Tripal)
    • Why Tripal
    • Current Status
    • Future Direction
• This proposal
  • Our databases (underserved crops)
  • Budget
  • Sustainability model
Genome Databases

- Primary Databases – NCBI, EMBL, DDJB
- Secondary Databases – Pfam, PDB
- Tertiary Databases
  - Comparative Genomics Databases
  - Community Databases
Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | Research | NCBI News

Get Started

- Tools: Analyze data using NCBI software
- Downloads: Get NCBI data or software
- How-Tos: Learn how to accomplish specific tasks at NCBI
- Submissions: Submit data to GenBank or other NCBI databases

NCBI Facebook page

Find out the latest news about NCBI resources and participate in community discussions.

Popular Resources
- PubMed
- Bookshelf
- PubMed Central
- PubMed Health
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI Announcements

New NCBI YouTube video: Create custom databases for BLAST

Mar 28, 2014

In the newest NCBI video on YouTube, we show you how to create custom databases.

Come to the NCBI Discovery Workshops on May 6th & 7th!

Mar 28, 2014
Molecule of the Month

Neurotransmitter Transporters

Nerve cells communicate with one another in two ways. Some neurons send an electrical signal directly to their neighbors, which is very fast. Most neurons, however, use chemical signals to transmit their messages, releasing small neurotransmitter molecules that are recognized by receptors on neighboring neurons. Neurotransmitters have two important advantages: since thousands of molecules are released, they amplify the signal, and since many different types of neurotransmitters are used, they can encode a variety of different types of signals.

Full Article

Protein Structure Initiative Featured System

Glucacon Receptor

(Protein-coupled receptors) come in many sizes and shapes, each recognizing its own type of
Genome Databases

- Primary Databases – NCBI, EMBL, DDJB
- Secondary Databases – Pfam, PDB
- Tertiary Databases
  - Comparative Genomics Databases
  - Community Databases
Gramene: A comparative resource for plants

Gramene is a curated, open-source, integrated data resource for comparative functional genomics in crops and model plant species. Our goal is to facilitate the study of cross-species comparisons using information generated from projects supported by public funds. Gramene currently hosts annotated whole genomes in over two dozen plant species and partial assemblies for almost a dozen wild rice species in the Ensembl browser, genetic and physical maps with genes, ESTs and QTLs locations, genetic diversity data sets, structure-function analysis of proteins, plant pathways databases (BioCyc and Plant Reactome platforms), and descriptions of phenotypic traits and mutations.

Gramene Portals

- **Genome Browser**: Browse gene annotations & diversity data
- **BLAST**: Align DNA & protein sequences
- **Plant Reactome**: Browse metabolic & regulatory pathways
- **Pathways databases**: BioCyc based cellular metabolic networks for 10 plant species
- **Cancer Markers**: Regulatory data on microRNAs

More
Welcome to Phytozome!

Phytozome is a joint project of the Department of Energy's Joint Genome Institute and the Center for Integrative Genomics to facilitate comparative genomic studies amongst green plants. Families of orthologous and paralogous genes that represent the modern descendents of ancestral gene sets are constructed at key phylogenetic nodes. These families allow easy access to clade specific orthology/paralogy relationships as well as clade specific genes and gene expansions. As of release v9.1, Phytozome provides...
Genome Databases

• Primary Databases – NCBI, EMBL, DDJB

• Secondary Databases – Pfam, PDB

• Tertiary Databases
  • Comparative Genomics Databases
  • Community Databases
Why Do We Need Community Databases?

• To organize, store, curate, integrate and disseminate associated genomic, genetic and breeding data

• To provide **centralized access** to data for basic, translational and applied researchers.

• To provide **data mining opportunities** via intuitive online tools.

• To provide **data sharing and communication opportunities** (community building)
Integrated Data Facilitates Discovery!

Basic Science
Structure and evolution of genome, gene function, genetic variability, mechanism underlying traits

Genetics
Breeding
Germplasm
Diversit
Genomics
Integrated Data & Tools

Translational Science
QTL /marker discovery, genetic mapping, Breeding values

Applied Science
Utilization of DNA information in breeding decisions
MaizeGDB is a community-oriented, long-term, federally funded informatics service to researchers focused on the crop plant and model organism Zea mays.

Maize Genome Sequencing

<table>
<thead>
<tr>
<th>Line</th>
<th>Genome Browsing</th>
<th>Other Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>B73</td>
<td>MaizeGDB Genome Browser</td>
<td>Sequence, Information, Publication</td>
</tr>
<tr>
<td>Mo17</td>
<td>phytozome</td>
<td>Sequence, Publication</td>
</tr>
<tr>
<td>Palomero</td>
<td></td>
<td>Sequence, Publication</td>
</tr>
</tbody>
</table>

For more information, please visit our [sequence resource page](http://maizegenome.org).

