NRSP_temp321

Database Resources for Crop Genomics, Genetics and Breeding Research

2014 SAAESD Spring Meeting
Savannah, GA

UNIVERSITY OF ARKANSAS

A&M

AUBURN UNIVERSITY

CLEMSON UNIVERSITY

UF UNIVERSITY of FLORIDA

The University of Georgia

UK UNIVERSITY OF KENTUCKY

LSU

MISSISSIPPI STATE UNIVERSITY

NC STATE UNIVERSITY

OKLAHOMA STATE UNIVERSITY

UNIVERSIDAD DE PUERTO RICO · RECINTO DE RIO PIEDRAS

THE UNIVERSITY of TENNESSEE KNOXVILLE

UNIVERSITY OF THE VIRGIN ISLANDS

Virginia Tech

NRSP_temp321

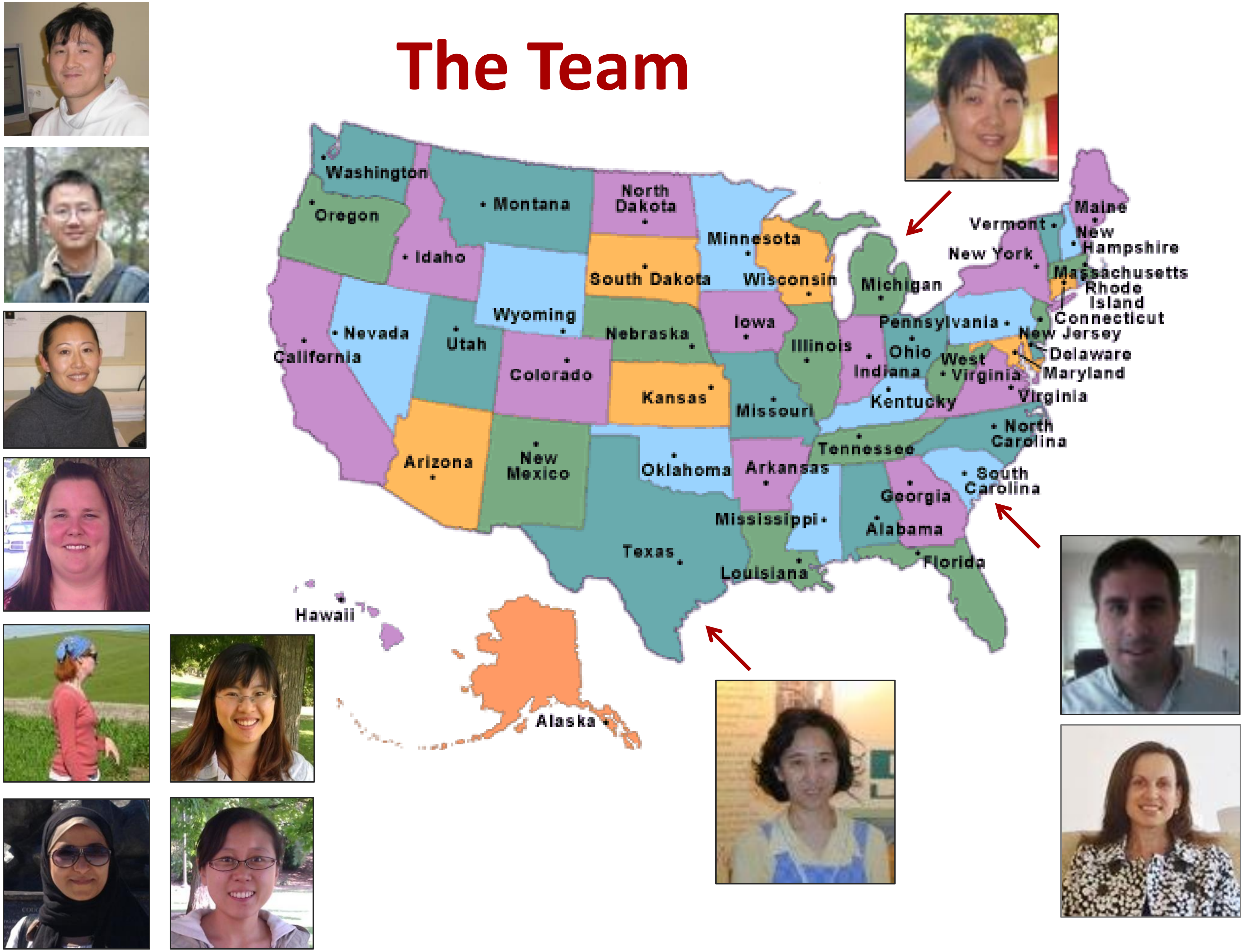
Database Resources for Crop Genomics, Genetics and Breeding Research

2014 SAAESD Spring Meeting
Savannah, GA

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

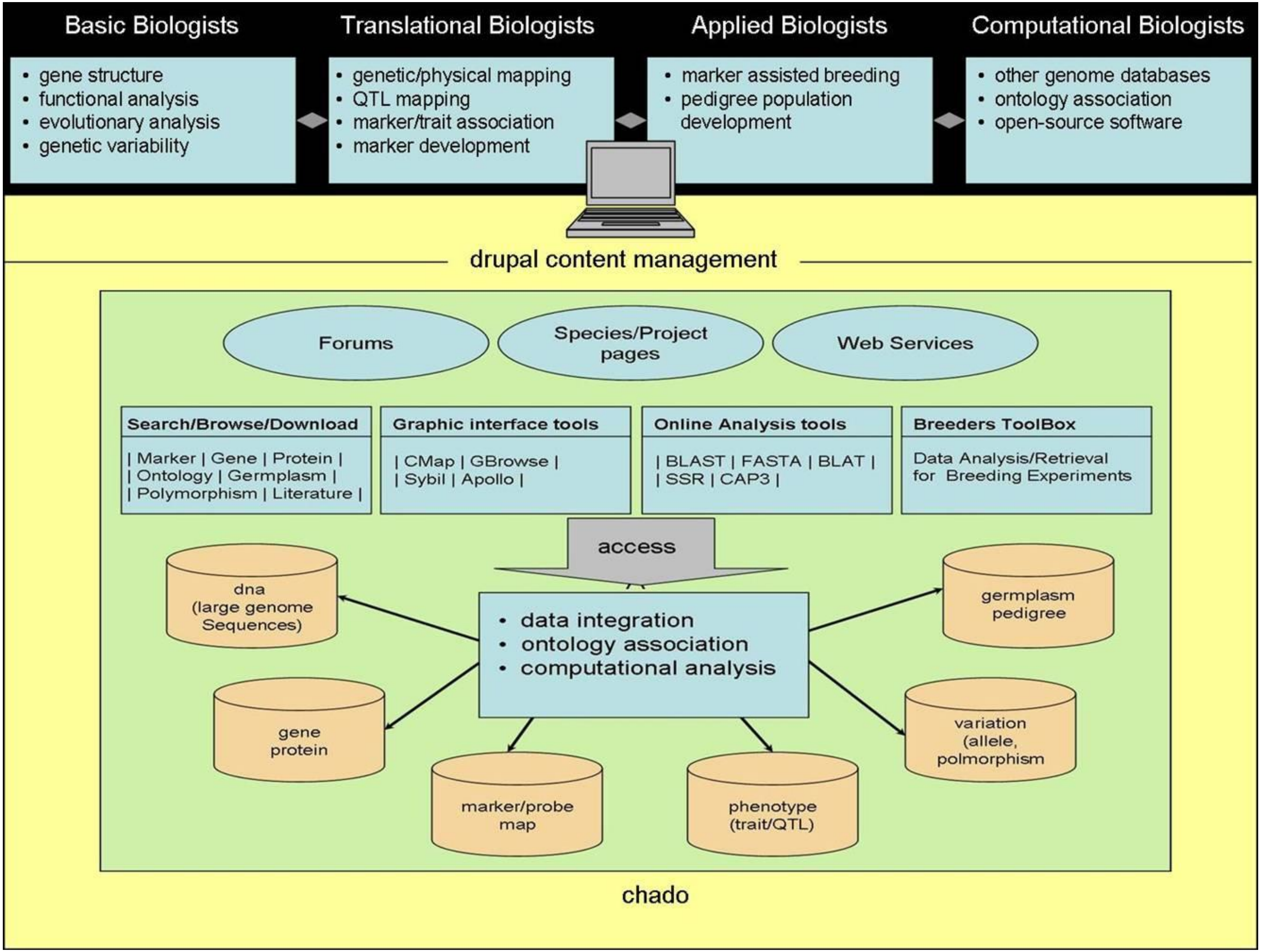
Writing Team
Dorrie Main(WSU)
Sook Jung (WSU)
Mike Kahn (WSU)
Cameron Peace (WSU)
Jim McFerson (WTRC)
Reviewers (US Wide)

The Team


Washington
Oregon
Montana
Idaho
North Dakota
South Dakota
Minnesota
Wisconsin
Michigan
Vermont
Maine
New Hampshire
New York
Massachusetts
Rhode Island
Connecticut
New Jersey
Delaware
Maryland
Pennsylvania
Ohio
West Virginia
Virginia
Indiana
Illinois
Iowa
Nebraska
Wyoming
Nevada
California
Utah
Colorado
Kansas
Missouri
Kentucky
Tennessee
North Carolina
South Carolina
Georgia
Alabama
Mississippi
Arkansas
Oklahoma
Arizona
New Mexico
Texas
Louisiana
Florida
Hawaii
Alaska

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


Basic Biologists
• gene structure
• functional analysis
• evolutionary analysis
• genetic variability
Translational Biologists
• genetic/physical mapping
• QTL mapping
• marker/trait association
• marker development
Applied Biologists
• marker assisted breeding
• pedigree population development
Computational Biologists
• other genome databases
• ontology association
• open-source software
drupal content management
Forums
Species/Project pages
Web Services
Search/Browse/Download
Marker	Gene	Protein
Ontology	Germplasm
Polymorphism	Literature
Graphic interface tools
CMap	GBrowse
Sybil	Apollo
Online Analysis tools
BLAST	FASTA	BLAT
SSR	CAP3
Breeders ToolBox
Data Analysis/Retrieval for Breeding Experiments
access
• data integration
• ontology association
• computational analysis
dna (large genome Sequences)
gene protein
marker/probe map
phenotype (trait/QTL)
variation (allele, polmorphism
germplasm pedigree
chado

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-To's: 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

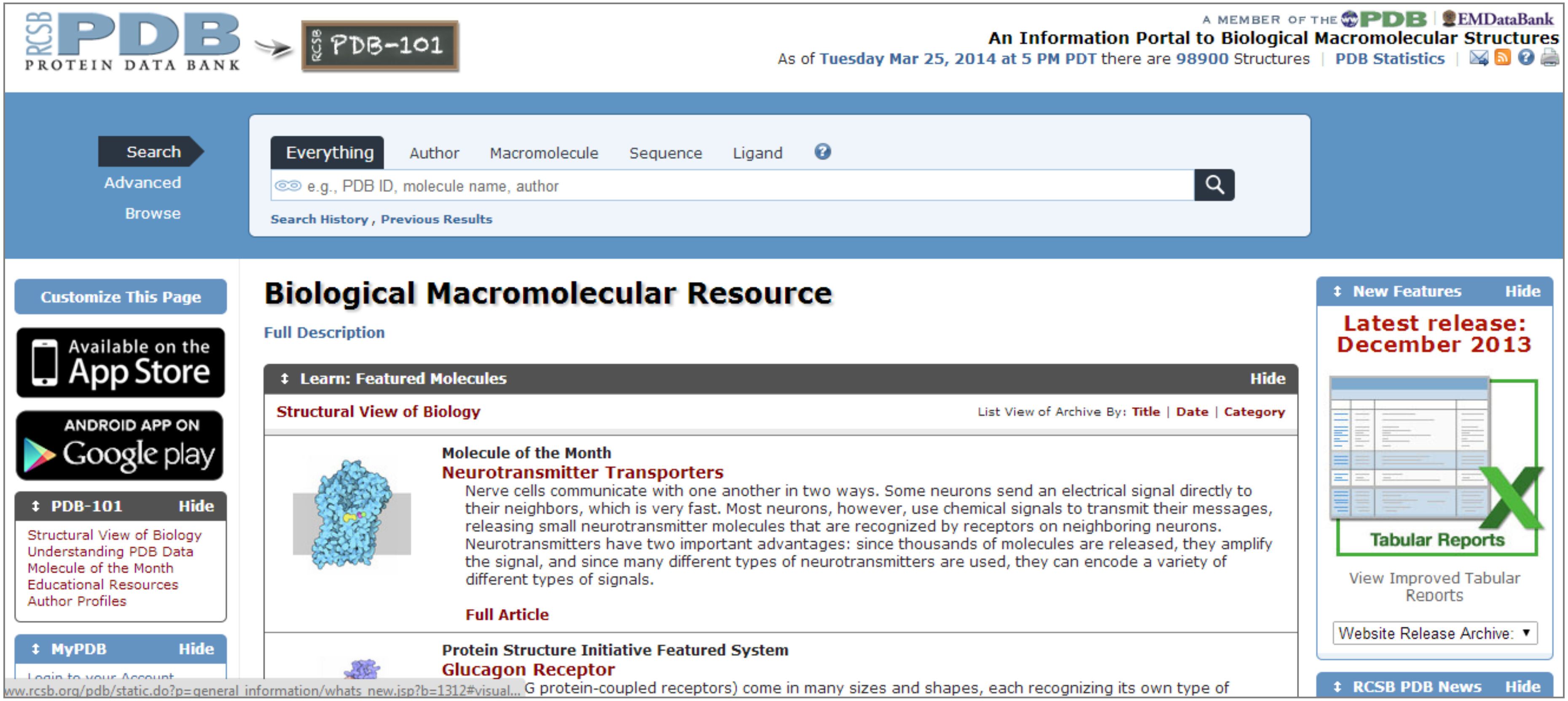
NCBI Announcements

New NCBI YouTube video: Create custom databases for BLAST

Mar 28, 2014

In the newest NCBI video on YouTube,

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



An Information Portal to Biological Macromolecular Structures

As of Tuesday Mar 25, 2014 at 5 PM PDT there are 98900 Structures | PDB Statistics

e.g., PDB ID, molecule name, author

Structural View of Biology
Understanding PDB Data
Molecule of the Month
Educational Resources
Author Profiles

Biological Macromolecular Resource

Full Description

Structural View of Biology

List View of Archive By: Title | Date | Category

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

Glucagon Receptor

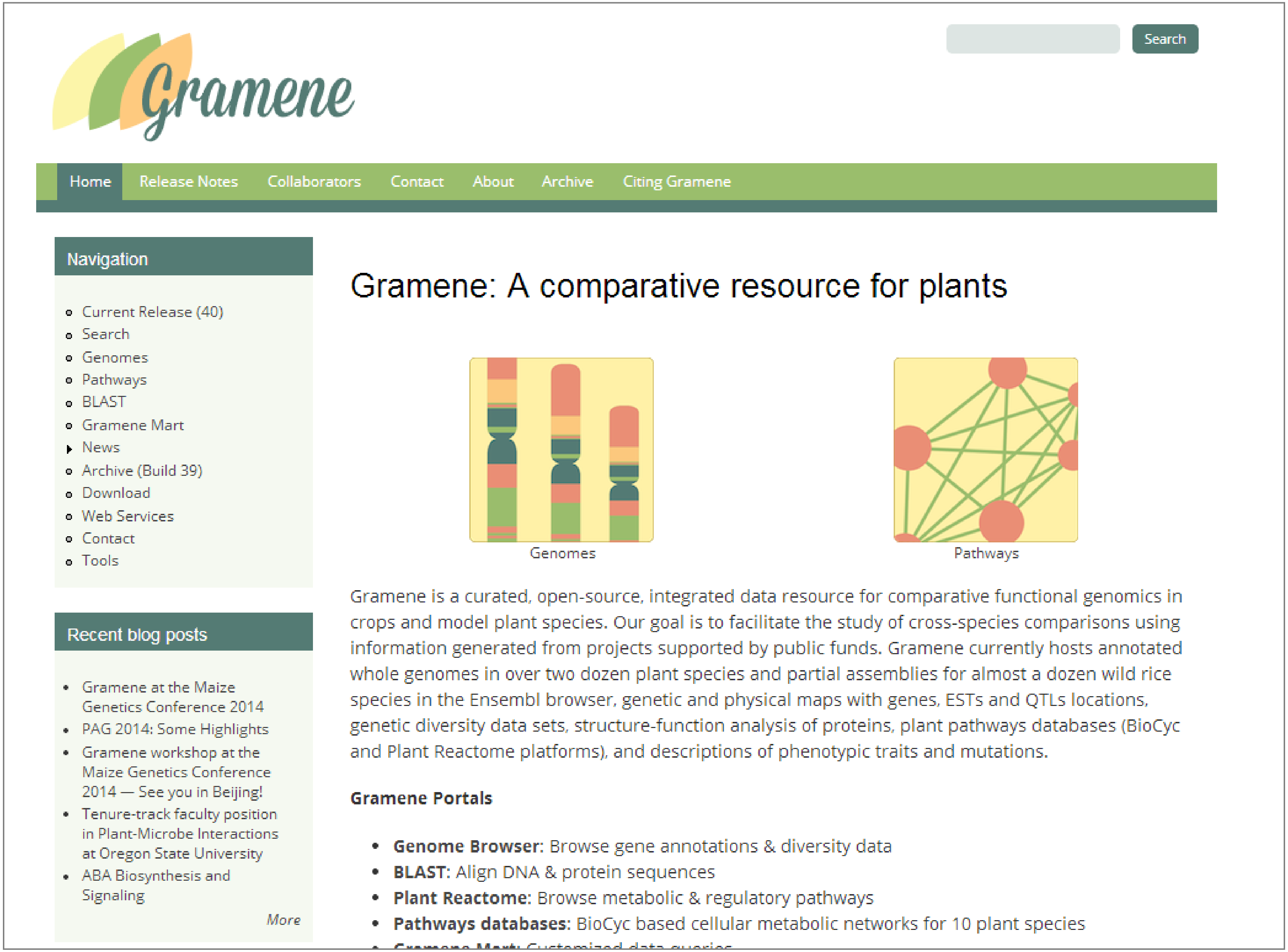
G protein-coupled receptors) come in many sizes and shapes, each recognizing its own type of

Latest release: December 2013

View Improved Tabular Reports

Genome Databases

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

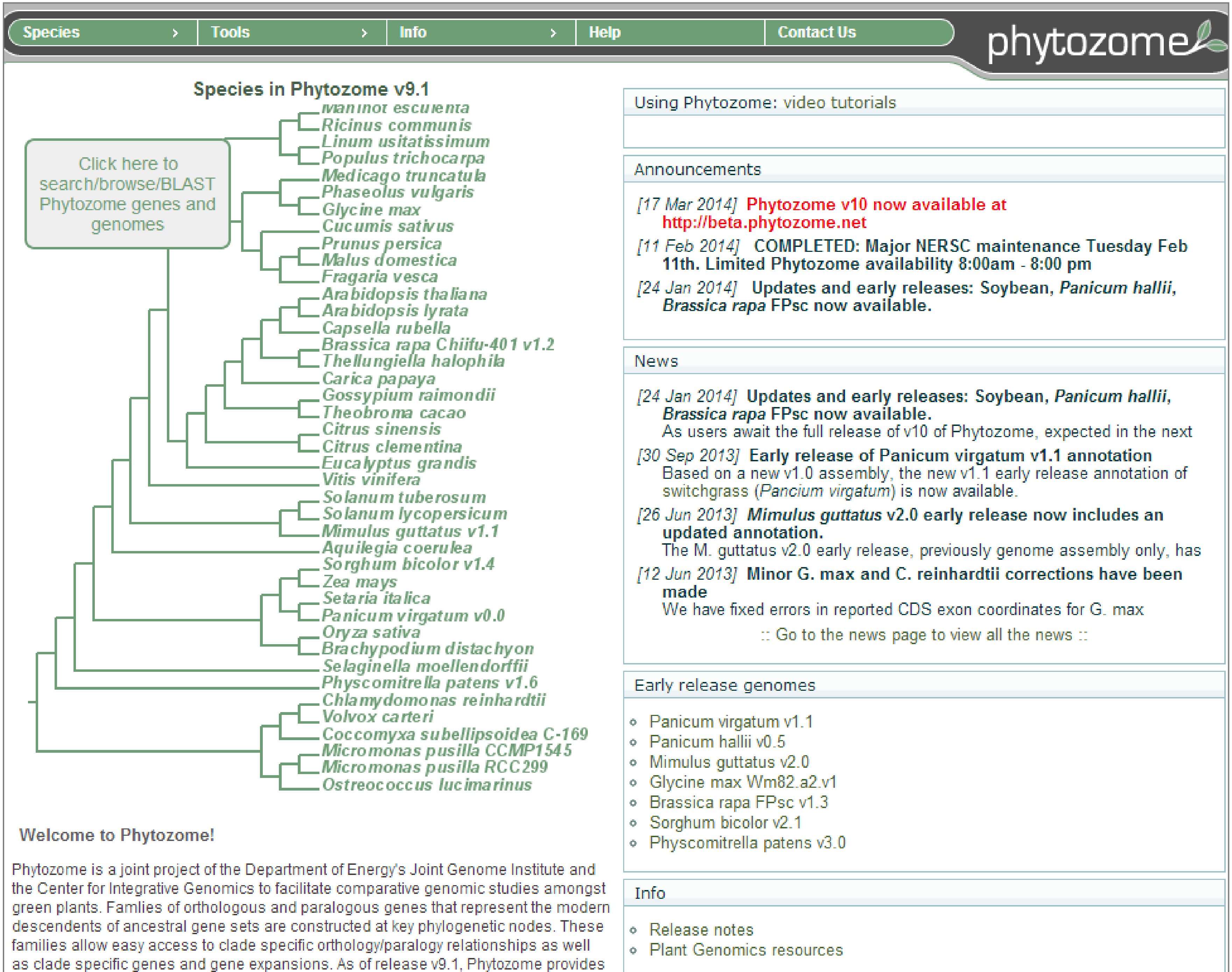


Gramene: A comparative resource for plants

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
- **Gramene Mart:** Customized data queries



Species in Phytozome v9.1

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. Famlies 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

Using Phytozome: video tutorials

Announcements

[17 Mar 2014] **Phytozome v10 now available at http://beta.phytozome.net**

[11 Feb 2014] **COMPLETED: Major NERSC maintenance Tuesday Feb 11th. Limited Phytozome availability 8:00am - 8:00 pm**

[24 Jan 2014] **Updates and early releases: Soybean, *Panicum hallii*, *Brassica rapa* FPsc now available.**

News

[24 Jan 2014] **Updates and early releases: Soybean, *Panicum hallii*, *Brassica rapa* FPsc now available.**
As users await the full release of v10 of Phytozome, expected in the next

[30 Sep 2013] **Early release of Panicum virgatum v1.1 annotation**
Based on a new v1.0 assembly, the new v1.1 early release annotation of switchgrass (*Pancium virgatum*) is now available.