Be on Top of Recent Advances!

- Maize Meeting
- Hot New Papers
- Newly Characterized Genes
- Maize Gene Review

Project Oversight

MaizeGDB is guided by members of the community of maize geneticists.

Funding Sources

USDA
**Featured Tool on GrainGenes**

**CMap.** CMap is the official map display for GrainGenes 2.0. Created by the Gramene project. CMap is a powerful tool for comparative mapping by visualizing the position of common markers on multiple maps in the GrainGenes collection. Visit the CMap page at [http://wheat.nw.usda.gov/cmap/](http://wheat.nw.usda.gov/cmap/).

**List of all Featured Tools**

**Hot Topics**
- €1M from Bayer CropScience to finish wheat physical map
- Genetic markers on the barley physical map - BLAST or search by name
- Wheat Transcriptome BLAST
- Ae. tauschii physical/functional map
- Wheat improvement: The Truth Unveiled
- Wheat initiative - an international research initiative for wheat improvement
- Genome sequences:
  - Barley
  - Morex barley 24x
  - T. urartu draft
  - Ae. tauschii draft
  - T. aestivum 5x
- Morex barley DH map and GBS data
- CerealsDB: Wheat SNPs, 5x genome sequence
- USDA wheat SNP database
- T3, The Triticeae Toolbox
- phenotyp genotype database
- 53 barley datasets at PLEXdb
- Wheat SK SNP assay available
- Wheat Mapping Populations
- European project TriticeaeGenome
- 2008 Wheat Gene Catalogue
- TRPP Release 10
- Barley QTl Community Curation Workbook
- Brassica website
- More...

**Meeting Announcements**
- Borlaug Summit on Wheat for Food Security, 25 Mar
- EUCARPIA Cereals Section - I TMI Joint Conference, 29 Jun
Entry Points

**Sequence Search**
- Search sequenced legume genomes »
- Search genes/proteins/gene families »
- Advanced sequence search »

**Species Access and Data**
Information, resources, and datasets for many legume species.

**Leggle Gene Family Pages**
Searchable gene families, alignments, and gene trees for the legumes and outgroup species.

**CVIT Genome Search & Synteny Views**
Search for sequences or genes and view against whole-genome views and synteny plots of three legume reference genomes.

**Genome Browsers**
Chromosome- to gene-scale views of the *Lotus japonicus*, *Medicago truncatula*, *Cajanus cajan* (pigeonpea), and *Glycine max* (soybean) genomes.

Welcome

The mission of LIS is to facilitate discoveries and crop improvement in the legumes — critical components of global agriculture. Please send us your questions, suggestions, and ideas for improving this site! lis_feedback@ncgr.org

Meetings

International Conference on Legume Genetics and Genomics (JCLGG)
Jul 07-11, 2014

Recent News

January 24, 2014 at 15:57 MST -- Open postdoctoral position in computational and plant evolutionary biology

The project involves...

Read More »

November 11, 2013 at 16:49 MST -- Cowpea Consortium SNP Genotyping Chip

Researchers at the University of California, Riverside,...
The Arabidopsis Information Resource

The Arabidopsis Information Resource (TAIR) maintains a database of genetic and molecular biology data for the model higher plant Arabidopsis thaliana. Data available from TAIR includes the complete genome sequence along with gene structure, gene product information, gene expression, DNA and seed stocks, genome maps, genetic and physical markers, publications, and information about the Arabidopsis research community. Gene product function data is updated every week from the latest published research literature and community data submissions. TAIR also provides extensive linkouts from our data pages to other Arabidopsis resources.

The Arabidopsis Biological Resource Center at The Ohio State University collects, reproduces, preserves and distributes seed and DNA resources of Arabidopsis thaliana and related species. Stock information and ordering for the ABRC are fully integrated into TAIR.

TAIR is located at the Carnegie Institution for Science Department of Plant Biology and funded by the National Science Foundation.

Beginning on October 1, 2013, access to TAIR by researchers at companies or for-profit institutions requires a subscription. Academic / non-profit access to TAIR will require a subscription after April 1, 2014. ABRC catalog, stock and ordering pages hosted at TAIR will remain accessible without a subscription. Please see our subscription page for further details.

Chinese National Science and Technology Library (NSTL) to subscribe to TAIR

通知

致中国用户的通知-2

Subscribe to news feed
Follow our Twitter feed
Join our Facebook group

Breaking News

Update on NSTL subscription for China
[March 28, 2014]
We have reached a tentative agreement with NSTL for a national subscription covering all mainland Chinese academic and non-profit institutions. While we finalize the details we will keep TAIR available to access from China.

Notice to Chinese users
[March 12, 2014]
We asked all TAIR users in China to fill out a survey. The results have been sent to the Chinese National Library of Science and Technology (NSTL), which is considering whether to move ahead with a national TAIR subscription covering all Chinese academic institutions.

ABRC tool for adding comments about stocks
[January 29, 2014]
We are encouraging users to add comments about stocks they have ordered. We are collecting information such as validation of insert/sequence,
Welcome to the Genome Database for Rosaceae

Initiated in 2003, the Genome Database for Rosaceae (GDR) is a curated and integrated web-based relational database providing centralized access to Rosaceae genomics, genetics and breeding data and analysis tools to facilitate basic, translational and applied Rosaceae research. GDR has been supported by the NSF Plant Genome Program, the USDA Specialty Crop Research Initiative, the Washington Tree Fruit Research Commission, Clemson University and Washington State University.

What's new in GDR?

- GDR has been redesigned! Check out our new search sites for gene, sequence, marker, molecular diversity data as well as new data pages for gene, sequence, marker, molecular diversity (09/19/13)
- New sequence retrieval tool available to enable retrieval of genes and sequences in fasta format, with added upstream and downstream sequence feature (09/15/13)

General Information

Report a problem | Ask us a question | Post a job | Post a meeting or event | Used GDR data or tools in your research - how to reference us.
Community Databases Even More Important!

Recent advances in sequencing, genotyping, and phenotyping technologies have led to a paradigm shift in crop science research.

**Individual scientists now routinely**

- Sequence and genotype genomes from populations, families, individuals of interest
- Pursue large-scale gene expression studies
- Create highly saturated genetic maps
- Identify loci influencing traits of interest
- Conduct large-scale standardized phenotyping.
Challenges for Community Databases

• Largely using legacy systems
  = difficult to add new data types
  = difficult to implement for other species.
  = generally resource inefficient

• Issues of data quality, storage, speed of querying, standardizing phenotyping, ontology associations

• Can not expect long term funding by NSF or USDA

• Need to develop sustainable funding models for underserved crops
Proposed Database Solution - Tripal

• Develop a common database platform that is open-source, efficient, flexible, modular and easy to implement, manage and use.

• Reviewed existing solutions and decided to further develop Tripal, a toolkit for building online biological databases that was initiated at Clemson University in 2008 (Stephen Ficklin - WSU and Meg Staton - University of Tennessee)

• Tripal utilizes Drupal and Chado, open-source software environments for content management and database construction.
Database Structure

- **Drupal**: Content Management System
- **Tripal**: Drupal modules as web front-end for Chado
- **Chado**: Generic Database schema
Database tool

The Chado Natural Diversity module: a new generic database schema for large-scale phenotyping and genotyping data

Sook Jung¹,*†, Naama Menda²,*†, Seth Redmond³,‡, Robert M. Buels², Maren Friesen⁴, Yuri Bendana⁴, Lacey-Anne Sanderson⁵, Hilmar Lapp⁶, Taein Lee¹, Bob MacCallum³, Kirstin E. Bett⁵, Scott Cain⁷, Dave Clements⁶,*⁷, Lukas A. Mueller² and Dorrie Main¹

¹Department of Horticulture and Landscape, Washington State University, Pullman, WA 99164, ²Boyle Thompson Institute for Plant Research, Ithaca, NY 14853, USA, ³Imperial College London, London SW7 2AZ, UK, ⁴University of Southern California, Los Angeles, CA 90089, USA, ⁵Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, S7N 5A8, Canada, ⁶National Evolutionary Synthesis Center (NEStCent), Durham, NC, USA and ⁷Ontario Institute for Cancer Research, Toronto, Ontario, M5G 0A3, Canada

*Corresponding author: Tel: 509-335-7093; Fax: 509-335-8660; Email: sook_jung@wsu.edu/ Correspondence may also be addressed to Naama Menda. Tel: 607-254-3569; Fax: 607-254-1242; Email: naama.menda@cornell.edu
Why Choose Drupal?
Use Drupal to build everything from personal blogs to enterprise applications. Thousands of add-on modules and designs let you build any site you can imagine.

Drupal is free, flexible, robust and constantly being improved by hundreds of thousands of passionate people from all over the world.