[26 Jun 2013] ***Mimulus guttatus* v2.0 early release now includes an updated annotation.**
The M. guttatus v2.0 early release, previously genome assembly only, has

[12 Jun 2013] **Minor G. max and C. reinhardtii corrections have been made**
We have fixed errors in reported CDS exon coordinates for G. max

:: Go to the news page to view all the news ::

Early release genomes

- Panicum virgatum v1.1
- Panicum hallii v0.5
- Mimulus guttatus v2.0
- Glycine max Wm82.a2.v1
- Brassica rapa FPsc v1.3
- Sorghum bicolor v2.1
- Physcomitrella patens v3.0

Info

- Release notes
- Plant Genomics resources

Genome Databases

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

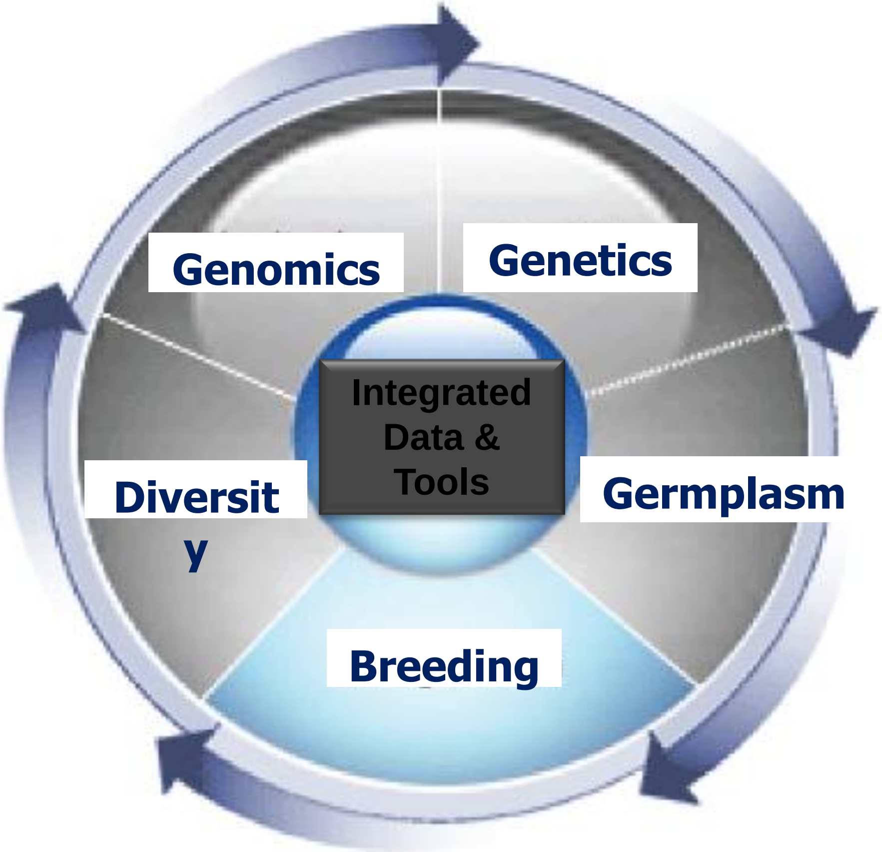
Why Do We Need Community Databases?

- 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

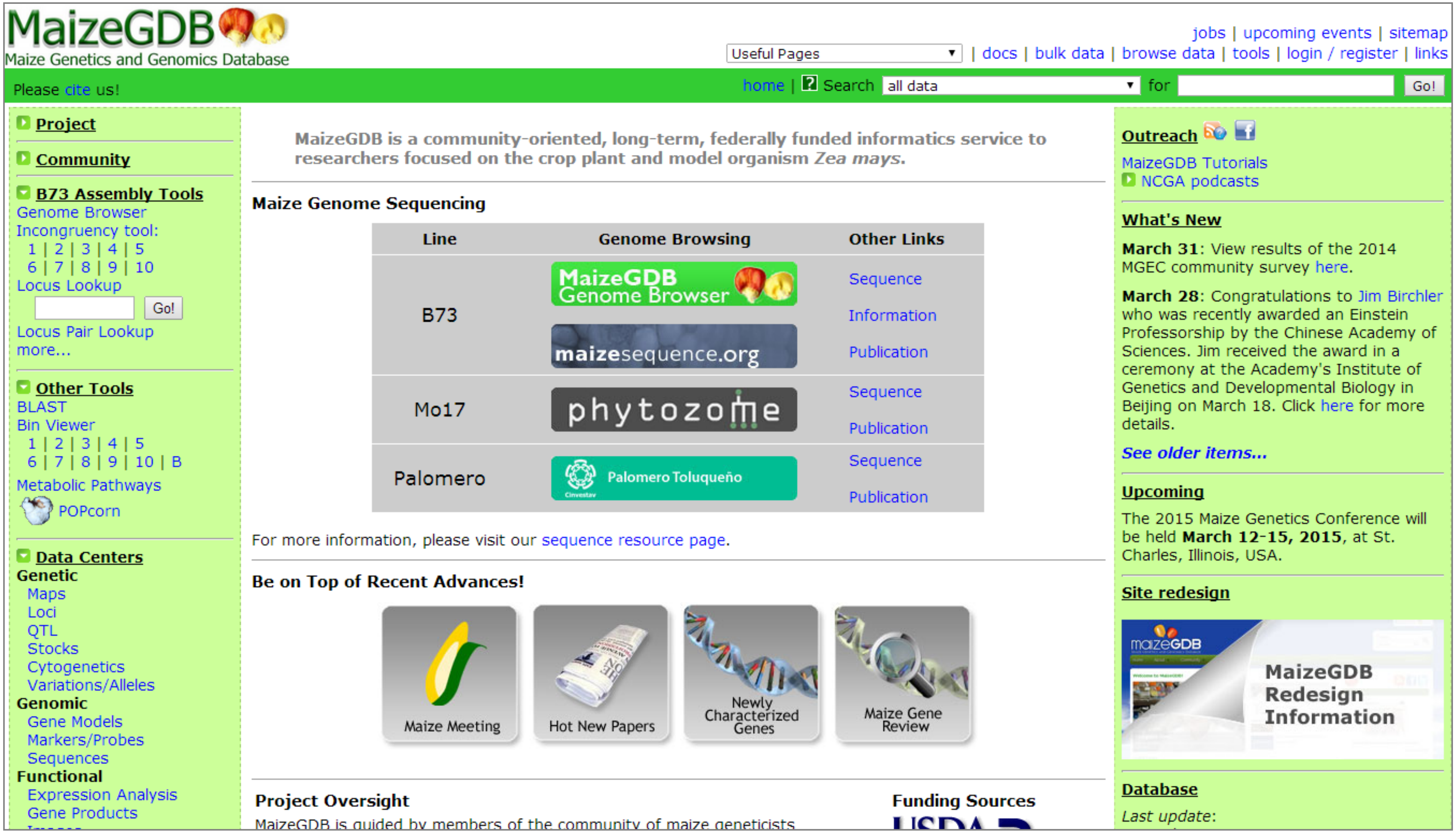


Translational Science

QTL /marker discovery, genetic mapping, Breeding values

Applied Science

Utilization of DNA information in breeding decisions



MaizeGDB

Maize Genetics and Genomics Database

jobs | upcoming events | sitemap

Useful Pages ▾ | docs | bulk data | browse data | tools | login / register | links

Please cite us!

home | Search all data ▾ for Go!

Project

Community

B73 Assembly Tools

Genome Browser

Incongruency tool:

1 | 2 | 3 | 4 | 5

6 | 7 | 8 | 9 | 10

Locus Lookup

Go!

Locus Pair Lookup

more...

Other Tools

BLAST

Bin Viewer

1 | 2 | 3 | 4 | 5

6 | 7 | 8 | 9 | 10 | B

Metabolic Pathways

POPcorn

Data Centers

Genetic

- Maps
- Loci
- QTL
- Stocks
- Cytogenetics
- Variations/Alleles

Genomic

- Gene Models
- Markers/Probes
- Sequences

Functional

- Expression Analysis
- Gene Products

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

Line	Genome Browsing	Other Links
B73	MaizeGDB Genome Browser maizesequence.org	Sequence Information Publication
Mo17	phytozome	Sequence Publication
Palomero	Palomero Toluqueño	Sequence Publication

For more information, please visit our sequence resource page.

Be on Top of Recent Advances!

Project Oversight

MaizeGDB is guided by members of the community of maize geneticists

Funding Sources

Outreach

MaizeGDB Tutorials

NCGA podcasts

What's New

March 31: View results of the 2014 MGEC community survey here.

March 28: Congratulations to Jim Birchler who was recently awarded an Einstein Professorship by the Chinese Academy of Sciences. Jim received the award in a ceremony at the Academy's Institute of Genetics and Developmental Biology in Beijing on March 18. Click here for more details.

See older items...

Upcoming

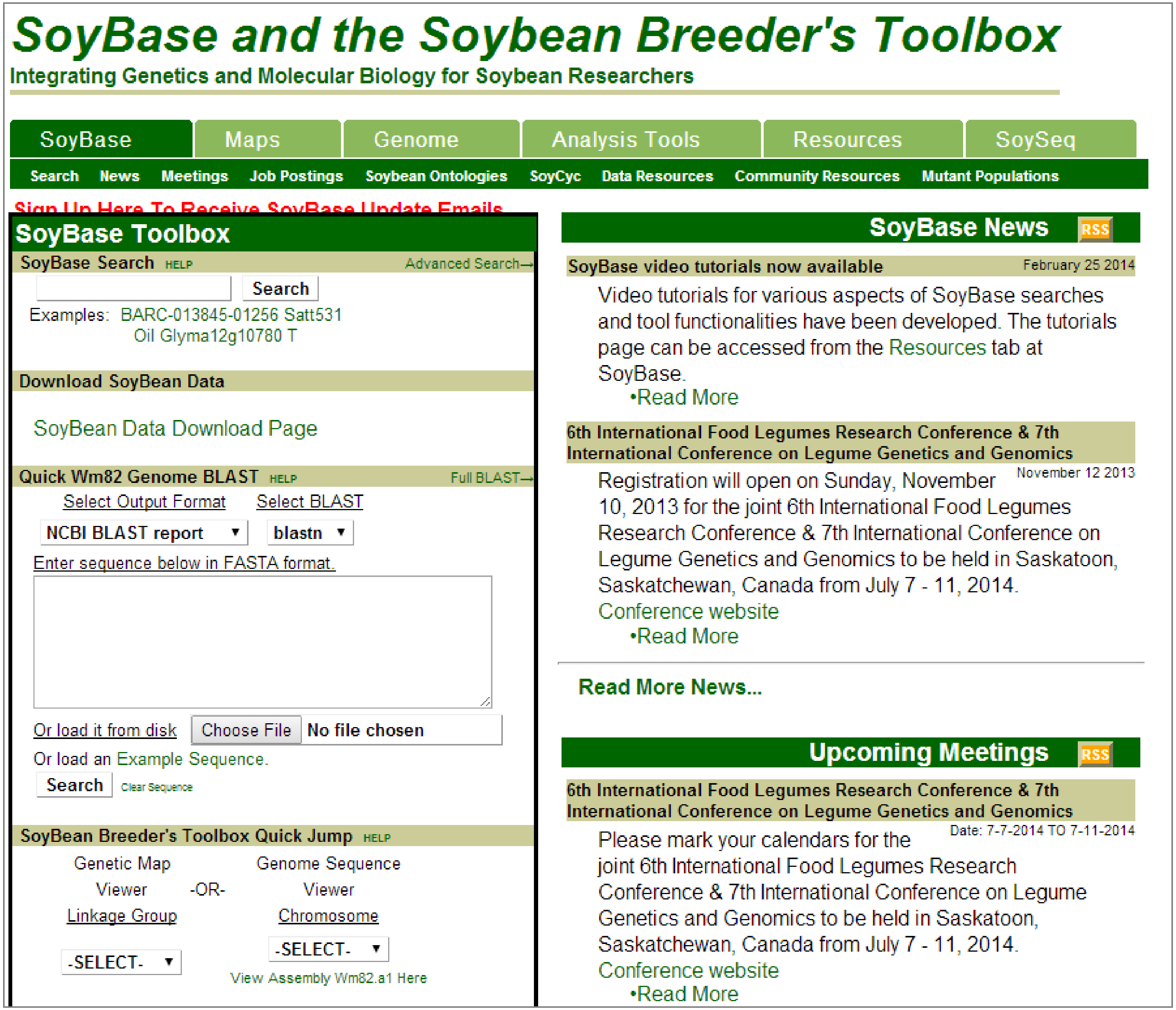
The 2015 Maize Genetics Conference will be held **March 12-15, 2015**, at St. Charles, Illinois, USA.

Site redesign

MaizeGDB Redesign Information

Database

Last update:



SoyBase and the Soybean Breeder's Toolbox

Integrating Genetics and Molecular Biology for Soybean Researchers

Sign Up Here To Receive SoyBase Update Emails

SoyBase Toolbox

SoyBase Search

Examples: BARC-013845-01256 Satt531 Oil Glyma12g10780 T

Download SoyBean Data

SoyBean Data Download Page

Quick Wm82 Genome BLAST

Enter sequence below in FASTA format.

Or load it from disk

Or load an Example Sequence.

SoyBean Breeder's Toolbox Quick Jump

Genetic Map Viewer -OR- Genome Sequence Viewer

Linkage Group

Chromosome

View Assembly Wm82.a1 Here

SoyBase News

SoyBase video tutorials now available February 25 2014

Video tutorials for various aspects of SoyBase searches and tool functionalities have been developed. The tutorials page can be accessed from the Resources tab at SoyBase.

•Read More

6th International Food Legumes Research Conference & 7th International Conference on Legume Genetics and Genomics November 12 2013

Registration will open on Sunday, November 10, 2013 for the joint 6th International Food Legumes Research Conference & 7th International Conference on Legume Genetics and Genomics to be held in Saskatoon, Saskatchewan, Canada from July 7 - 11, 2014.

Conference website

•Read More

Read More News...

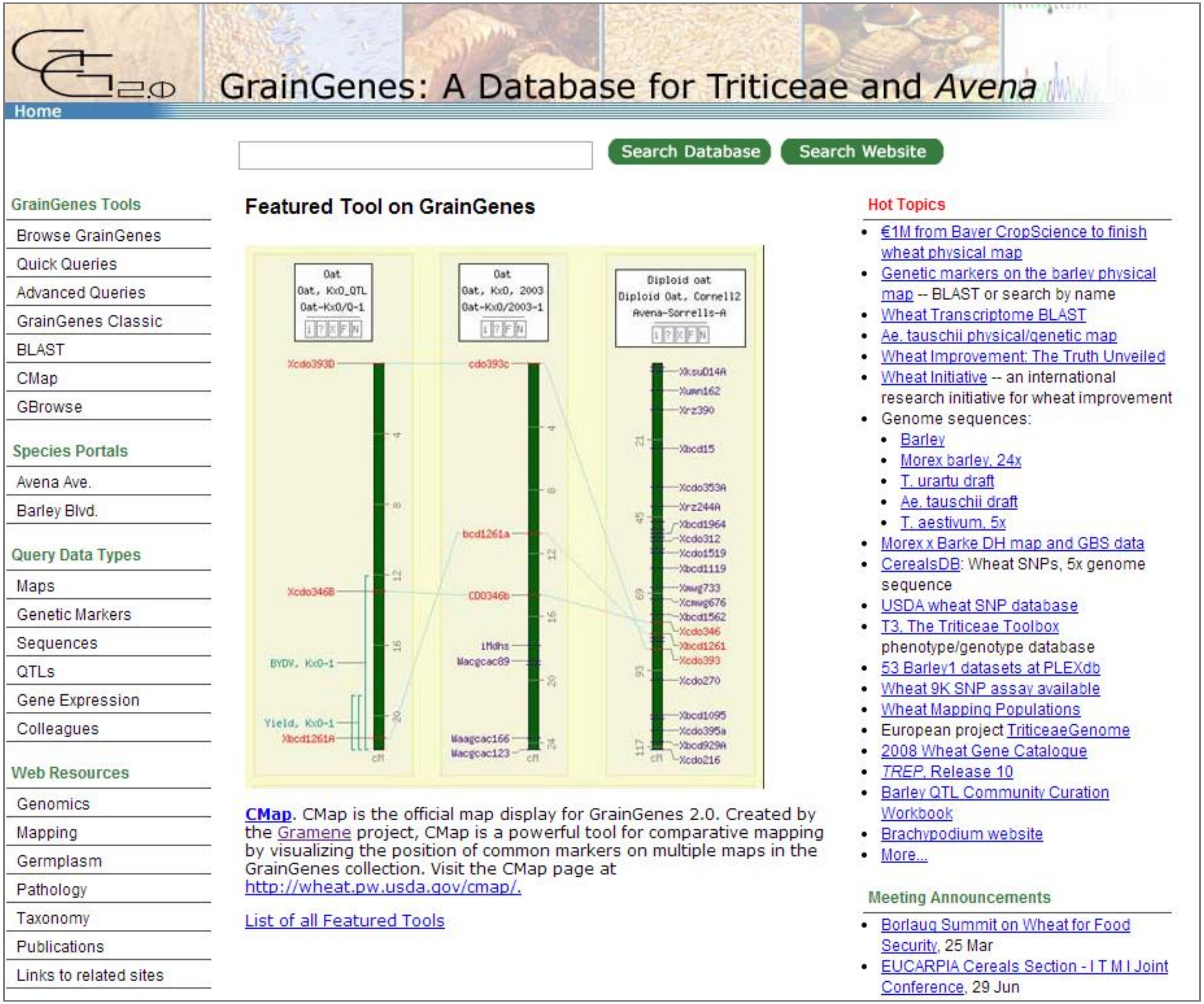
Upcoming Meetings

6th International Food Legumes Research Conference & 7th International Conference on Legume Genetics and Genomics Date: 7-7-2014 TO 7-11-2014

Please mark your calendars for the joint 6th International Food Legumes Research Conference & 7th International Conference on Legume Genetics and Genomics to be held in Saskatoon, Saskatchewan, Canada from July 7 - 11, 2014.

Conference website

•Read More



GrainGenes: A Database for Triticeae and *Avena*

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.pw.usda.gov/cmap/.