Join us!

Sites Made with Drupal

Pagebuild Case Study
Drupal is used by some of the biggest sites on the Web, like The Economist, Examiner.com and The White House. Read more Drupal success stories.

Develop with Drupal
14,121 Modules
1,363 Themes
14,553 Developers

This week
3,250 Code commits
6,506 Issue comments

Drupal Core
Security Info
Developer Docs
API Docs

Join the Drupal Association
Help build a successful ecosystem around Drupal.
Original article

Tripal v1.1: a standards-based toolkit for construction of online genetic and genomic databases

Lacey-Anne Sanderson¹,†, Stephen P. Ficklin²,†, Chun-Huai Cheng², Sook Jung², Frank A. Feltus³, Kirstin E. Bett¹ and Dorrie Main²,⁎

¹Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK Canada, ²Department of Horticulture, Washington State University, Pullman, WA, USA and ³Department of Genetics and Biochemistry, Clemson University, Clemson, SC, USA

⁎Correspondence author: Tel: +1 509 335 9000; Fax: +1 509 335 8690; Email: dorrie@wsu.edu

†These authors contributed equally to this work.

Submitted 26 June 2013; Revised 9 September 2013; Accepted 27 September 2013

**Tripal**

Tripal is a collection of open-source freely available Drupal modules and is a member of the GMOD family of tools. Tripal serves as a web interface for the GMOD Chado database and is designed to allow anyone with genomic data to quickly create an online genomic database using community supported tools. Tripal is licensed under the GNU General Public License version 2.

Sites currently using Tripal:

- Banana Genome Hub
- Cacao Genome Database
- Citrus Genome Database
- Cool Season Food Legume Genome Database
- CottonGen
- Fagaceae Genomics Web
- Genome Database for Rosaceae
- Genome Database for Vaccinium
- Hardwood Genomics Project
- KnowPulse: Pulse Crop Genomics & Breeding
- Legume Information System
- PeanutBase

Send an email to the Tripal mailing list if you would like your Tripal/Drupal/Chado site listed above!

**To cite Tripal:**


Tripal Timeline

• 2008: Tripal was used for development of the Marine Genomics Network and the Fagaceae Genomics Network. Clemson University

• 2008 – 2011: Development of the Cacao Genome Database ($435K from USDA-ARS/MARS Inc. WSU

• 2008-2013: Development of the Citrus Genome Database and conversion of the Genome Database for Rosaceae to Tripal (~$4 m from USDA NIFA SCRI Program, WA Tree Fruit Research Commission, Florida Citrus Research Commission, WSU, UF and Clemson)
Tripal Timeline

- From 2010: Development of the Cool Season Food Legume Database ($48 – 100 K from USA Dry Pea and Lentil Council) WSU

- From 2009: Development of the KnowPulse Database. University of Saskatchewan

- 2011 – 2016: Development of CottonGen ($835K from Cotton Incorporated, USDA-ARS, Southern Association of Experiment Station Directors, Monsanto, Dow, Bayer)

- From 2011: Development of the Genome Database for Vaccinium ($20K from NC State). WSU, NCSU, UF
Tripal Timeline

- From 2013: Development of the Peanut database and the common bean database, conversion of the Legume Information System, Iowa State, NCGR
- 2014: 26 databases now using Tripal
Converting to Tripal
Converting to Tripal
Converting to Tripal
Arabidopsis Information Portal Implemented in Tripal
Considering implementing a Tripal Instance
### Other Confirmed Tripal Databases