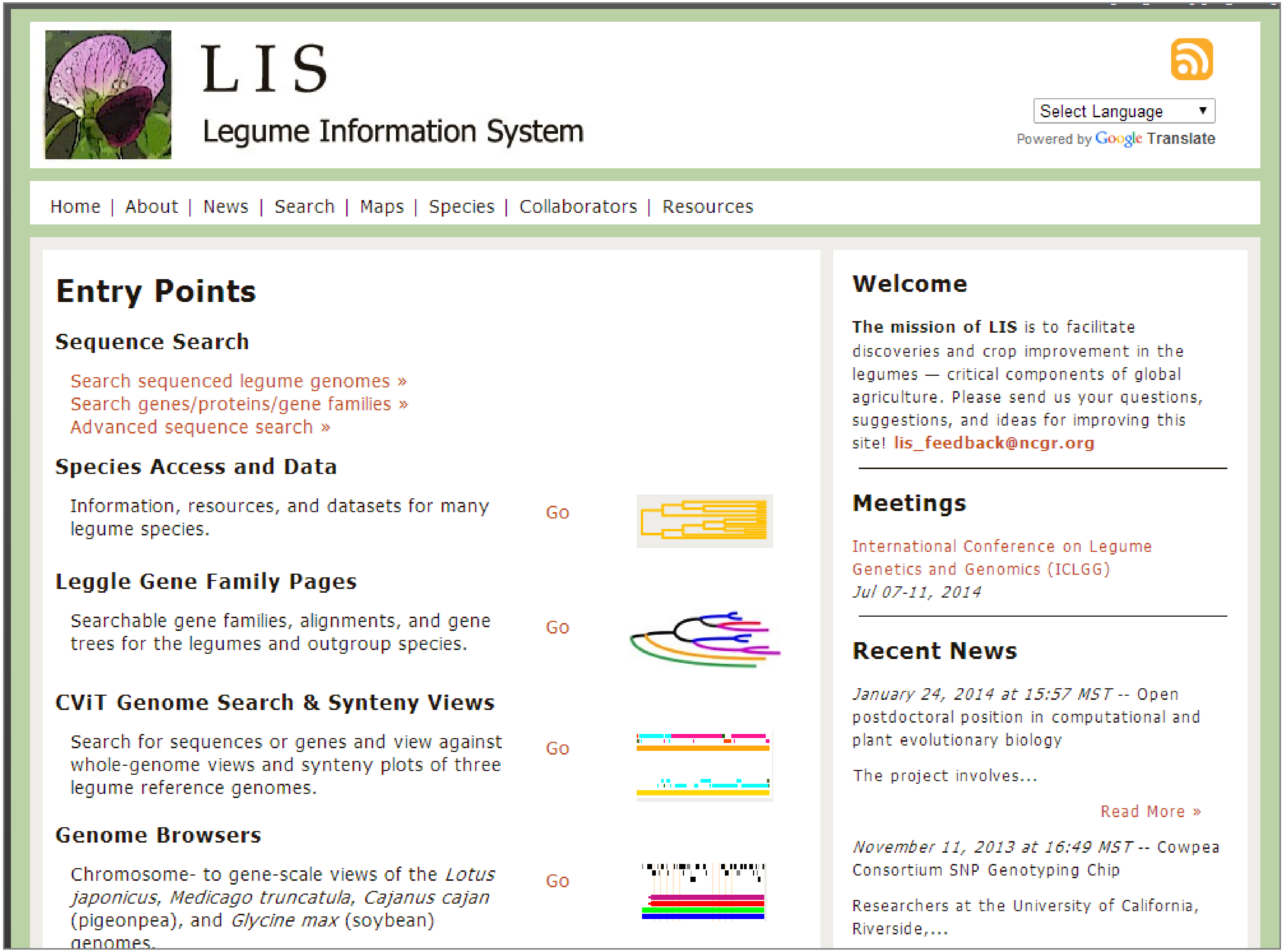
List of all Featured Tools

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/genetic 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 x Barke DH map and GBS data
- CerealsDB: Wheat SNPs, 5x genome sequence
- USDA wheat SNP database
- T3, The Triticeae Toolbox phenotype/genotype database
- 53 Barley1 datasets at PLEXdb
- Wheat 9K SNP assay available
- Wheat Mapping Populations
- European project TriticeaeGenome
- 2008 Wheat Gene Catalogue
- *TREP*, Release 10
- Barley QTL Community Curation Workbook
- Brachypodium website
- More...

Meeting Announcements

- Borlaug Summit on Wheat for Food Security, 25 Mar
- EUCARPIA Cereals Section - I T M I 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 (ICLGG)
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 / nonprofit 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

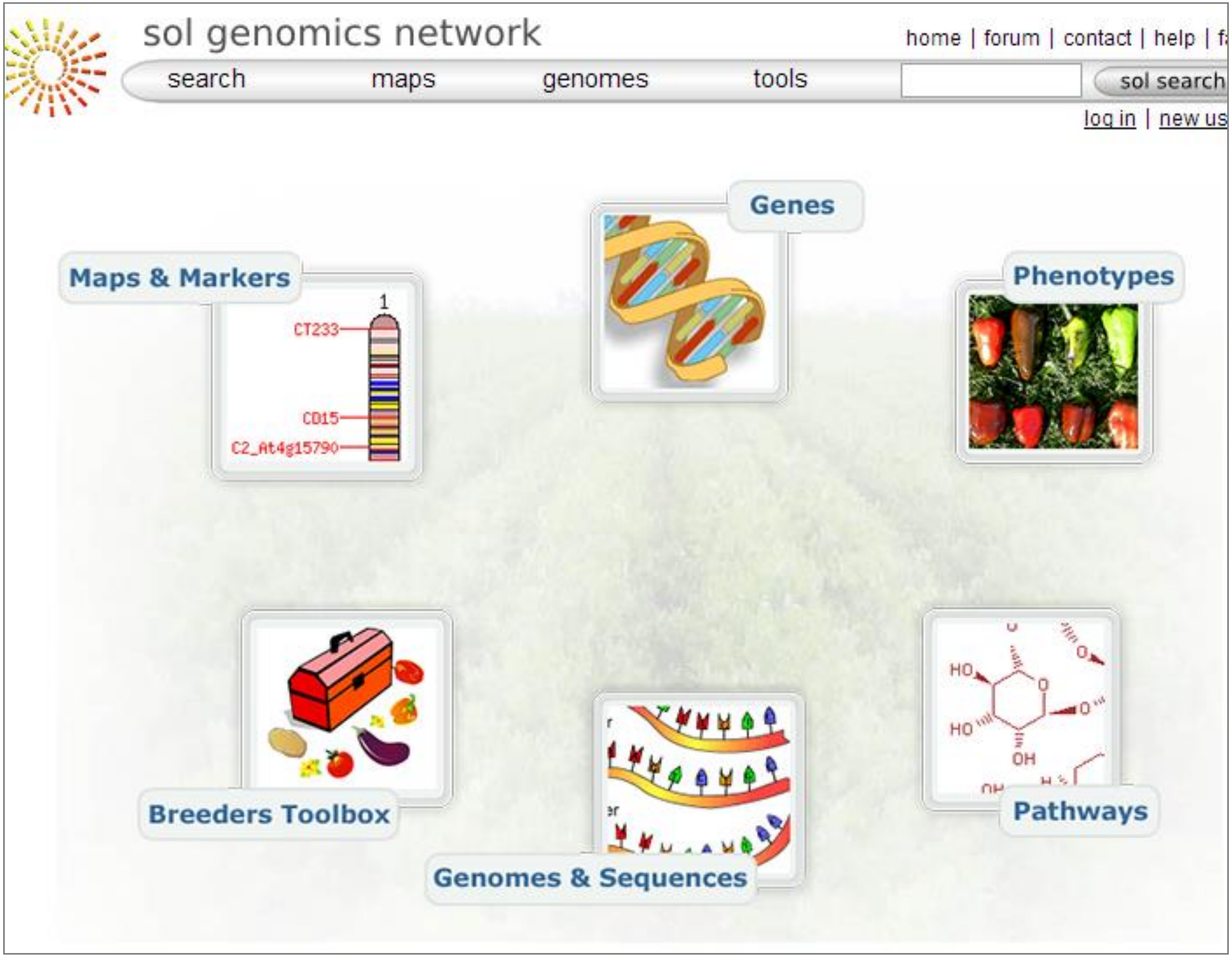
Subscribe to news feed
Follow our Twitter feed
Join our Facebook group

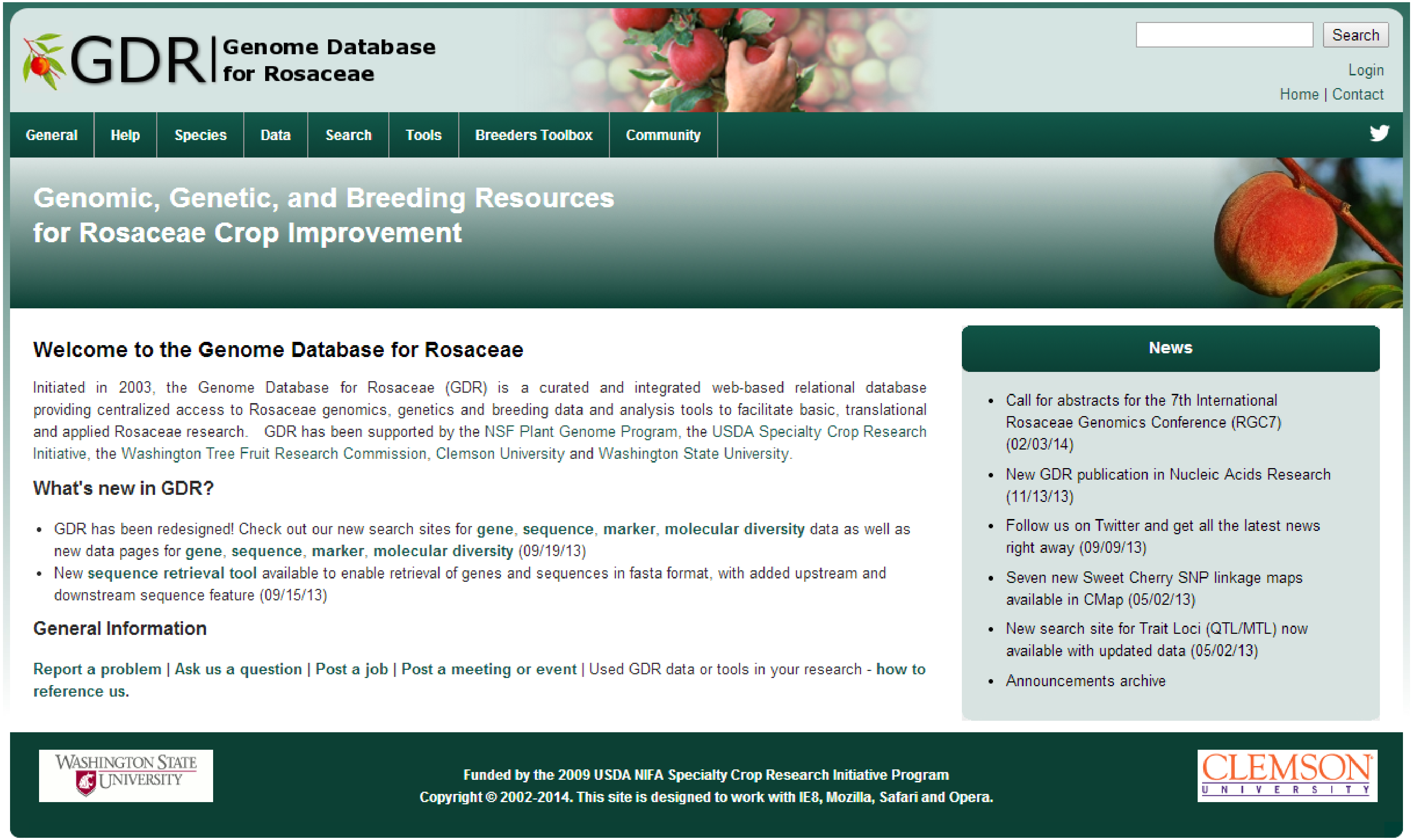
Breaking News

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 that they have ordered. We are collecting information such as validation of insert/sequence,


sol genomics network
home | forum | contact | help | f
search
maps
genomes
tools
sol search
log in | new us
Genes
Maps & Markers
1
CT233
CD15
C2_At4g15790
Phenotypes
HO
HO
OH
O
Breeders Toolbox
Pathways
Genomes & Sequences



Genomic, Genetic, and Breeding Resources for Rosaceae Crop Improvement

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.**

News

- Call for abstracts for the 7th International Rosaceae Genomics Conference (RGC7) (02/03/14)
- New GDR publication in Nucleic Acids Research (11/13/13)
- Follow us on Twitter and get all the latest news right away (09/09/13)
- Seven new Sweet Cherry SNP linkage maps available in CMap (05/02/13)
- New search site for Trait Loci (QTL/MTL) now available with updated data (05/02/13)
- Announcements archive

Funded by the 2009 USDA NIFA Specialty Crop Research Initiative Program

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

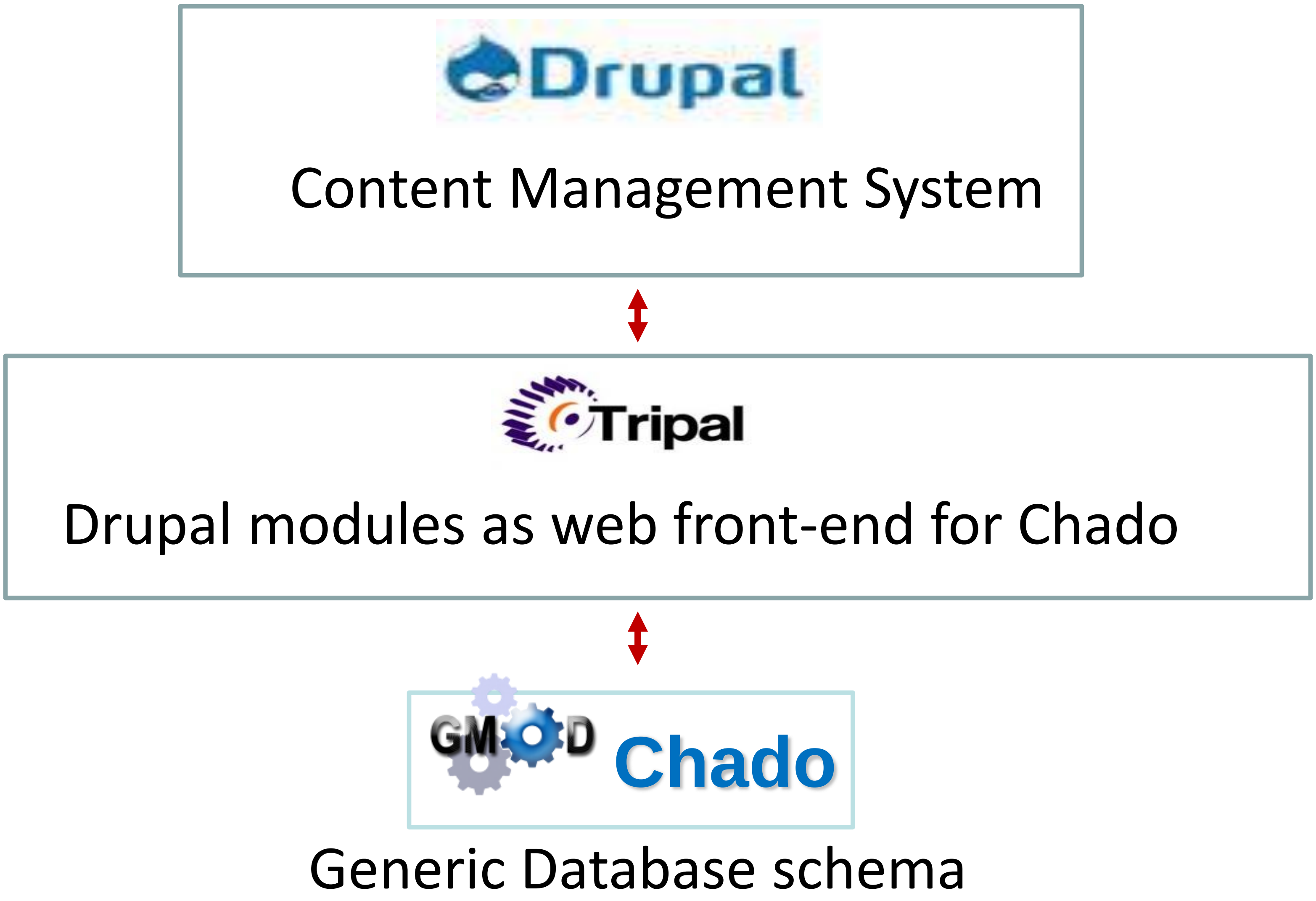
  = difficult to add new data types
  = difficult to implement for other species.
  = generally resource inefficient

- 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

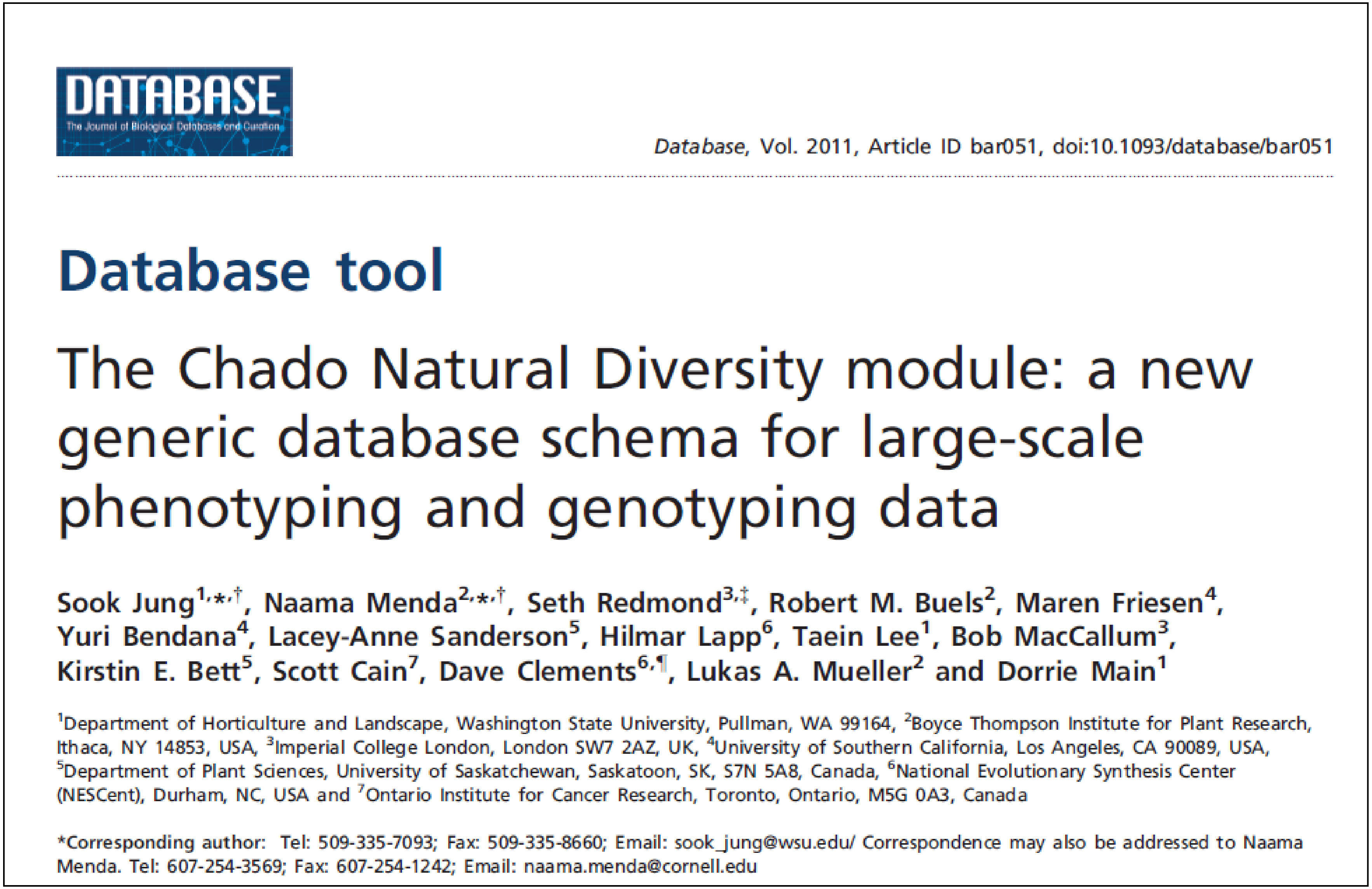
Content Management System

Tripal

Drupal modules as web front-end for Chado

GMOD Chado

Generic Database schema



DATABASE The Journal of Biological Databases and Curation

Database, Vol. 2011, Article ID bar051, doi:10.1093/database/bar051

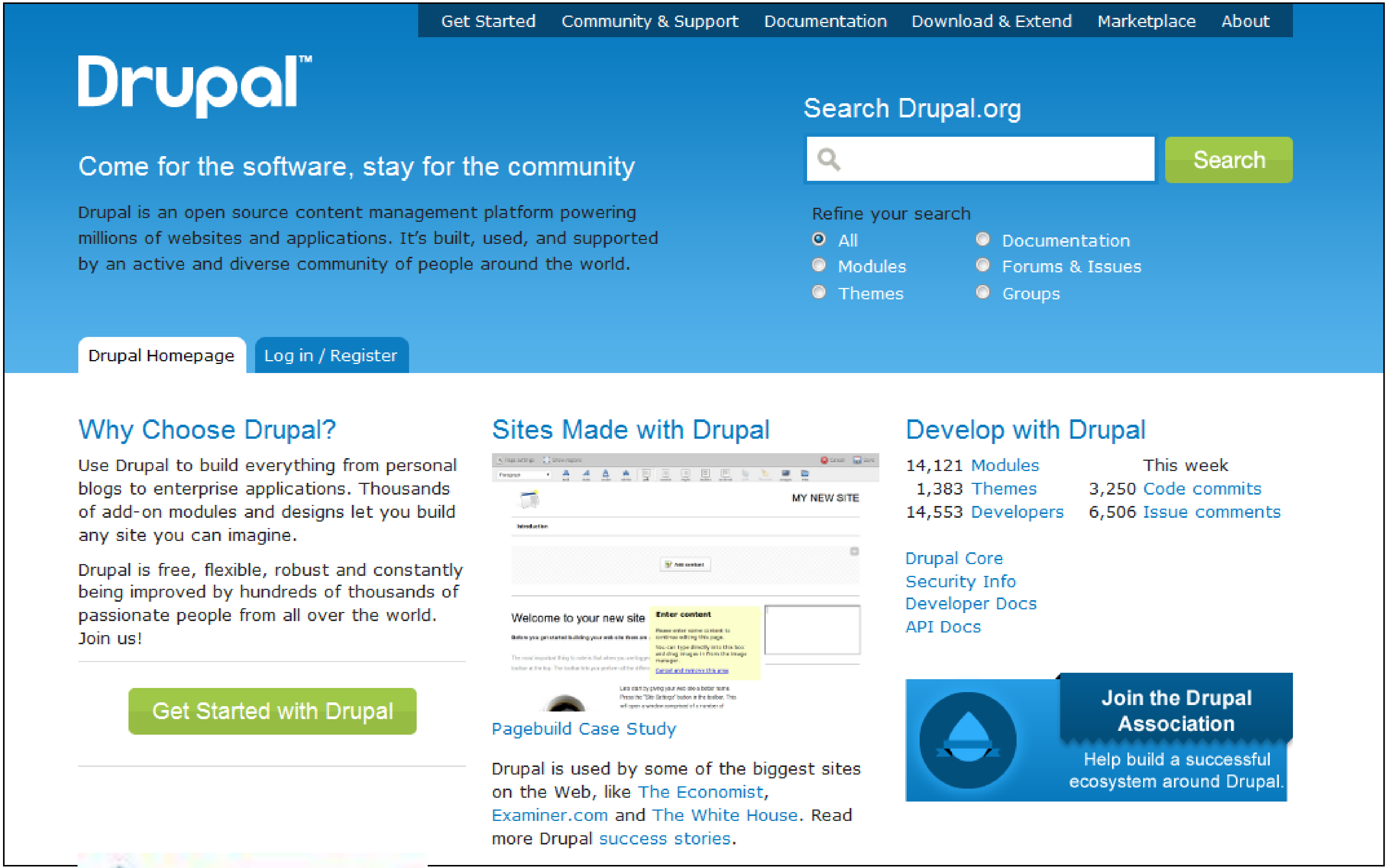
Database tool

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

Sook Jung[1,*,†], Naama Menda[2,*,†], Seth Redmond[3,‡], Robert M. Buels[2], Maren Friesen[4], Yuri Bendana[4], Lacey-Anne Sanderson[5], Hilmar Lapp[6], Taein Lee[1], Bob MacCallum[3], Kirstin E. Bett[5], Scott Cain[7], Dave Clements[6,¶], Lukas A. Mueller[2] and Dorrie Main[1]

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



Drupal™

Come for the software, stay for the community

Drupal is an open source content management platform powering millions of websites and applications. It's built, used, and supported by an active and diverse community of people around the world.