<table>
<thead>
<tr>
<th>Site</th>
<th>Species</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Arabidopsis Information Portal</td>
<td>Arabidopsis</td>
<td>Rockville MD, USA</td>
</tr>
<tr>
<td>2. Cacao Genome Database</td>
<td>Cacao matina</td>
<td>Ames IA, USA</td>
</tr>
<tr>
<td>3. PeanutBase</td>
<td>Arachis spp</td>
<td>Ames IA, USA</td>
</tr>
<tr>
<td>4. Legume Information System</td>
<td>various legumes</td>
<td>Ames IA, USA</td>
</tr>
<tr>
<td>5. i5K Workspace @ USDA NAL</td>
<td>30 insect genomes</td>
<td>Beltville, MD USA</td>
</tr>
<tr>
<td>6. Fagaceae Genomics Web</td>
<td>Fagaceae spp</td>
<td>Clemson SC, USA</td>
</tr>
<tr>
<td>7. MarineGenomics.org</td>
<td>various species</td>
<td>Clemson SC, USA</td>
</tr>
<tr>
<td>8. GeneNet Engine</td>
<td>various species</td>
<td>Clemson SC, USA</td>
</tr>
<tr>
<td>9. PeanutBase</td>
<td>Arachis spp</td>
<td>Ames IA, USA</td>
</tr>
<tr>
<td>10. Banana Genome Hub</td>
<td>Arachis spp</td>
<td>Ames IA, USA</td>
</tr>
<tr>
<td>11. Hardwood Genomics</td>
<td>various species</td>
<td>Clemson SC, USA</td>
</tr>
<tr>
<td>12. Fragaria x ananassa strawberry</td>
<td>strawberry</td>
<td>Malaga, Spain</td>
</tr>
<tr>
<td>13. NECC Little Skate Gnome</td>
<td>Leucoraja erinacea</td>
<td>Newark, DE</td>
</tr>
<tr>
<td>14. LiceBase</td>
<td>Salmon louse</td>
<td>Norway</td>
</tr>
<tr>
<td>15. Wild Strawberry</td>
<td>Fragaria</td>
<td>OSU Orgeon, USA</td>
</tr>
<tr>
<td>16. Chlamydomonas database</td>
<td>Chlamydomonas</td>
<td>Palo Alto, CA USA</td>
</tr>
<tr>
<td>17. Amborella Genome</td>
<td>Amborella trichopoda</td>
<td>PennState PA/Athens GA, USA</td>
</tr>
<tr>
<td>18. Ruditapes decussatus db</td>
<td>Ruditapes decussatus</td>
<td>Portugal</td>
</tr>
<tr>
<td>19. Know Pulse</td>
<td>various legumes</td>
<td>Saskatoon SK, Canada</td>
</tr>
<tr>
<td>20. Koala Genome Consortium</td>
<td>Phascolarctos cinereus</td>
<td>Sydney Australia</td>
</tr>
</tbody>
</table>
Vision

• Enable basic, translational and applied crop research by expanding existing online databases currently housing high-quality genomics, genetics and breeding data for Rosaceae, Citrus, Cotton, Cool Season Food Legumes and Vaccinium crops.

• Provide a complete open-source, flexible, database solution for other organisms.

• Develop a model for long term sustainability of community databases.
• Crops annual production value in 2012 = $12.6 B
• Database established 2003 (NSF, USDA, Industry, University)
• **14,237 users** (from 52 US States/territories, 130 countries)
  176,259 pages accessed
• Crops annual production value in 2012 = $3.44 B

• Database established 2009 (NSF, USDA, Industry, University)

• 5,244 users (from 49 US states/territories, 125 countries) 34,475 pages accessed
Crops annual production value in 2012 = $5.97 B

Database established 2011 (NSF, USDA, Industry, University)

2,320 users (from 43 US states, 74 countries) 46,279 pages accessed
Welcome to CottonGen

CottonGen is a new cotton community genomics, genetics and breeding database being developed to enable basic, translational and applied research in cotton. It is being built using the open-source Tripal database infrastructure. CottonGen will initially consolidate the data from CottonDB and the Cotton Marker Database, which includes sequences, genetic and physical maps, genotypic and phenotypic markers and polymorphisms, QTLs, pathogens, germplasm collections and trait evaluations, pedigrees, and relevant bibliographic citations. It will be expanded to include annotated transcriptome, genome sequence, marker-trait-locus and breeding data, as well as enhanced tools for easy querying and visualizing research data. It will facilitate community communication and collaboration through housing the ICGI website, forums and mailing lists. This project is funded by Cotton Incorporated, the USDA-ARS Crop Improvement program, and the Cotton Research and Development Institute.
• Crops annual production value in 2012 = $0.4 B

• Database established 2003 (NSF, USDA, Industry, University)

• **2,273 users** (from 50 US states, 101 countries) 11,009 pages accessed
• Crops annual production value in 2012 = $1.23B

• Database established 2003 (NSF, USDA, Industry, University)

• **1,120 users** (from 45 US states, 84 countries) 5,898 pages accessed
# Phenotyping Data Search by Varieties

**Choose Crop**
- Sweet Cherry (RosBREED)

**Choose Breeding Group**
- SWEET CHERRY CRS/BPS WA

**Names of Varieties**
- exactly

**Upload List of Varieties**
- Choose File - No file chosen

**Restrict by Datasets**
- Sweet_Cherry_CRS_phenotyping_2011
- Sweet_Cherry_CRS_phenotyping_2010
- Sweet_Cherry_BPS_WA_phenotyping_2011
- Sweet_Cherry_BPS_WA_phenotyping_2010

*Example 1:*
- select 'exactly' and type 9814-105

*Example 2:*
- select 'start with' and type 9814
- 4.10
Phenotyping Data Search by Traits

Search by Traits

Search for varieties with a certain set of phenotypic values. You can set a range for numeric phenotypic values and choose multiple coded values by holding down the control key (see help).

Choose Crop

Choose Breeding Group

Private Data

Appearance

<table>
<thead>
<tr>
<th>Flesh_C</th>
<th>FreeStone</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: white</td>
<td>1: clingy</td>
</tr>
<tr>
<td>2: pink</td>
<td>2: --</td>
</tr>
<tr>
<td>3: orange</td>
<td>3: --</td>
</tr>
<tr>
<td>4: red</td>
<td>4: --</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pull_Force</th>
<th>Skin-C_Blush</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ Pull_Force ≤</td>
<td>1: 0-25%</td>
</tr>
<tr>
<td></td>
<td>2: 25-50%</td>
</tr>
<tr>
<td></td>
<td>3: 50-75%</td>
</tr>
<tr>
<td></td>
<td>4: 75-100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Skin_C_maho</th>
<th>SSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ Skin_C_maho ≤</td>
<td>≤ SSC ≤</td>
</tr>
</tbody>
</table>

| Stem_L | |
|--------||
| ≤ Stem_L ≤ | |
Phenotyping Data Search by Parentage

Search Varieties by Parentage

Choose Crop
Sweet Cherry (RosBREED)

Choose Breeding Group
SWEET CHERRY CRS/BPS WA

Maternal Parents
- Ambrunes
- BB
- Benton
- Bing
- BlackRepublican
- CC
- Cowiche
- DD
- EE
- EmperorFrancis
- GG
- Glacier
- Hedelfingen
- Index

Paternal Parents
- Ambrunes
- Benton
- Cashmere
- Chelan
- EmperorFrancis
- Index
- Lambert
- Lapins
- Moreau
- Napoleon
- PMR-1
- Rainier
- Regina
- Schmidt

Search
Reset
### Phenotyping Data Trait Search Example

#### Download Options

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Variety ID</th>
<th>Clone ID</th>
<th>Sample ID</th>
<th>Site</th>
<th>Position</th>
<th>Flesh C</th>
<th>FreeStone</th>
<th>Pull Force</th>
<th>Skin C</th>
<th>Mahogany</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24058</td>
<td>Roza</td>
<td>2</td>
<td>4</td>
<td>8.37</td>
<td>4.4</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24722</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>3.57</td>
<td>6.2</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>9.88</td>
<td>3</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>8.35</td>
<td>3.2</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>10.02</td>
<td>5.6</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>7.71</td>
<td>19.9</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>8.9</td>
<td>2.25</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>11.43</td>
<td>3.4</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>6.52</td>
<td>4.67</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>7.06</td>
<td>4.4</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>12.24</td>
<td>4.4</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>10.55</td>
<td>3.6</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>11.29</td>
<td>3.6</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>8.01</td>
<td>3.6</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>2.6</td>
<td>3.6</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>8.3</td>
<td>3.6</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>7.22</td>
<td>1.8</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>10.15</td>
<td>5</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>15.93</td>
<td>4</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>9.69</td>
<td>3.6</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>10.1</td>
<td>3</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>10.1</td>
<td>3</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>7.78</td>
<td>3</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>9.45</td>
<td>4.3</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>12.8</td>
<td>4.3</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>6.1</td>
<td>3</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>6.45</td>
<td>3</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet Cherry CRS photonov 2011</td>
<td>4.10.19-001</td>
<td>4.10.19-001_Roza_11</td>
<td>24728</td>
<td>Roza</td>
<td>2</td>
<td>2</td>
<td>6.45</td>
<td>3</td>
<td>16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Texture**

- select / deselect
- Firmness_1
### Goldrush

<table>
<thead>
<tr>
<th>Alias</th>
<th>n/a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother</td>
<td>GoldenDel</td>
</tr>
<tr>
<td>Father</td>
<td>Coop17</td>
</tr>
<tr>
<td>Cross Number</td>
<td>n/a</td>
</tr>
</tbody>
</table>

### Evaluation Data

**Phenotyping Data**: [download phenotypic characters]

**Genotyping Data**: [download]

*click image for larger view*

### Associated Markers / Alleles

<table>
<thead>
<tr>
<th>Marker</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACO1</td>
<td>2</td>
</tr>
<tr>
<td>CH05c06</td>
<td>116, 118</td>
</tr>
<tr>
<td>Md-Exp2</td>
<td>295</td>
</tr>
<tr>
<td>Hi04e04</td>
<td>222, 246</td>
</tr>
<tr>
<td>CH02b03</td>
<td>79, 97</td>
</tr>
</tbody>
</table>
Cross Assist: Generates a list of parents and the number of seedlings to get the progeny with desired traits