Search Drupal.org

Refine your search

● All	○ Documentation
○ Modules	○ Forums & Issues
○ Themes	○ Groups

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

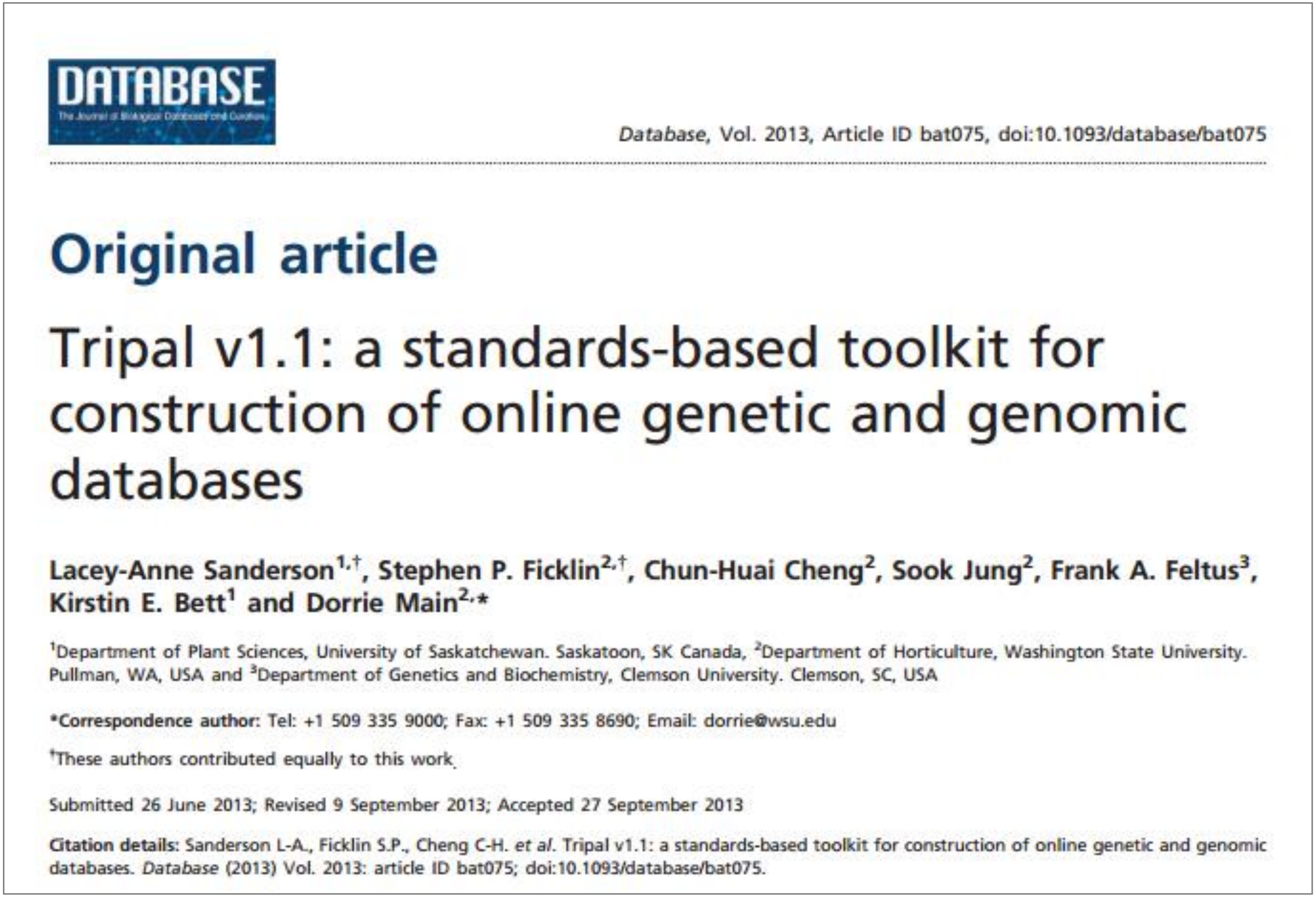
14,121	Modules		This week
1,383	Themes	3,250	Code commits
14,553	Developers	6,506	Issue comments

Drupal Core
Security Info
Developer Docs
API Docs

Join the Drupal Association

Help build a successful ecosystem around Drupal.

Drupal



DATABASE

Database, Vol. 2013, Article ID bat075, doi:10.1093/database/bat075

Original article

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

Lacey-Anne Sanderson[1,†], Stephen P. Ficklin[2,†], Chun-Huai Cheng[2], Sook Jung[2], Frank A. Feltus[3], Kirstin E. Bett[1] and Dorrie Main[2,*]

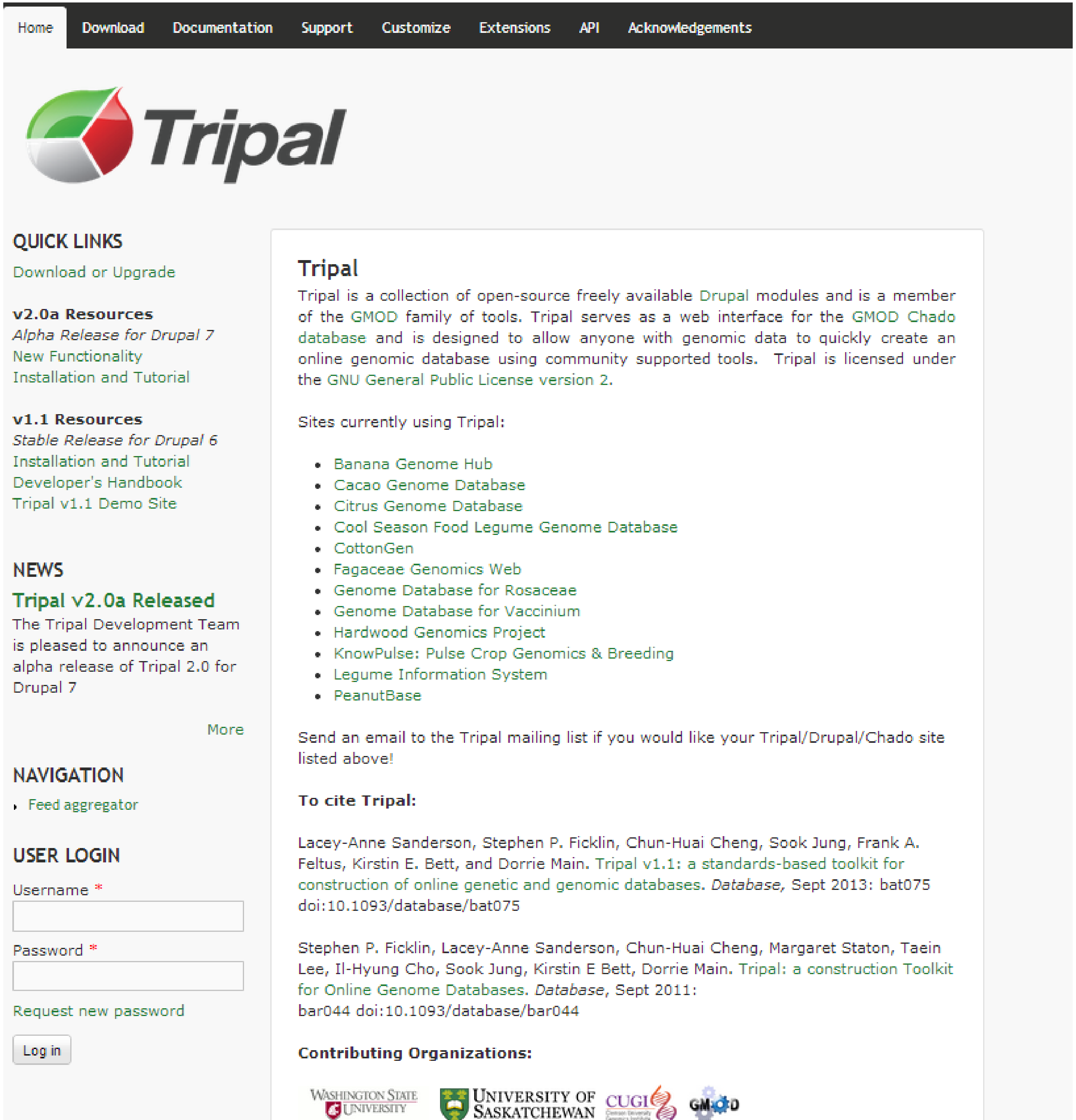
[1]Department of Plant Sciences, University of Saskatchewan. Saskatoon, SK Canada, [2]Department of Horticulture, Washington State University. Pullman, WA, USA and [3]Department of Genetics and Biochemistry, Clemson University. Clemson, SC, USA

***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

Citation details: Sanderson L-A., Ficklin S.P., Cheng C-H. *et al*. Tripal v1.1: a standards-based toolkit for construction of online genetic and genomic databases. *Database* (2013) Vol. 2013: article ID bat075; doi:10.1093/database/bat075.



QUICK LINKS

Download or Upgrade

v2.0a Resources
Alpha Release for Drupal 7
New Functionality
Installation and Tutorial

v1.1 Resources
Stable Release for Drupal 6
Installation and Tutorial
Developer's Handbook
Tripal v1.1 Demo Site

NEWS

Tripal v2.0a Released

The Tripal Development Team is pleased to announce an alpha release of Tripal 2.0 for Drupal 7

More

NAVIGATION

- Feed aggregator

USER LOGIN

Username *

Password *

Request new password

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:

Lacey-Anne Sanderson, Stephen P. Ficklin, Chun-Huai Cheng, Sook Jung, Frank A. Feltus, Kirstin E. Bett, and Dorrie Main. Tripal v1.1: a standards-based toolkit for construction of online genetic and genomic databases. *Database,* Sept 2013: bat075 doi:10.1093/database/bat075

Stephen P. Ficklin, Lacey-Anne Sanderson, Chun-Huai Cheng, Margaret Staton, Taein Lee, Il-Hyung Cho, Sook Jung, Kirstin E Bett, Dorrie Main. Tripal: a construction Toolkit for Online Genome Databases. *Database*, Sept 2011: bar044 doi:10.1093/database/bar044

Contributing Organizations:

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)

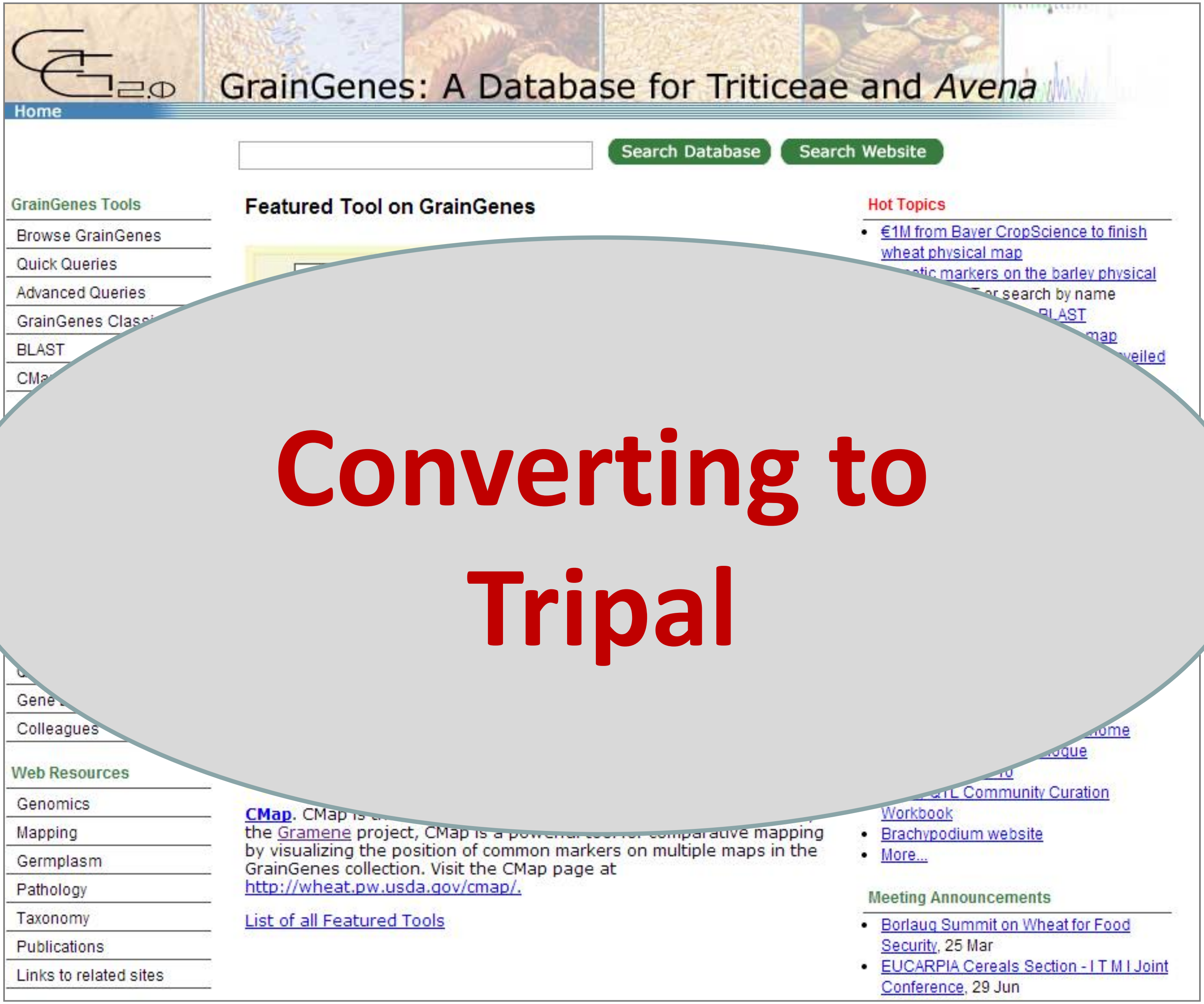
Tripal Timeline

- 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

- 2011: Development of the GeneNet Engine database. Clemson University (Alex Feltus/Stephen Ficklin)
- 2013 - 2015: Development of the WSU Cereals Database. (**$200K** Washington Cereals Commission, WSU)
- 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



GrainGenes: A Database for Triticeae and *Avena*

Home

Search Database
Search Website

GrainGenes Tools

- Browse GrainGenes
- Quick Queries
- Advanced Queries
- GrainGenes Class
- BLAST
- CMa
- Gene
- Colleagues

Web Resources

- Genomics
- Mapping
- Germplasm
- Pathology
- Taxonomy
- Publications
- Links to related sites

Featured Tool on GrainGenes

CMap. CMap is
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.pw.usda.gov/cmap/.

List of all Featured Tools

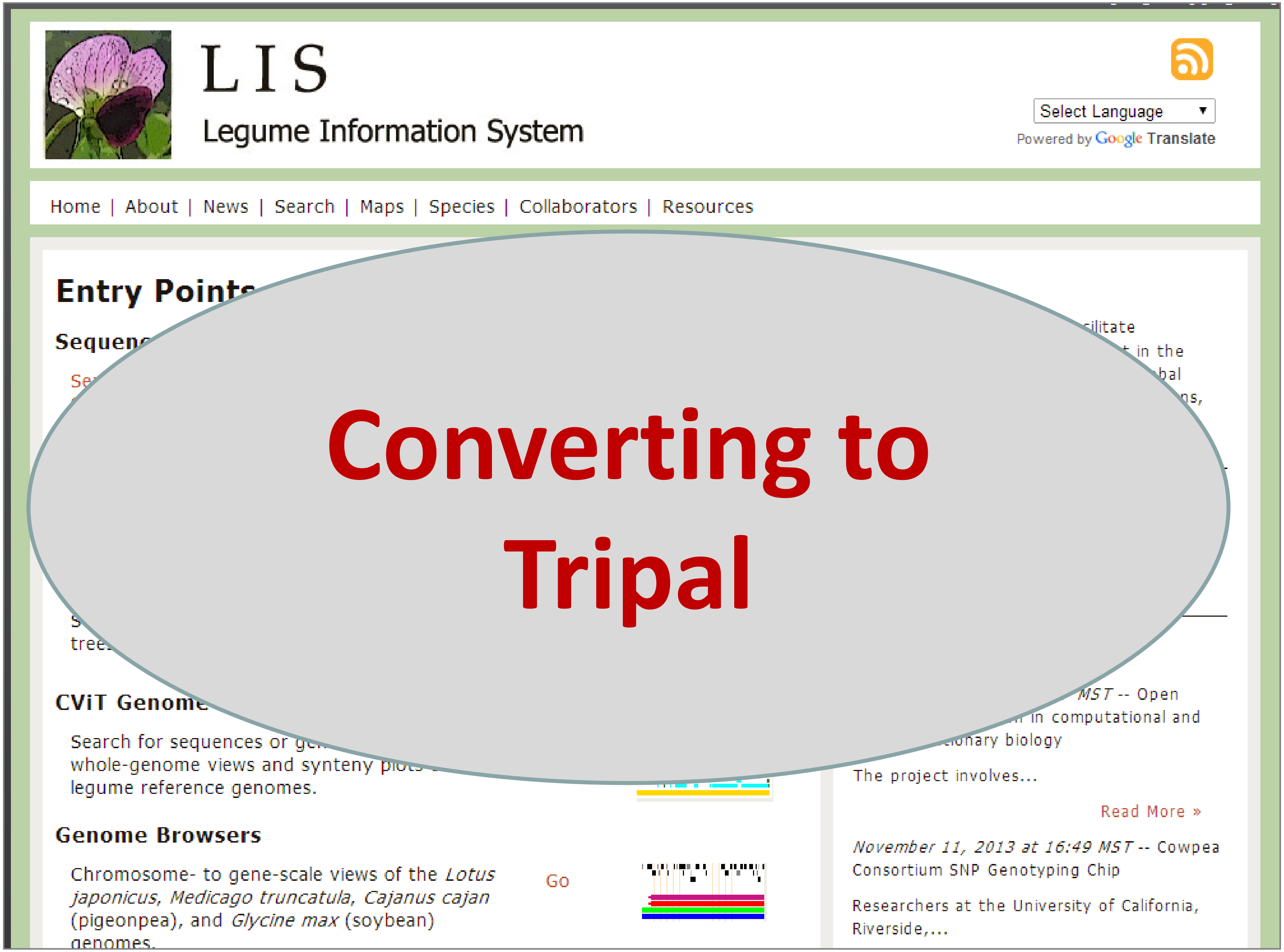
Hot Topics

- €1M from Bayer CropScience to finish wheat physical map
- tic markers on the barley physical
  or search by name
  BLAST
  map
  veiled
- ome
- ogue
- QTL Community Curation Workbook
- Brachypodium website
- More...

Meeting Announcements

- Borlaug Summit on Wheat for Food Security, 25 Mar
- EUCARPIA Cereals Section - I T M I Joint Conference, 29 Jun

Converting to Tripal



LIS

Legume Information System

Select Language ▼

Powered by Google Translate

Home | About | News | Search | Maps | Species | Collaborators | Resources

Entry Points

Sequen

Se

S
tree

CViT Genome

Search for sequences or ge
whole-genome views and synteny plots
legume reference genomes.

Genome Browsers

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

Go

cilitate
t in the
bal
ns,

MST -- Open
in computational and
ionary 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,...