<table>
<thead>
<tr>
<th>Choices Made</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parental Dataset</td>
<td>SWEET_CHERRY_CRS</td>
</tr>
<tr>
<td>Target # of Seedlings</td>
<td>100</td>
</tr>
<tr>
<td># of Parents</td>
<td>60</td>
</tr>
<tr>
<td># of Pairs of Parents</td>
<td>1770</td>
</tr>
<tr>
<td>Method</td>
<td>Phenotype</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quantitative (continuous)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fruit_Wt</td>
<td>&gt;= 12</td>
</tr>
<tr>
<td>Firmness_1</td>
<td>&gt;= 300</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quantitative (discontinuous)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>not selected</td>
<td></td>
</tr>
</tbody>
</table>

- **Estimate # Seedlings**
- **Progress Bar**: 43%

- **Range**
  - >= 12
  - >= 300
### Filter Results

- **Flag**: 
- **Number of Seedlings**: 
- **Mother**: cowiche
- **Father**: 

- Both parents have data for all selected traits
- One parent has no data for at least one trait

### Cross List

<table>
<thead>
<tr>
<th># of Seedlings Required</th>
<th>Mother</th>
<th>Father</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>187</td>
<td>Cowiche</td>
<td>Selah</td>
<td>$f_x$</td>
</tr>
<tr>
<td>212</td>
<td>Cowiche</td>
<td>Lapins</td>
<td>$f_x$</td>
</tr>
<tr>
<td>250</td>
<td>Cowiche</td>
<td>Rainier</td>
<td>$f_x$</td>
</tr>
<tr>
<td>283</td>
<td>Cowiche</td>
<td>Sweetheart</td>
<td>$f_x$</td>
</tr>
<tr>
<td>342</td>
<td>Cowiche</td>
<td>PMR-1</td>
<td>$f_x$</td>
</tr>
<tr>
<td>351</td>
<td>Cowiche</td>
<td>Ambrunes</td>
<td>$f_x$</td>
</tr>
<tr>
<td>372</td>
<td>Cowiche</td>
<td>Chelan</td>
<td>$f_x$</td>
</tr>
<tr>
<td>379</td>
<td>Cowiche</td>
<td>Glacier</td>
<td>$f_x$</td>
</tr>
<tr>
<td>379</td>
<td>Cowiche</td>
<td>Van</td>
<td>$f_x$</td>
</tr>
<tr>
<td>381</td>
<td>Cowiche</td>
<td>9816-078</td>
<td>$f_x$</td>
</tr>
<tr>
<td>403</td>
<td>Cowiche</td>
<td>JJ</td>
<td>$f_x$</td>
</tr>
<tr>
<td>432</td>
<td>Cowiche</td>
<td>Kiona</td>
<td>$f_x$</td>
</tr>
<tr>
<td>462</td>
<td>Cowiche</td>
<td>9816-103</td>
<td>$f_x$</td>
</tr>
<tr>
<td>473</td>
<td>Cowiche</td>
<td>9816-083</td>
<td>$f_x$</td>
</tr>
</tbody>
</table>
Breeding Program Outline

The Washington State University apple breeding program began in 1994 to develop new varieties suitable to the unique climate of central Washington. Washington is the leading apple producing state with over 50% of U.S. production. Unfortunately, many of the new varieties developed in the world are not well adapted to growing conditions in central Washington or available to the majority of Washington growers.

The goal is to produce apples of a high eating quality with particular factors of outstanding flavor, texture and juiciness. The breeding program is a traditional breeding program, hybridizing parents with desirable traits. Promising seedlings are selected from large populations and their fruit is evaluated in the laboratory for eating quality and suitability for long-term storage. This program is one of the 12 core US breeding programs of the SCRI RosBREED project, enabling the application of marker-assisted breeding within the 4 years of the project.
<table>
<thead>
<tr>
<th>Role</th>
<th>View</th>
<th>Edit</th>
<th>Delete</th>
</tr>
</thead>
<tbody>
<tr>
<td>anonymous user</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>authenticated user</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breeders Toolbox Admin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breeders Toolbox TP Manager</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breeding Group Participant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FruitBreedomics Manager</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FruitBreedomics Member</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>peach guest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>publication curator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGC7 abstract moderator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGC7 editor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGC7 Planning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RosComm chief-editor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RosComm moderator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RosComm submitter</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Breeder without an up to date, comprehensive database

Button-clicking energized Breeder using an up to date database to help make breeding-decisions
GenSAS

• It is a web-based Genome Sequence Annotation Server
• A one-stop website with a single graphical interface for running multiple structural and functional annotation tools
• Enables the visualization and manual curation of genome sequences
• Funded by the USDA funded PineRefSeq project
Tasks are given custom names and added to the task queue.