Converting to Tripal

Home Site Map Site Stats Contact Us

Dendrome

A Forest Tree Genome Database

Welcome Research TreeGenes DiversiTree FTGSC iPlant Resources Events News Jobs Links Help Downloads

Welcome to the Dendrome Proje

Dendrome is a collection
databases and oth
resources for
communi
collab
for m

In
Th
res
CGN
throu
conifer
the Dendr

Learn more abou
Join CGN

NATURE GENETICS

EVENTS

ium on Biology of Rare

sembly (v1.01) of

ee (v2.0.0), a map interface that works with DiversiTree to bring together genomics, ecological, and trait data is now live!

CartograTree(v3.0.0) Beta is available for testing.

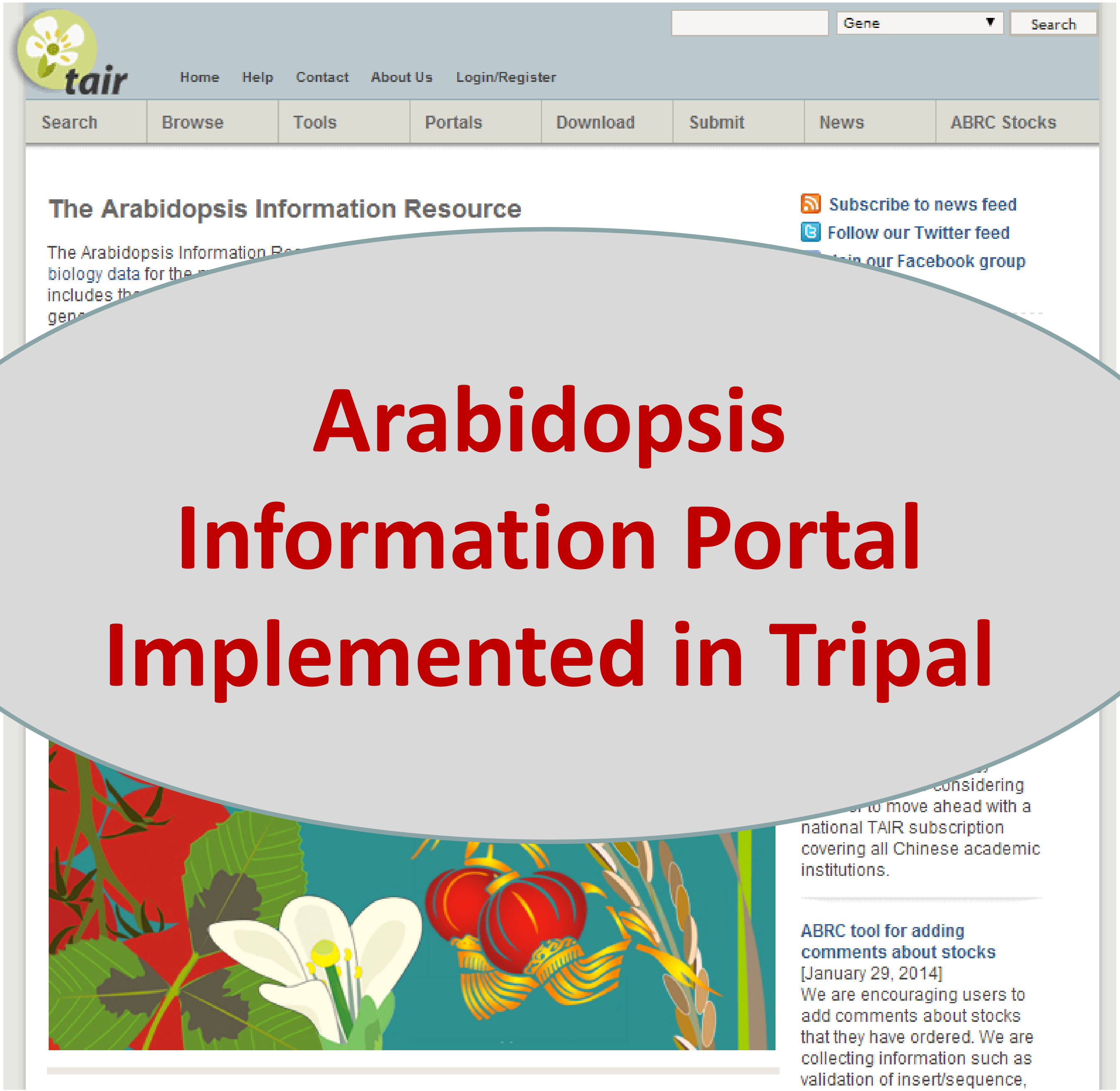
SMarTForests Project releases the third assembly of the Picea glauca genome.

TREEGENES CONNECT

Project Resources

- AdapTree
- Conifer Translational Genomics Network

Arabidopsis Information Portal Implemented in Tripal



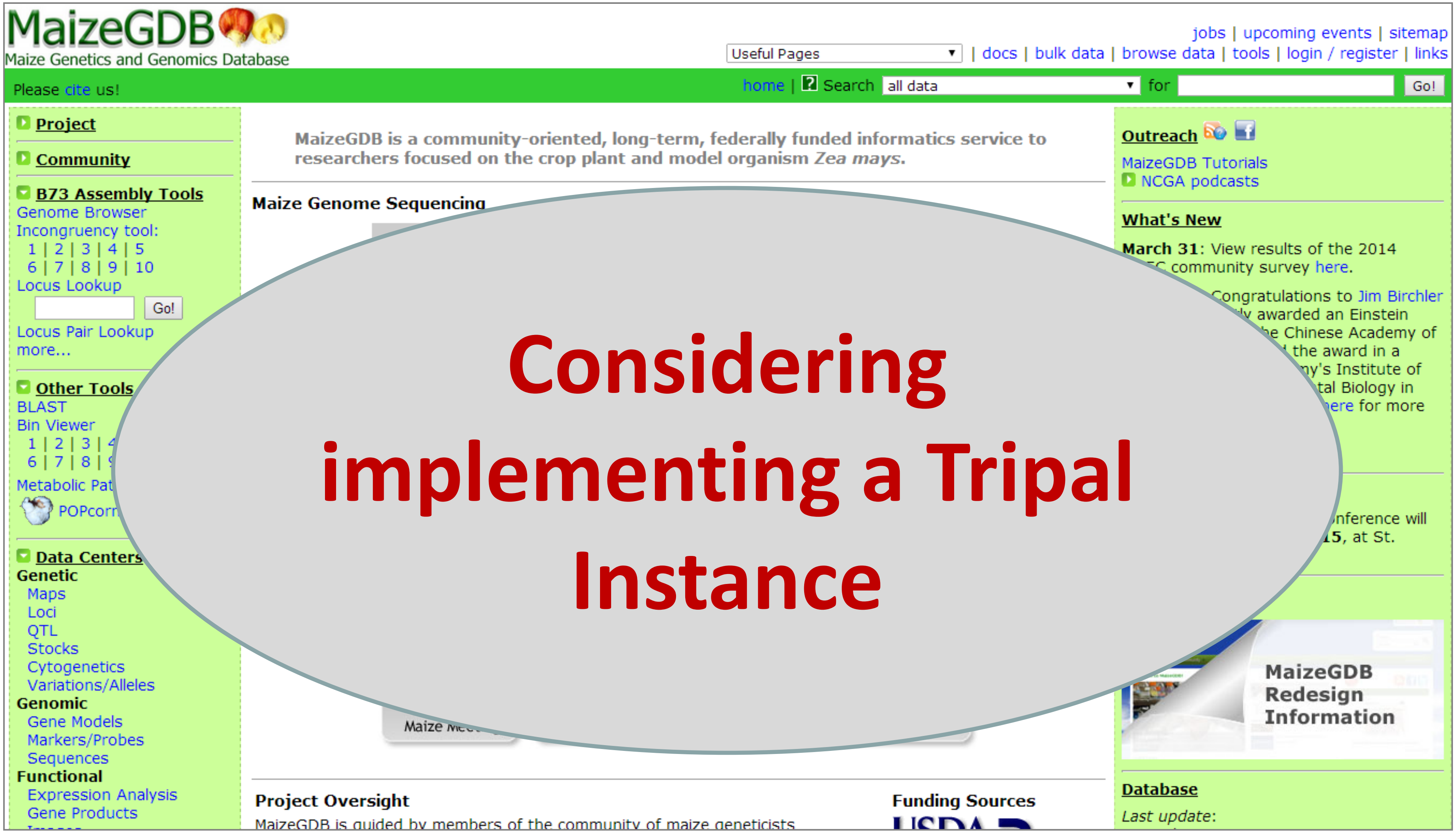
The Arabidopsis Information Resource

The Arabidopsis Information Re
biology data for the m
includes th
gen

considering
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 that they have ordered. We are collecting information such as validation of insert/sequence,

Considering implementing a Tripal Instance



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

March 31: View results of the 2014 …C community survey here.

Congratulations to Jim Birchler … awarded an Einstein … the Chinese Academy of … the award in a … my's Institute of … tal Biology in … here for more

… nference will … **15**, at St.

Project Oversight

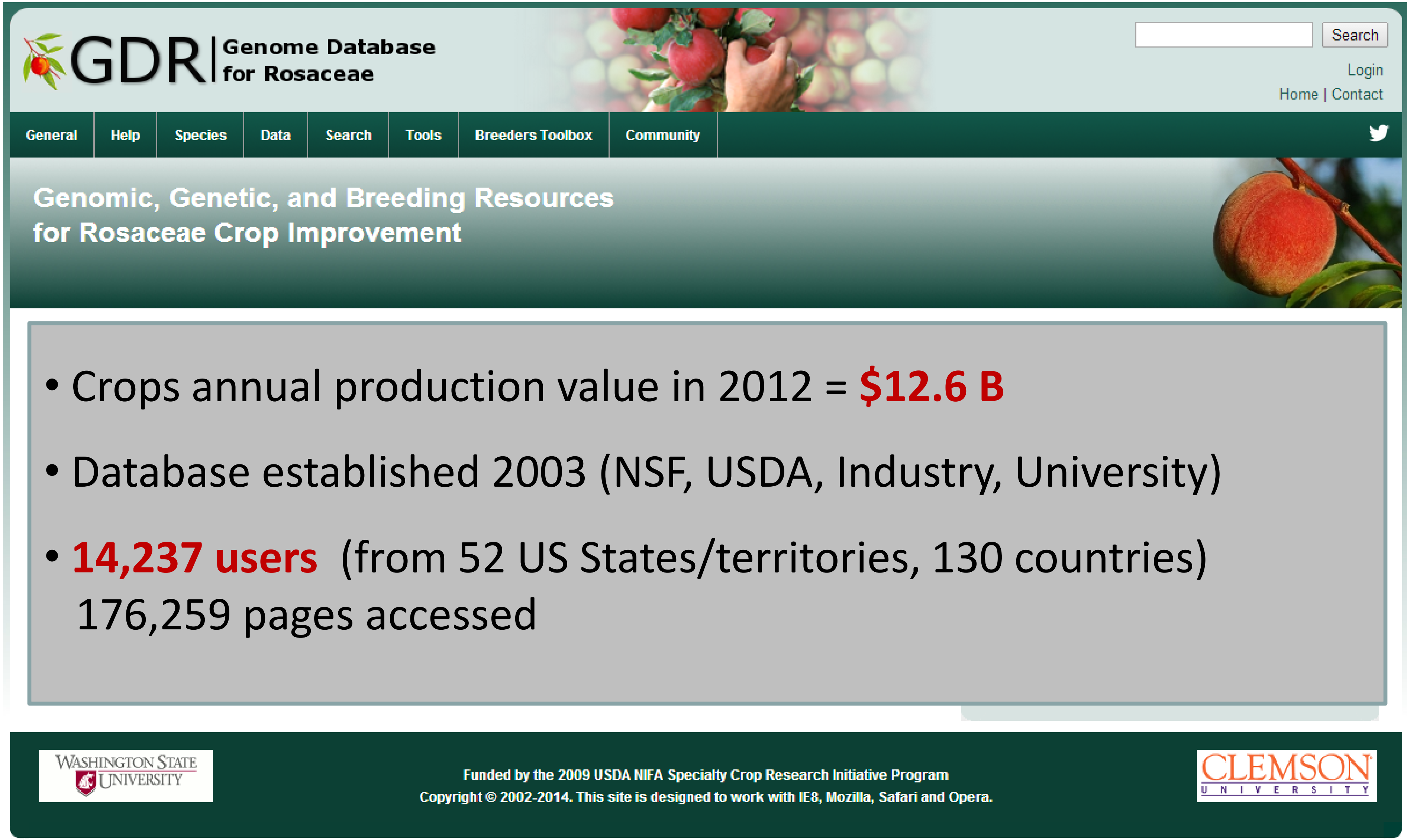
MaizeGDB is guided by members of the community of maize geneticists

Other Confirmed Tripal Databases

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

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.



GDR | Genome Database for Rosaceae

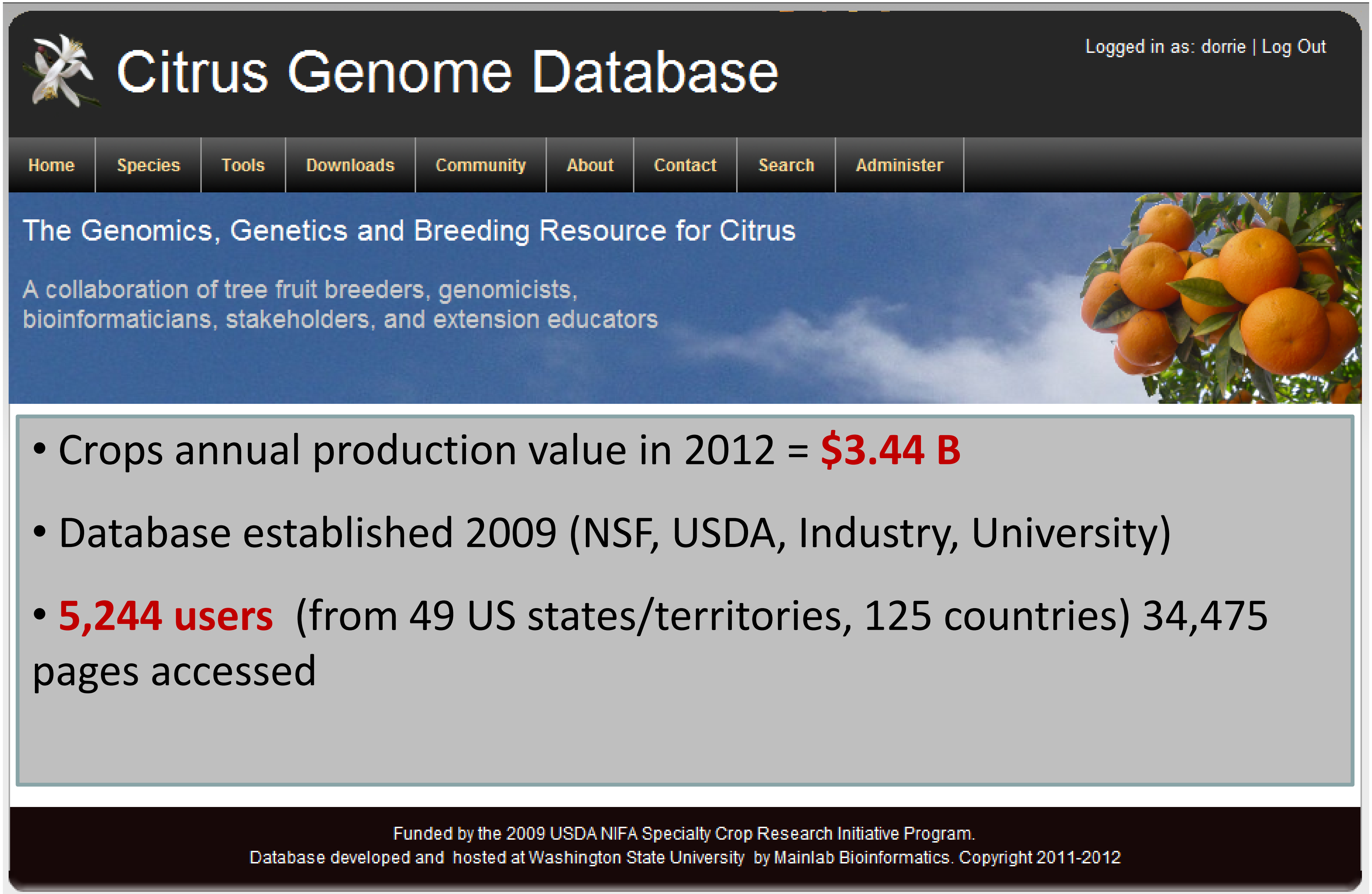
Genomic, Genetic, and Breeding Resources for Rosaceae Crop Improvement

- 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

Washington State University

Funded by the 2009 USDA NIFA Specialty Crop Research Initiative Program

Clemson University



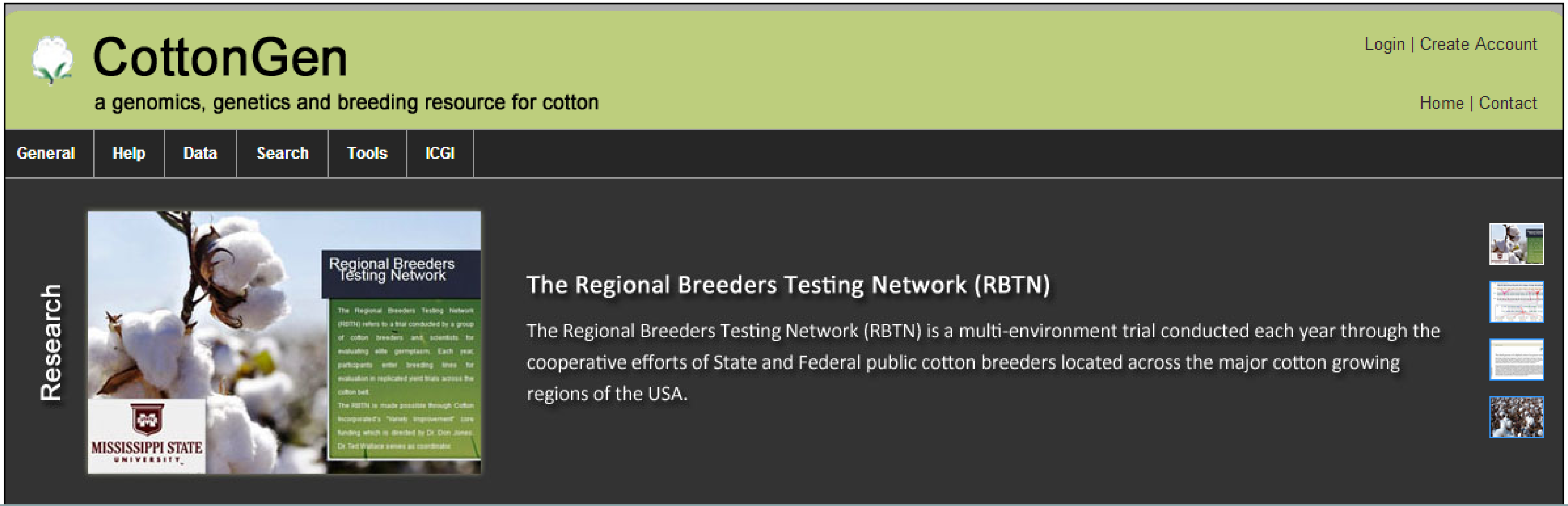
Citrus Genome Database

The Genomics, Genetics and Breeding Resource for Citrus

A collaboration of tree fruit breeders, genomicists, bioinformaticians, stakeholders, and extension educators

- 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

Funded by the 2009 USDA NIFA Specialty Crop Research Initiative Program.
Database developed and hosted at Washington State University by Mainlab Bioinformatics.



CottonGen

a genomics, genetics and breeding resource for cotton

The Regional Breeders Testing Network (RBTN)

The Regional Breeders Testing Network (RBTN) is a multi-environment trial conducted each year through the cooperative efforts of State and Federal public cotton breeders located across the major cotton growing regions of the USA.