- Multiple tasks can be added.
- Users are sent email notifications upon task execution and completion.
Specific Objectives

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

2. Develop a tablet application to collect phenotypic data from field and laboratory studies

3. Develop a Tripal Application Programming Interface for building breeding databases

4. Convert GenSAS, a community genome annotation tool, to Tripal

5. Develop Web Services to promote database interoperability
Tripal Databases Sustainability

• Database development consists of two components
  – Core development activities
  – Data analysis and curation activities

• Database costs can be split into 4 types
  – Core development (developers, db/sys administrators)
  – Data analysis and curation (data curators)
  – Operational costs (equipment, software, space, etc.)
  – Interaction costs (investigators, travel, etc.)
Tripal Databases Sustainability Model

- Core database developer salaries funded by NRSP for 5 years, benefits funded by WSU
- Data curators salaries and benefits funded by stakeholders (commodity commissions, grants, etc) - **Steering Committee Input**
- Curator positions can be located anywhere
- Other orphan crops can buy into this model or implement a Tripal database themselves (and we will provide support)
**Budget Request** ($1,991,190)

<table>
<thead>
<tr>
<th>Description</th>
<th>Yr1</th>
<th>Yr2</th>
<th>Yr3</th>
<th>Yr4</th>
<th>Yr5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salaries</td>
<td>303,631</td>
<td>315,165</td>
<td>326,834</td>
<td>338,969</td>
<td>351,591</td>
</tr>
<tr>
<td>Travel</td>
<td>20,000</td>
<td>20,000</td>
<td>20,000</td>
<td>20,000</td>
<td>20,000</td>
</tr>
<tr>
<td>Supplies</td>
<td>35,000</td>
<td>35,000</td>
<td>35,000</td>
<td>35,000</td>
<td>35,000</td>
</tr>
<tr>
<td>Hardware</td>
<td>40,000</td>
<td>40,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>398,631</td>
<td>370,165</td>
<td>381,834</td>
<td>433,969</td>
<td>406,591</td>
</tr>
</tbody>
</table>

- Within 3 years, **25%** of these core activities will be funded alternatively
- Within 5 years **50%** of these core activities will be funded alternatively
- Within 10 years, databases will be **self-sustaining** (but hopefully sooner 😊)
## Aligned Support ($2,166,942)

<table>
<thead>
<tr>
<th>Description</th>
<th>Yr1</th>
<th>Yr2</th>
<th>Yr3</th>
<th>Yr4</th>
<th>Yr5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salaries</td>
<td>184,523</td>
<td>280,893</td>
<td>69,003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fringe</td>
<td>170,632</td>
<td>219,097</td>
<td>132,026</td>
<td>105,068</td>
<td>109,272</td>
</tr>
<tr>
<td>Maintenance</td>
<td>197,327</td>
<td>197,288</td>
<td>148,216</td>
<td>134,759</td>
<td>128,960</td>
</tr>
<tr>
<td>Travel</td>
<td>5,000</td>
<td>17,000</td>
<td>5,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplies</td>
<td>19,327</td>
<td>18,000</td>
<td>5,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hardware</td>
<td>20,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>597,354</strong></td>
<td><strong>732,278</strong></td>
<td><strong>359,245</strong></td>
<td><strong>239,827</strong></td>
<td><strong>238,238</strong></td>
</tr>
</tbody>
</table>
Acknowledgements

- Mainlab Bioinformatics Team 😊
- Project coPIs/Pis
  - tfGDR (GDR and Citrus); Cacao Genome Database; Pine Genome Sequencing Project; Genome Database for Vaccinium; Cool Season Food Legume Database; CottonGen
- Rosaceae, Citrus, Cacao, Blueberry, Pea, Chickpea, Lentil, Cotton and Bioinformatics Community 😊
- USDA NIFA SCRI, USDA DOE, NSF Plant Genome Program, USDA-ARS, Mars Inc, Washington Tree Fruit Research Commission, USA Dry Pea and Lentil Commission
- US Land Grant University researchers and extension agents
Thanks for listening 😊