- 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

CottonGen Homepage

dii published

oid cotton *Gossypium raimondii* published

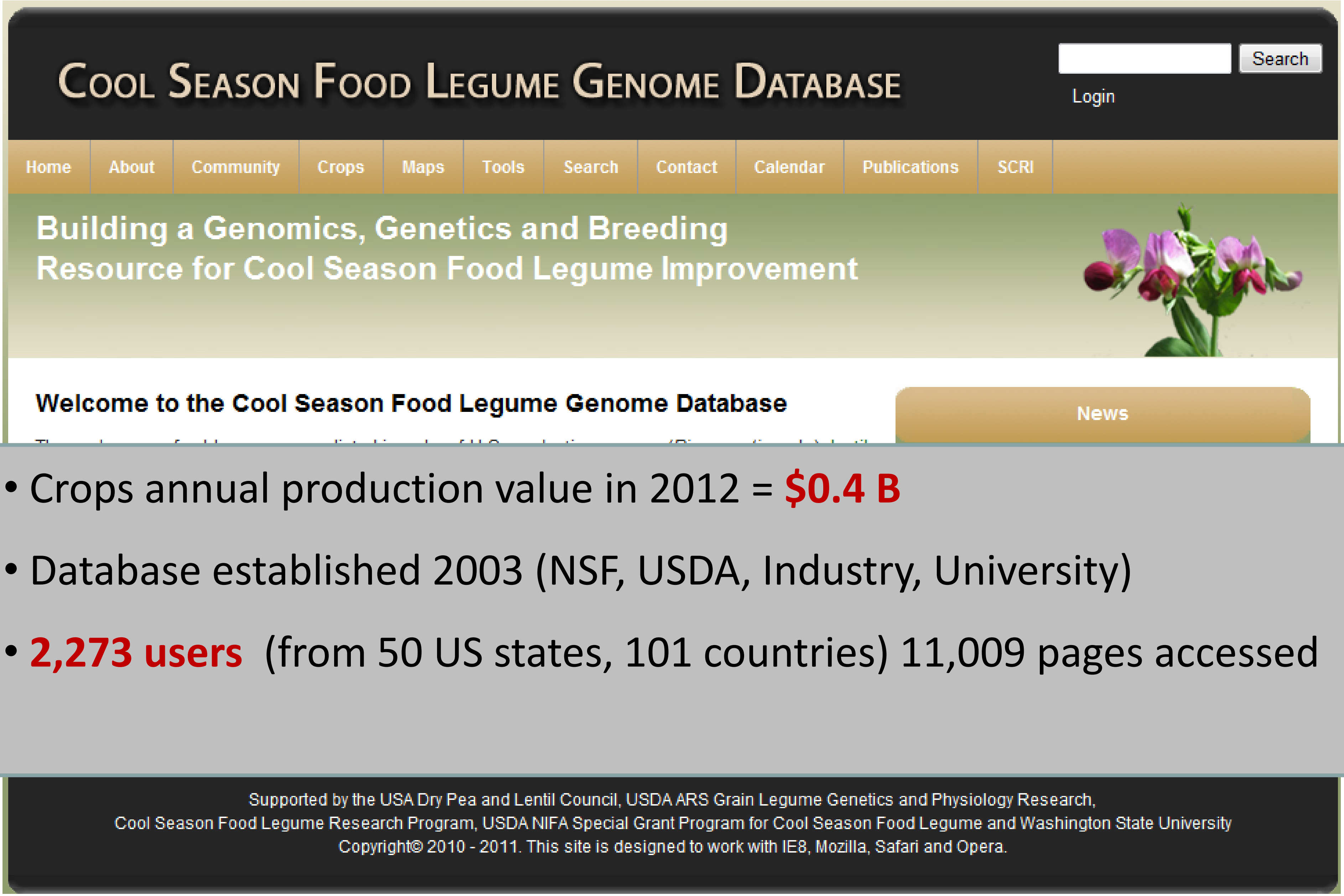
g *et al.*, 2012.

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 faciliate community communication and collaboration through housing the ICGI website, forums and mailing lists. This project is funded by Cotton Incorporated, the USDA-ARS Crop

News

- Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Paterson et al., Nature 492:423-427
- International Cotton Genome Initiative (ICGI) Workshop at PAG XXI
- Beltwide Cotton Conference 2013
- Draft G. raimondii genome sequence and



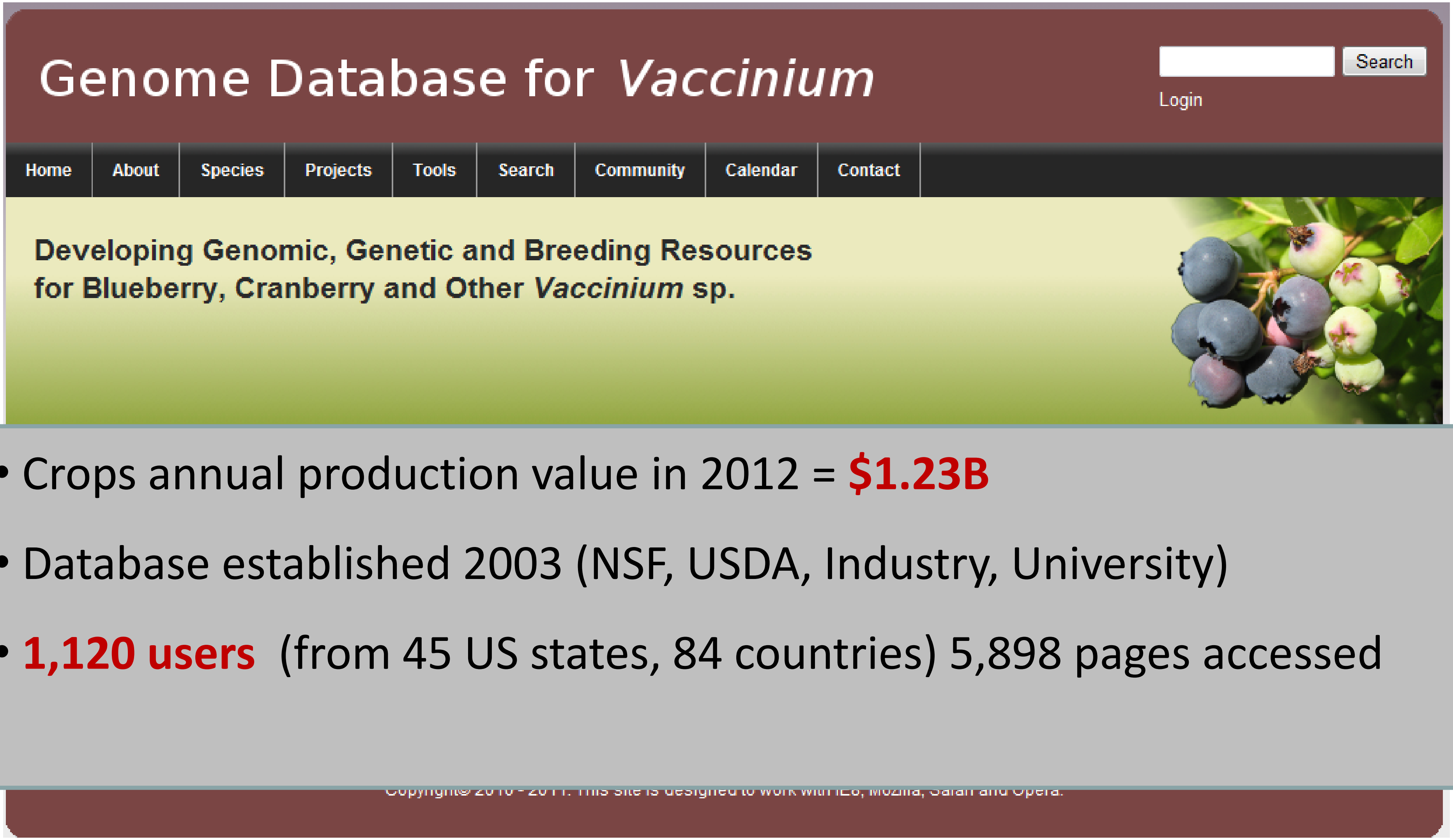
COOL SEASON FOOD LEGUME GENOME DATABASE

Building a Genomics, Genetics and Breeding Resource for Cool Season Food Legume Improvement

Welcome to the Cool Season Food Legume Genome Database

- 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

Supported by the USA Dry Pea and Lentil Council, USDA ARS Grain Legume Genetics and Physiology Research, Cool Season Food Legume Research Program, USDA NIFA Special Grant Program for Cool Season Food Legume and Washington State University

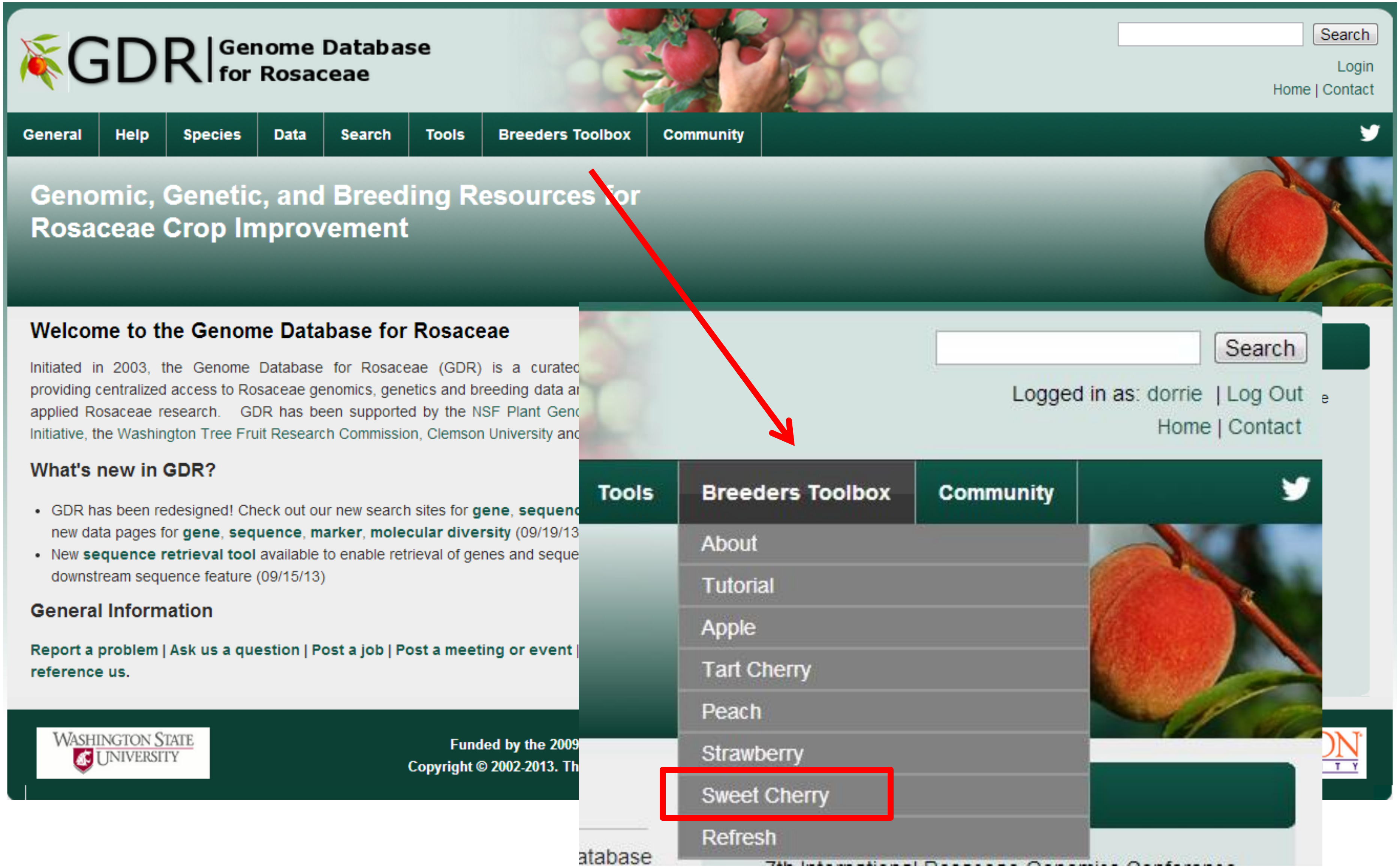


Genome Database for *Vaccinium*

Developing Genomic, Genetic and Breeding Resources for Blueberry, Cranberry and Other *Vaccinium* sp.

- 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

Current Functionality of PNWSCBP ToolBox



Genomic, Genetic, and Breeding Resources for Rosaceae Crop Improvement

Welcome to the Genome Database for Rosaceae

Initiated in 2003, the Genome Database for Rosaceae (GDR) is a curated providing centralized access to Rosaceae genomics, genetics and breeding data a applied Rosaceae research. GDR has been supported by the NSF Plant Gen Initiative, the Washington Tree Fruit Research Commission, Clemson University and

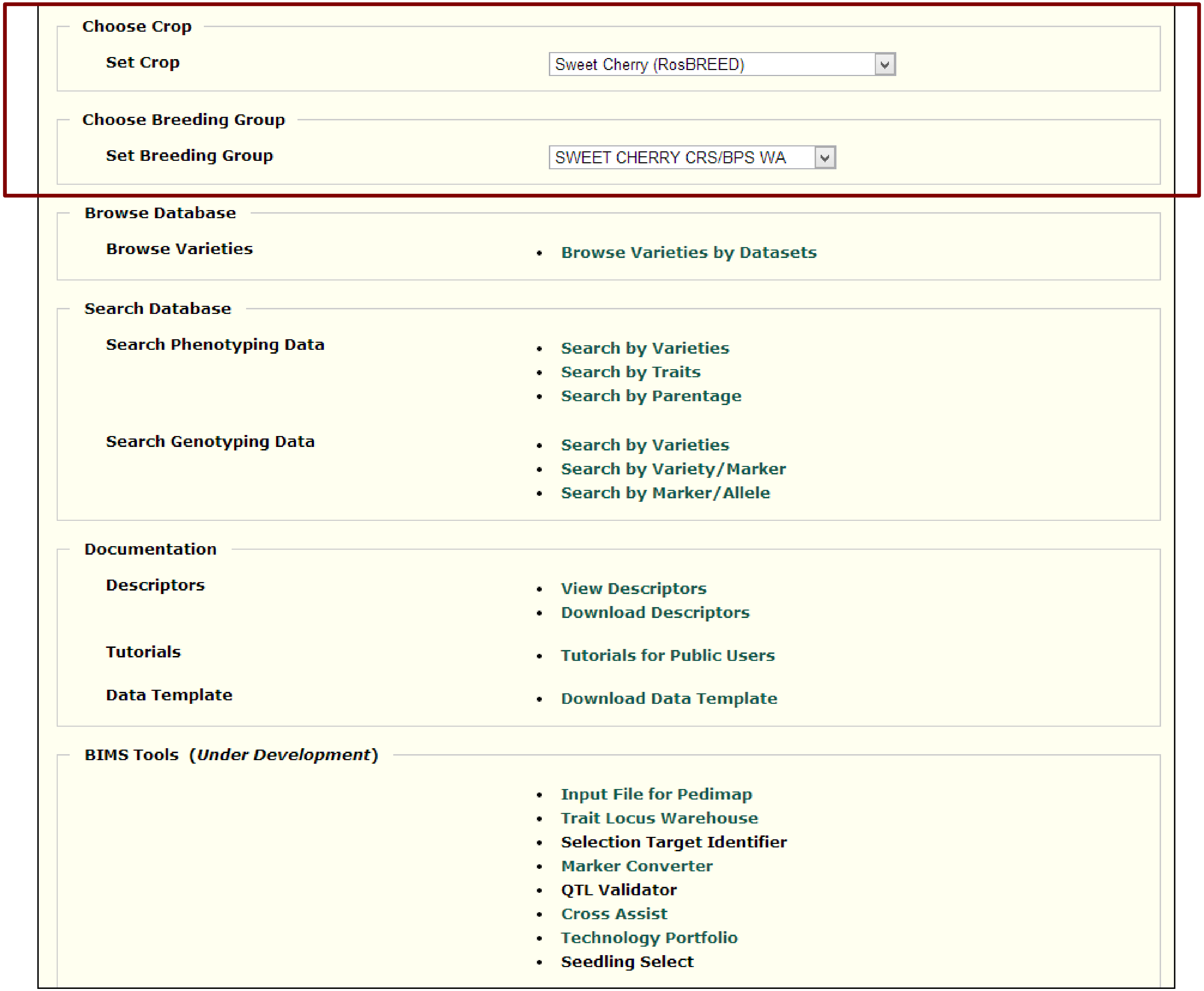
What's new in GDR?

- GDR has been redesigned! Check out our new search sites for gene, sequen new data pages for gene, sequence, marker, molecular diversity (09/19/13
- New sequence retrieval tool available to enable retrieval of genes and seque downstream sequence feature (09/15/13)

General Information

Report a problem | Ask us a question | Post a job | Post a meeting or event | reference us.

Funded by the 2009
Copyright © 2002-2013. Th



Choose Crop

Set Crop Sweet Cherry (RosBREED)

Choose Breeding Group

Set Breeding Group SWEET CHERRY CRS/BPS WA

Browse Database

Browse Varieties
- Browse Varieties by Datasets

Search Database

Search Phenotyping Data
- Search by Varieties
- Search by Traits
- Search by Parentage

Search Genotyping Data
- Search by Varieties
- Search by Variety/Marker
- Search by Marker/Allele

Documentation

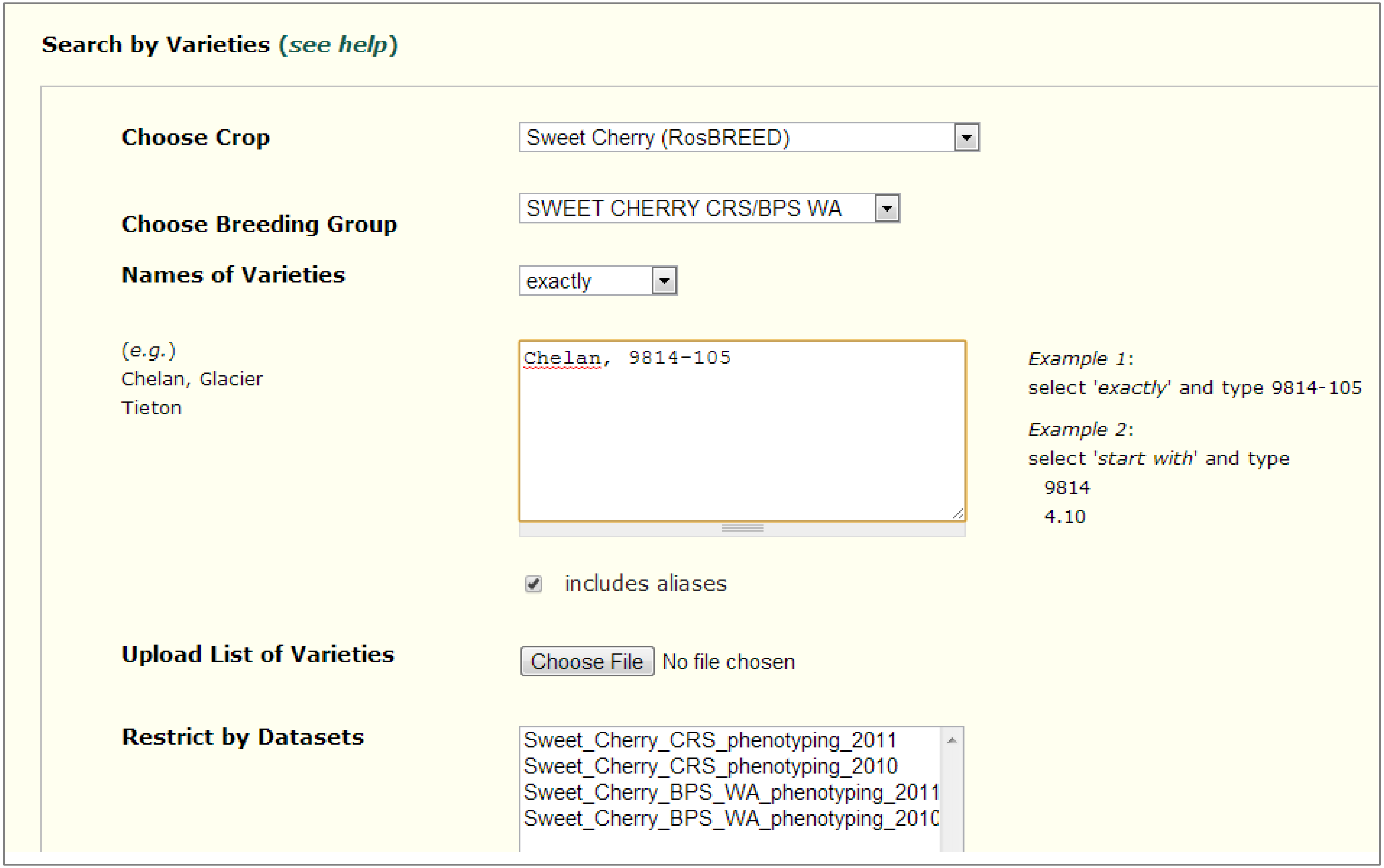
Descriptors
- View Descriptors
- Download Descriptors

Tutorials
- Tutorials for Public Users

Data Template
- Download Data Template

BIMS Tools (*Under Development*)
- Input File for Pedimap
- Trait Locus Warehouse
- Selection Target Identifier
- Marker Converter
- QTL Validator
- Cross Assist
- Technology Portfolio
- Seedling Select

Phenotyping Data Search by Varieties



Search by Varieties (*see help*)

Choose Crop: Sweet Cherry (RosBREED)

Choose Breeding Group: SWEET CHERRY CRS/BPS WA

Names of Varieties: exactly

(*e.g.*)
Chelan, Glacier
Tieton

```
Chelan, 9814-105
```

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

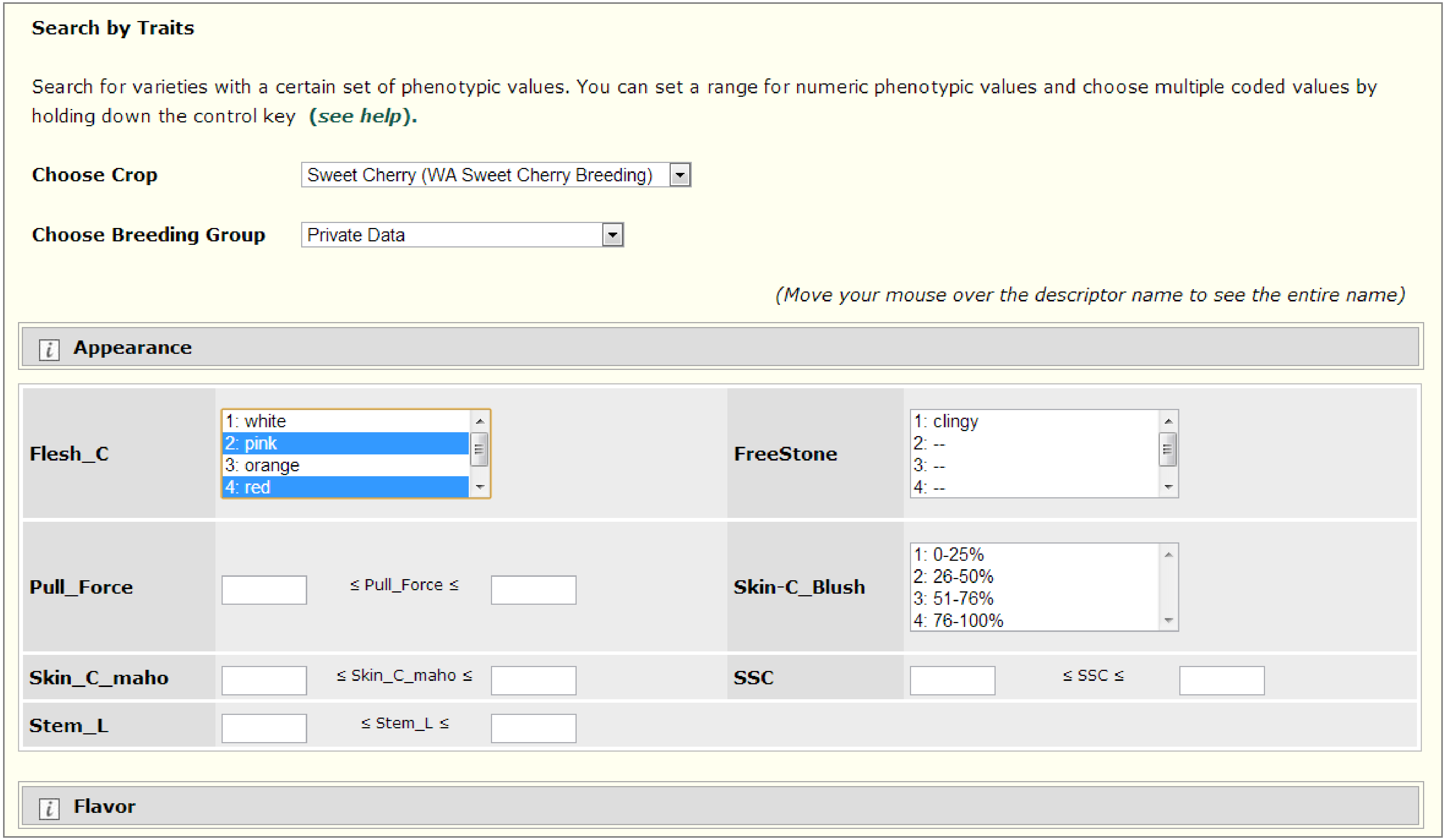
Example 2:
select '*start with*' and type
9814
4.10

☑ includes aliases

Upload List of Varieties: 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

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 Sweet Cherry (WA Sweet Cherry Breeding)

Choose Breeding Group Private Data

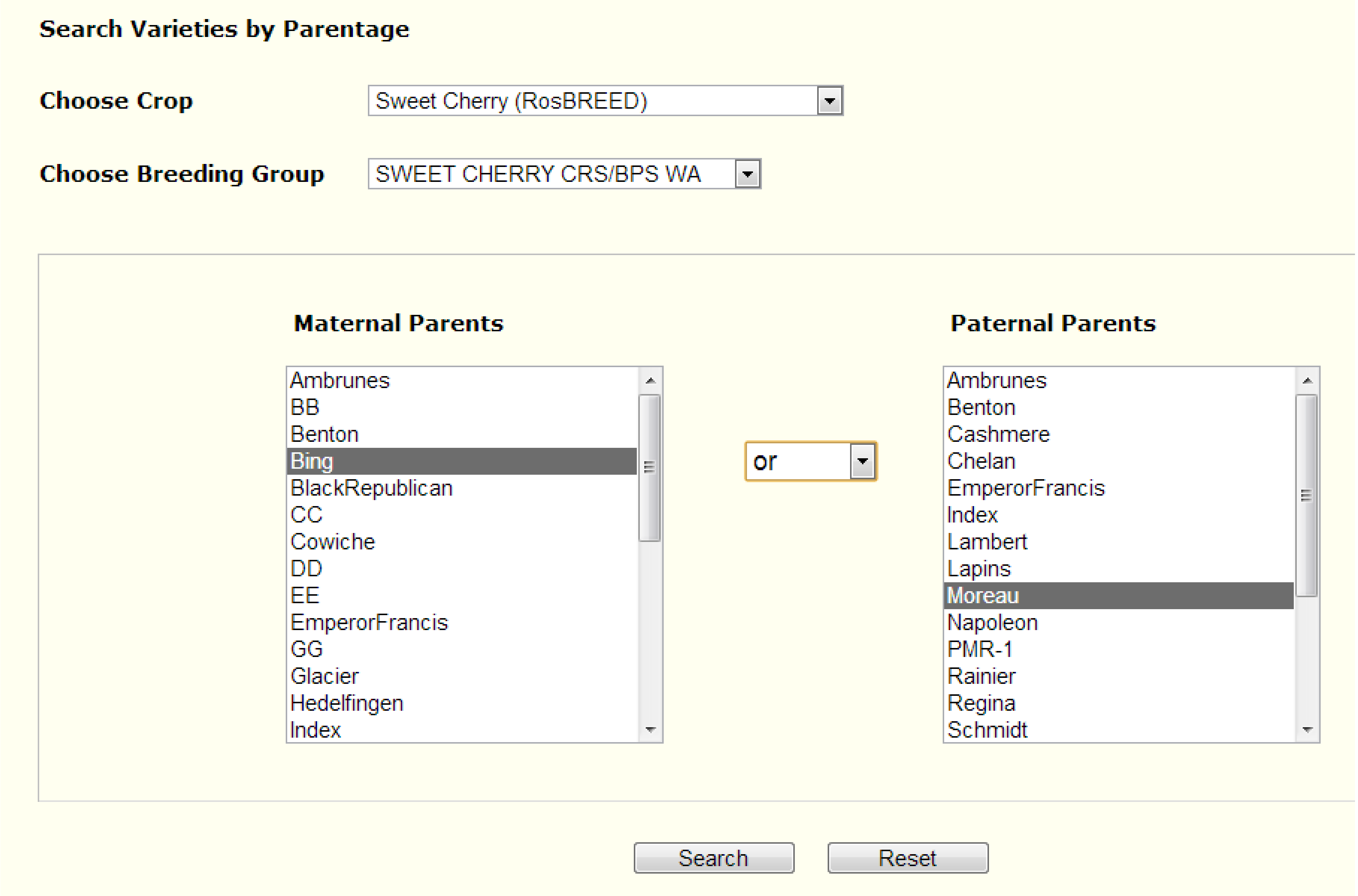
(Move your mouse over the descriptor name to see the entire name)

Appearance

Flesh_C	1: white 2: pink 3: orange 4: red	**FreeStone**	1: clingy 2: -- 3: -- 4: --
Pull_Force	≤ Pull_Force ≤	**Skin-C_Blush**	1: 0-25% 2: 26-50% 3: 51-76% 4: 76-100%
Skin_C_maho	≤ Skin_C_maho ≤	**SSC**	≤ SSC ≤
Stem_L	≤ Stem_L ≤		

Flavor

Phenotyping Data Search by Parentage



Search Varieties by Parentage

Choose Crop: Sweet Cherry (RosBREED)

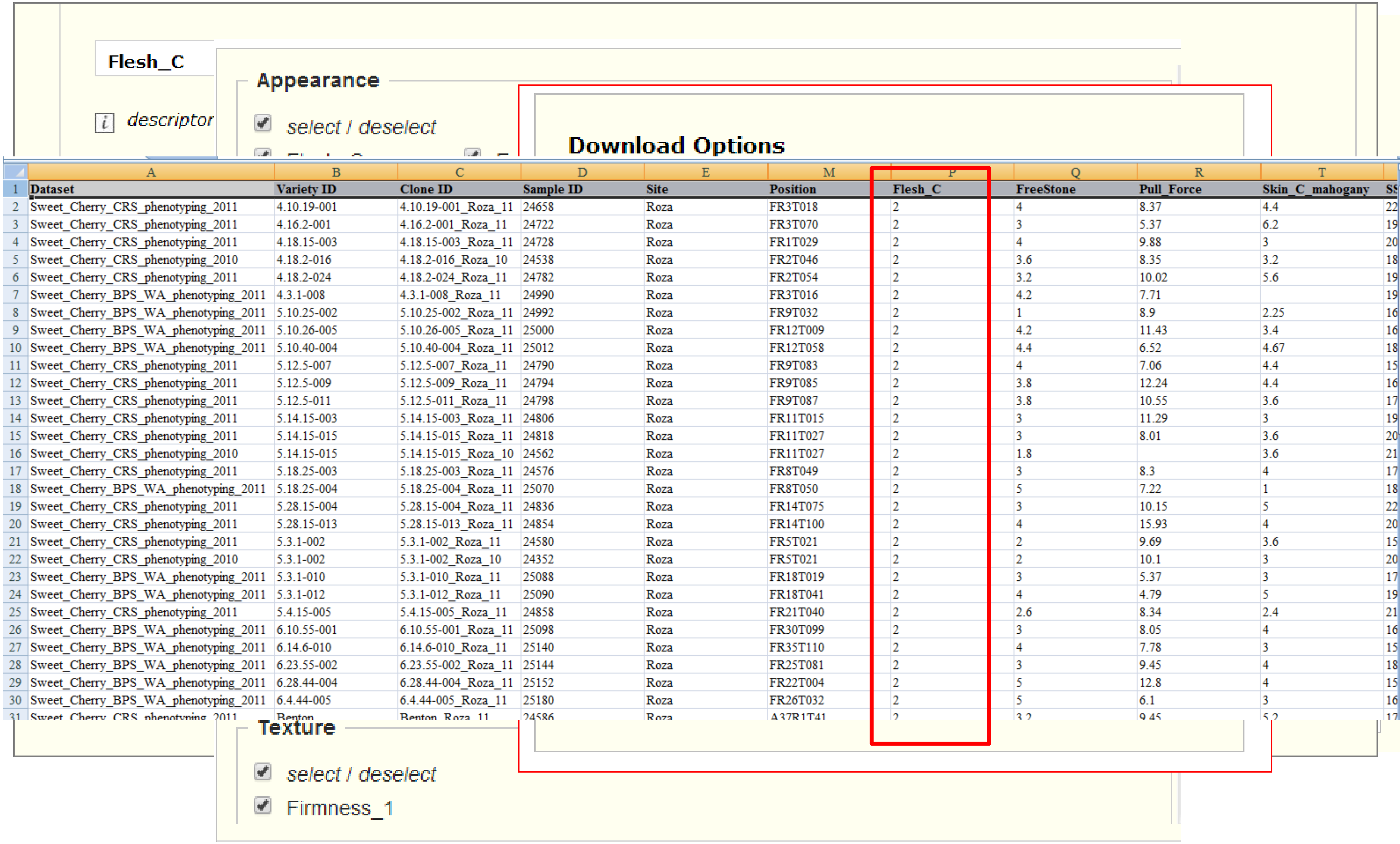
Choose Breeding Group: SWEET CHERRY CRS/BPS WA

Maternal Parents: Bing

or

Paternal Parents: Moreau

Phenotyping Data Trait Search Example



Flesh_C

descriptor

Appearance

☑ select / deselect

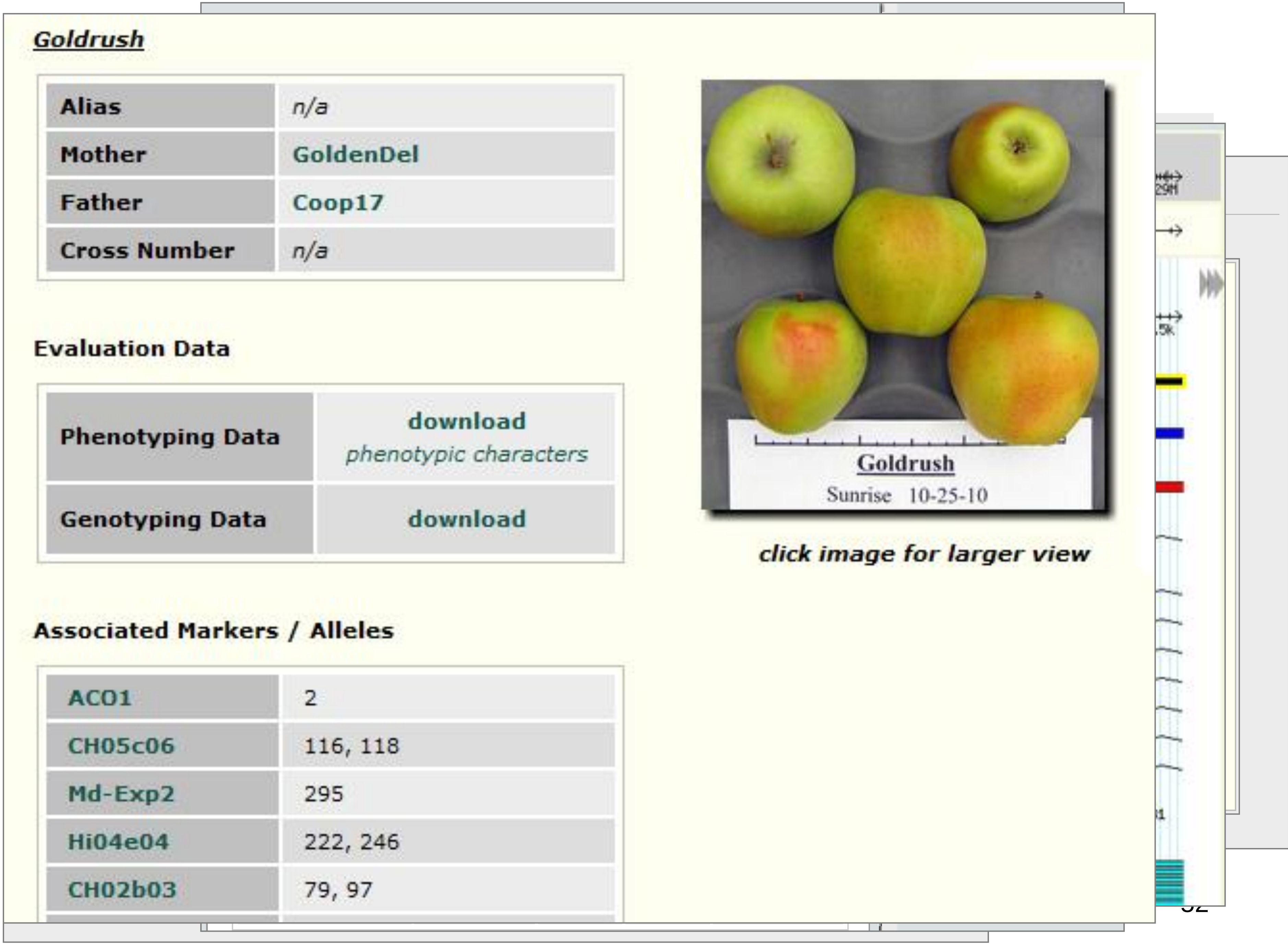
Download Options

	A	B	C	D	E	M	P	Q	R	T	
1	**Dataset**	**Variety ID**	**Clone ID**	**Sample ID**	**Site**	**Position**	**Flesh_C**	**FreeStone**	**Pull_Force**	**Skin_C_mahogany**	**SS**
2	Sweet_Cherry_CRS_phenotyping_2011	4.10.19-001	4.10.19-001_Roza_11	24658	Roza	FR3T018	2	4	8.37	4.4	22
3	Sweet_Cherry_CRS_phenotyping_2011	4.16.2-001	4.16.2-001_Roza_11	24722	Roza	FR3T070	2	3	5.37	6.2	19
4	Sweet_Cherry_CRS_phenotyping_2011	4.18.15-003	4.18.15-003_Roza_11	24728	Roza	FR1T029	2	4	9.88	3	20
5	Sweet_Cherry_CRS_phenotyping_2010	4.18.2-016	4.18.2-016_Roza_10	24538	Roza	FR2T046	2	3.6	8.35	3.2	18
6	Sweet_Cherry_CRS_phenotyping_2011	4.18.2-024	4.18.2-024_Roza_11	24782	Roza	FR2T054	2	3.2	10.02	5.6	19
7	Sweet_Cherry_BPS_WA_phenotyping_2011	4.3.1-008	4.3.1-008_Roza_11	24990	Roza	FR3T016	2	4.2	7.71		19
8	Sweet_Cherry_BPS_WA_phenotyping_2011	5.10.25-002	5.10.25-002_Roza_11	24992	Roza	FR9T032	2	1	8.9	2.25	16
9	Sweet_Cherry_BPS_WA_phenotyping_2011	5.10.26-005	5.10.26-005_Roza_11	25000	Roza	FR12T009	2	4.2	11.43	3.4	16
10	Sweet_Cherry_BPS_WA_phenotyping_2011	5.10.40-004	5.10.40-004_Roza_11	25012	Roza	FR12T058	2	4.4	6.52	4.67	18
11	Sweet_Cherry_CRS_phenotyping_2011	5.12.5-007	5.12.5-007_Roza_11	24790	Roza	FR9T083	2	4	7.06	4.4	15
12	Sweet_Cherry_CRS_phenotyping_2011	5.12.5-009	5.12.5-009_Roza_11	24794	Roza	FR9T085	2	3.8	12.24	4.4	16
13	Sweet_Cherry_CRS_phenotyping_2011	5.12.5-011	5.12.5-011_Roza_11	24798	Roza	FR9T087	2	3.8	10.55	3.6	17
14	Sweet_Cherry_CRS_phenotyping_2011	5.14.15-003	5.14.15-003_Roza_11	24806	Roza	FR11T015	2	3	11.29	3	19
15	Sweet_Cherry_CRS_phenotyping_2011	5.14.15-015	5.14.15-015_Roza_11	24818	Roza	FR11T027	2	3	8.01	3.6	20
16	Sweet_Cherry_CRS_phenotyping_2010	5.14.15-015	5.14.15-015_Roza_10	24562	Roza	FR11T027	2	1.8		3.6	21
17	Sweet_Cherry_CRS_phenotyping_2011	5.18.25-003	5.18.25-003_Roza_11	24576	Roza	FR8T049	2	3	8.3	4	17
18	Sweet_Cherry_BPS_WA_phenotyping_2011	5.18.25-004	5.18.25-004_Roza_11	25070	Roza	FR8T050	2	5	7.22	1	18
19	Sweet_Cherry_CRS_phenotyping_2011	5.28.15-004	5.28.15-004_Roza_11	24836	Roza	FR14T075	2	3	10.15	5	22
20	Sweet_Cherry_CRS_phenotyping_2011	5.28.15-013	5.28.15-013_Roza_11	24854	Roza	FR14T100	2	4	15.93	4	20
21	Sweet_Cherry_CRS_phenotyping_2011	5.3.1-002	5.3.1-002_Roza_11	24580	Roza	FR5T021	2	2	9.69	3.6	15
22	Sweet_Cherry_CRS_phenotyping_2010	5.3.1-002	5.3.1-002_Roza_10	24352	Roza	FR5T021	2	2	10.1	3	20
23	Sweet_Cherry_BPS_WA_phenotyping_2011	5.3.1-010	5.3.1-010_Roza_11	25088	Roza	FR18T019	2	3	5.37	3	17
24	Sweet_Cherry_BPS_WA_phenotyping_2011	5.3.1-012	5.3.1-012_Roza_11	25090	Roza	FR18T041	2	4	4.79	5	19
25	Sweet_Cherry_CRS_phenotyping_2011	5.4.15-005	5.4.15-005_Roza_11	24858	Roza	FR21T040	2	2.6	8.34	2.4	21
26	Sweet_Cherry_BPS_WA_phenotyping_2011	6.10.55-001	6.10.55-001_Roza_11	25098	Roza	FR30T099	2	3	8.05	4	16
27	Sweet_Cherry_BPS_WA_phenotyping_2011	6.14.6-010	6.14.6-010_Roza_11	25140	Roza	FR35T110	2	4	7.78	3	15
28	Sweet_Cherry_BPS_WA_phenotyping_2011	6.23.55-002	6.23.55-002_Roza_11	25144	Roza	FR25T081	2	3	9.45	4	18
29	Sweet_Cherry_BPS_WA_phenotyping_2011	6.28.44-004	6.28.44-004_Roza_11	25152	Roza	FR22T004	2	5	12.8	4	15
30	Sweet_Cherry_BPS_WA_phenotyping_2011	6.4.44-005	6.4.44-005_Roza_11	25180	Roza	FR26T032	2	5	6.1	3	16
31	Sweet_Cherry_CRS_phenotyping_2011	Benton	Benton_Roza_11	24586	Roza	A37R1T41	2	3.2	9.45	5.2	17

Texture

☑ select / deselect

☑ Firmness_1



Goldrush

Alias	*n/a*
Mother	GoldenDel
Father	Coop17
Cross Number	*n/a*

Evaluation Data

Phenotyping Data	download *phenotypic characters*
Genotyping Data	download

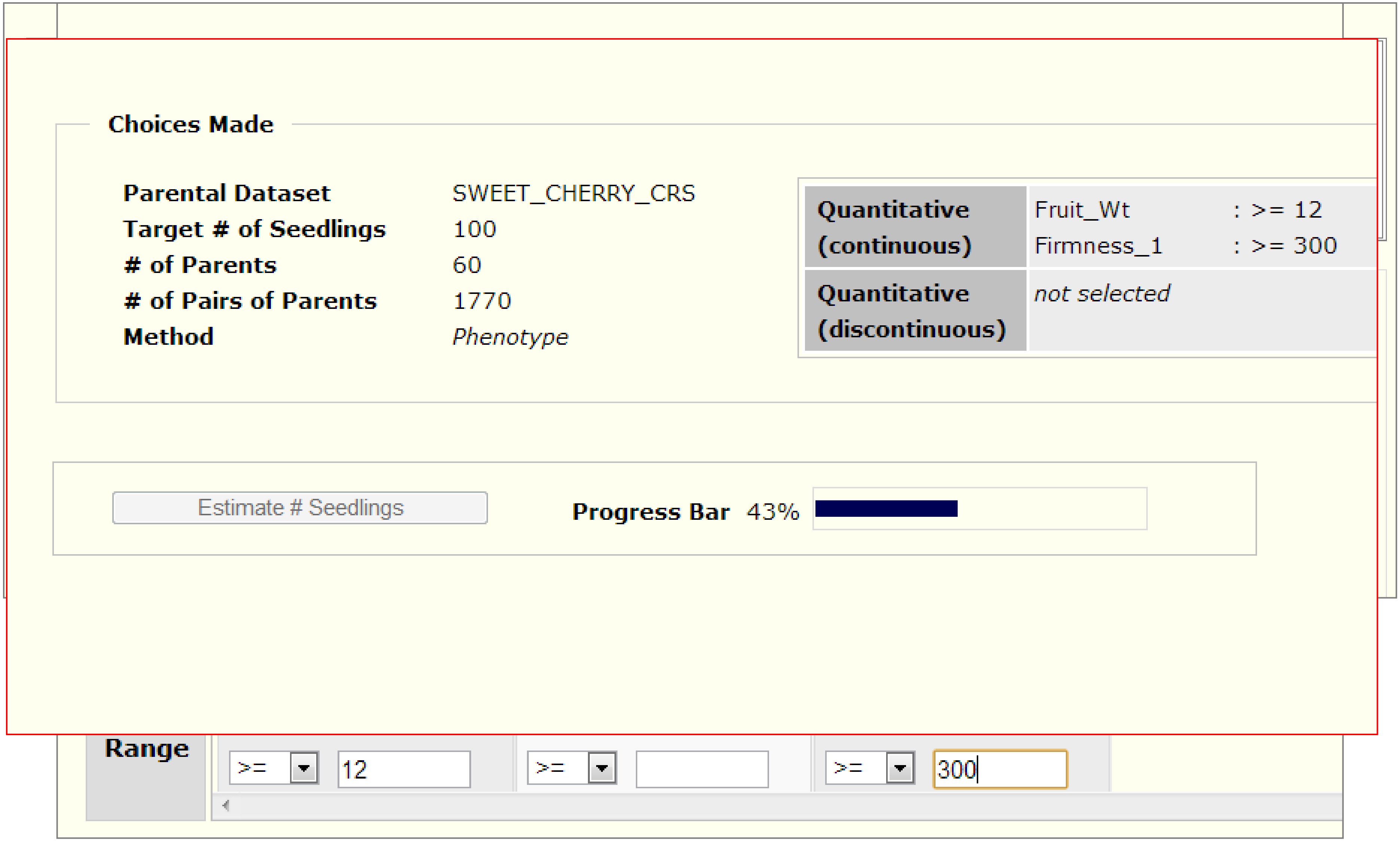
Associated Markers / Alleles

ACO1	2
CH05c06	116, 118
Md-Exp2	295
Hi04e04	222, 246
CH02b03	79, 97

Goldrush
Sunrise 10-25-10

click image for larger view

Cross Assist: Generates a list of parents and the number of seedlings to get the progeny with desired traits



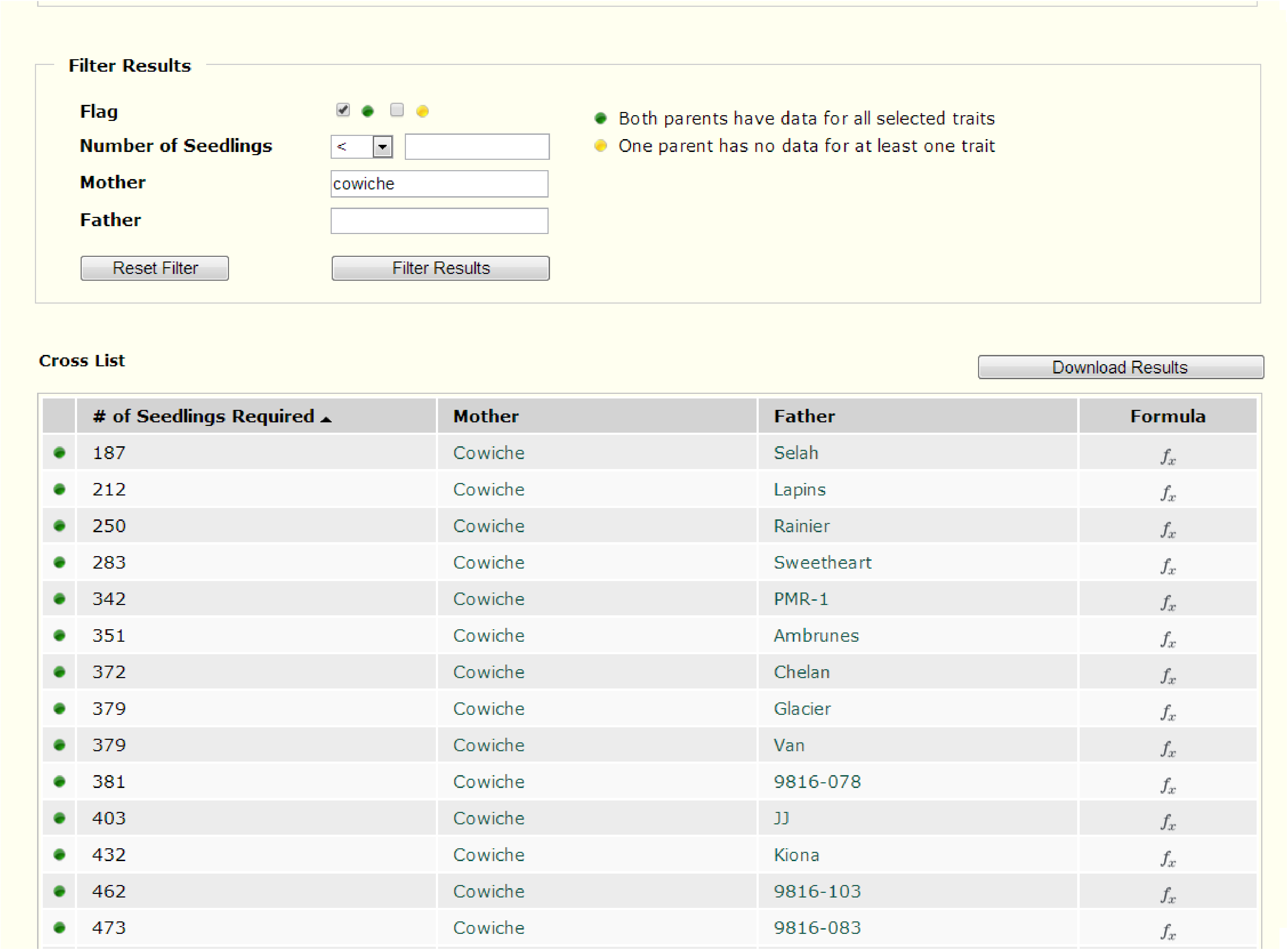
Choices Made

Parental Dataset SWEET_CHERRY_CRS
Target # of Seedlings 100
# of Parents 60
# of Pairs of Parents 1770
Method *Phenotype*

Quantitative (continuous)	Fruit_Wt : >= 12 Firmness_1 : >= 300
Quantitative (discontinuous)	*not selected*

Progress Bar 43%

Range >= 12 >= >= 300



Filter Results

Flag ☑ 🟢 ☐ 🟡

🟢 Both parents have data for all selected traits

🟡 One parent has no data for at least one trait

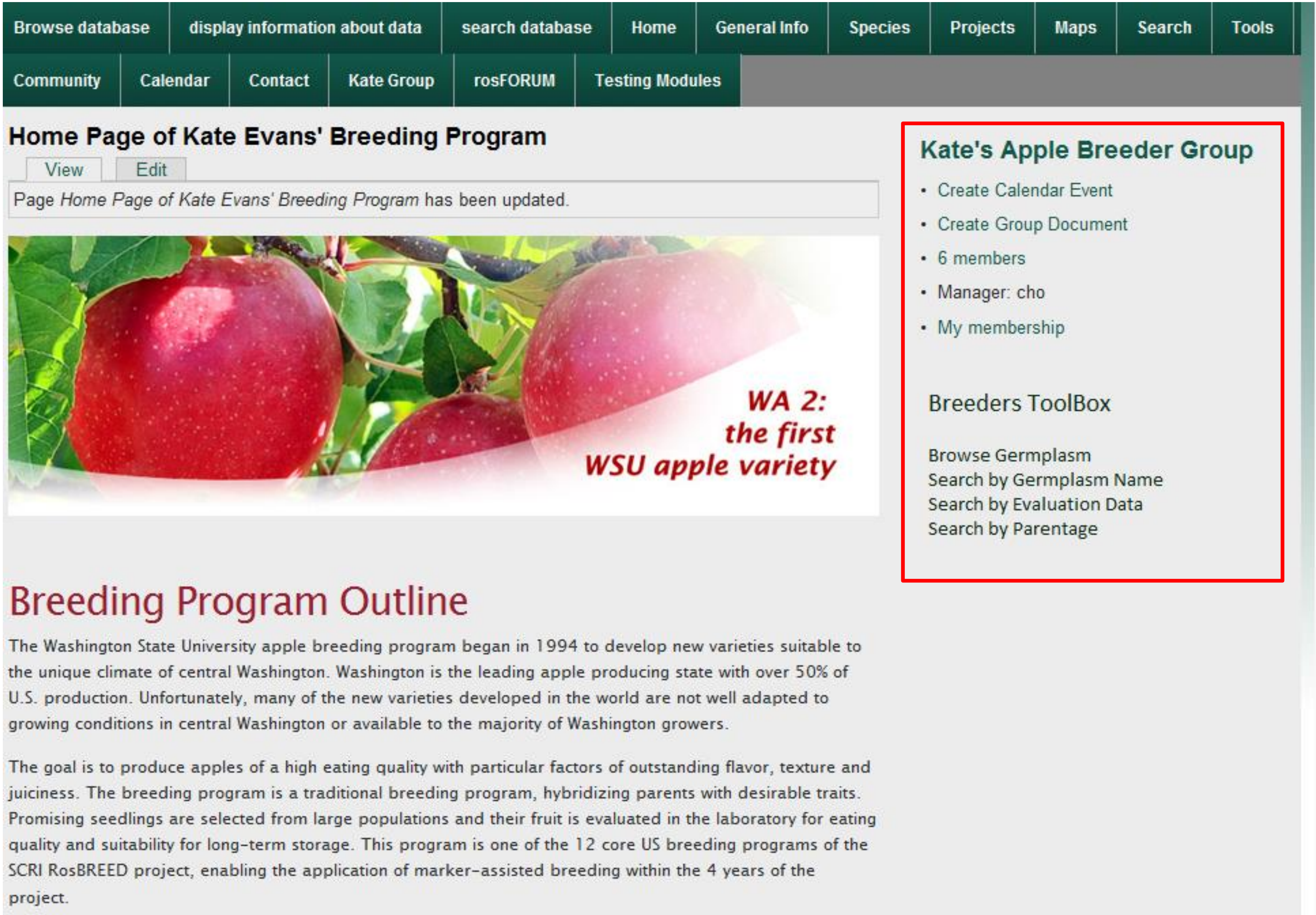
Number of Seedlings <

Mother cowiche

Father

Cross List

	# of Seedlings Required ▲	Mother	Father	Formula
🟢	187	Cowiche	Selah	f_x
🟢	212	Cowiche	Lapins	f_x
🟢	250	Cowiche	Rainier	f_x
🟢	283	Cowiche	Sweetheart	f_x
🟢	342	Cowiche	PMR-1	f_x
🟢	351	Cowiche	Ambrunes	f_x
🟢	372	Cowiche	Chelan	f_x
🟢	379	Cowiche	Glacier	f_x
🟢	379	Cowiche	Van	f_x
🟢	381	Cowiche	9816-078	f_x
🟢	403	Cowiche	JJ	f_x
🟢	432	Cowiche	Kiona	f_x
🟢	462	Cowiche	9816-103	f_x
🟢	473	Cowiche	9816-083	f_x



Home Page of Kate Evans' Breeding Program

Page *Home Page of Kate Evans' Breeding Program* has been updated.

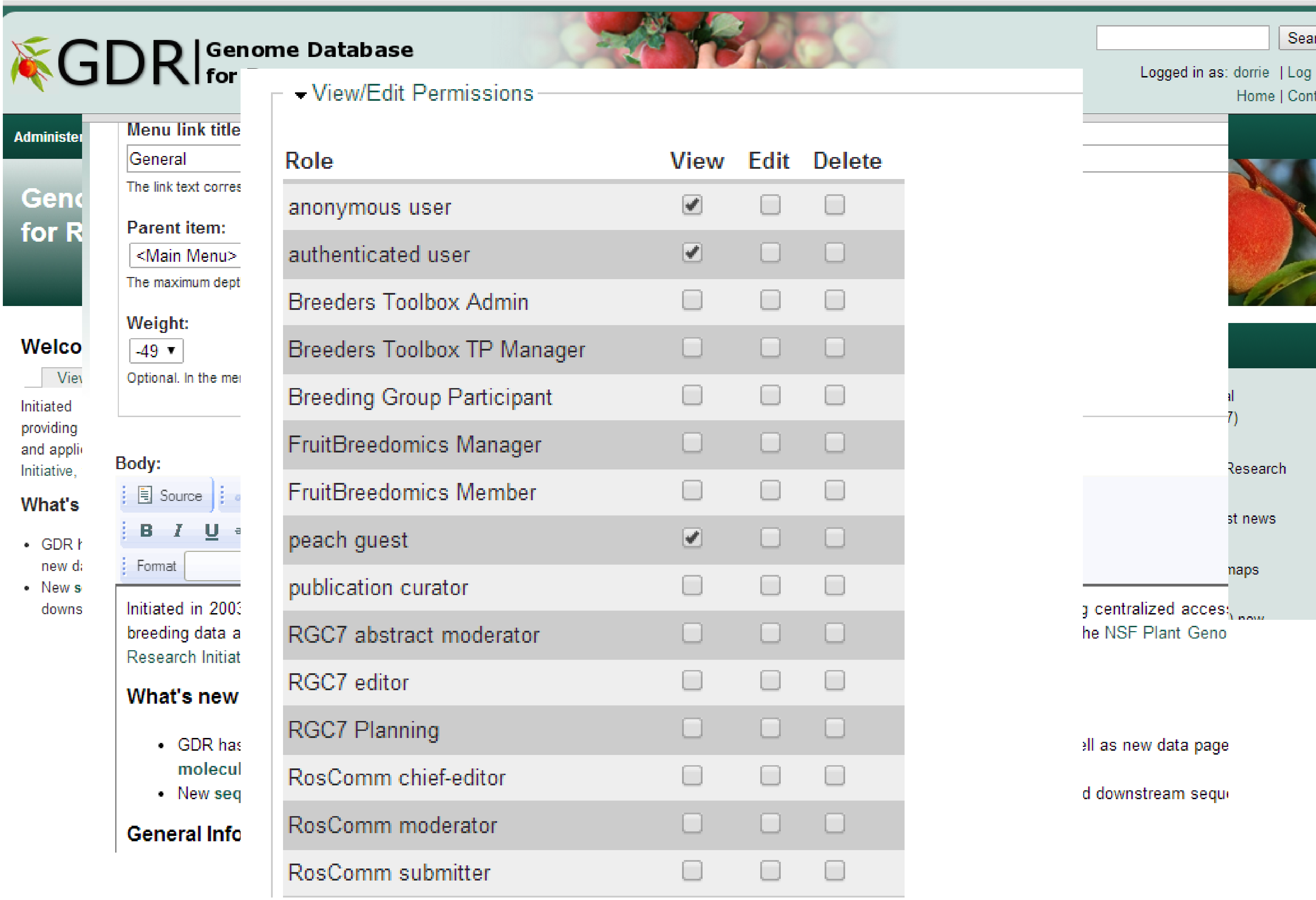
Kate's Apple Breeder Group

Breeders ToolBox

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.



View/Edit Permissions

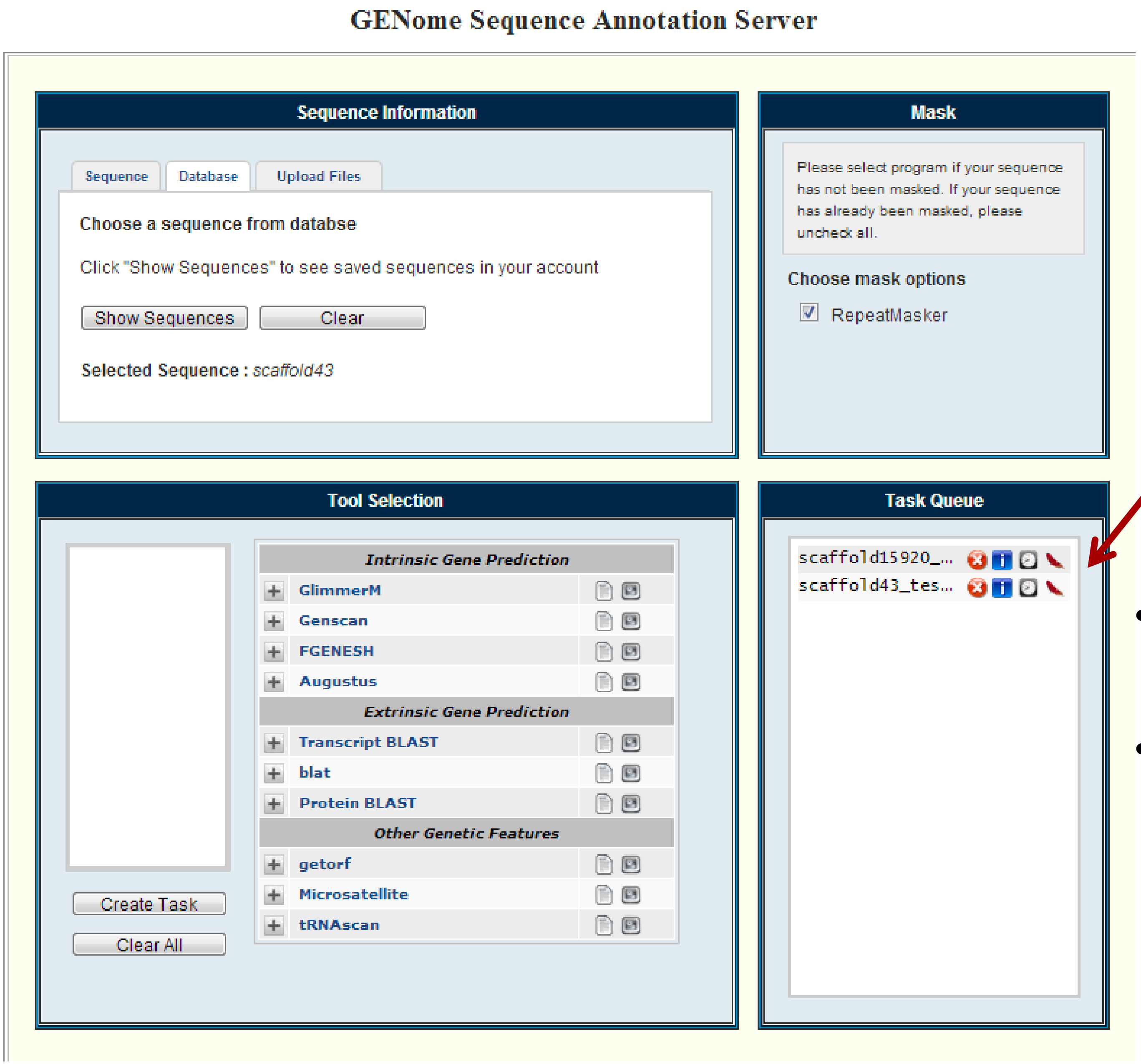
Role	View	Edit	Delete
anonymous user	☑	☐	☐
authenticated user	☑	☐	☐
Breeders Toolbox Admin	☐	☐	☐
Breeders Toolbox TP Manager	☐	☐	☐
Breeding Group Participant	☐	☐	☐
FruitBreedomics Manager	☐	☐	☐
FruitBreedomics Member	☐	☐	☐
peach guest	☑	☐	☐
publication curator	☐	☐	☐
RGC7 abstract moderator	☐	☐	☐
RGC7 editor	☐	☐	☐
RGC7 Planning	☐	☐	☐
RosComm chief-editor	☐	☐	☐
RosComm moderator	☐	☐	☐
RosComm submitter	☐	☐	☐

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



Choose a sequence from databse

Click "Show Sequences" to see saved sequences in your account

Selected Sequence : *scaffold43*

Please select program if your sequence has not been masked. If your sequence has already been masked, please uncheck all.

Choose mask options

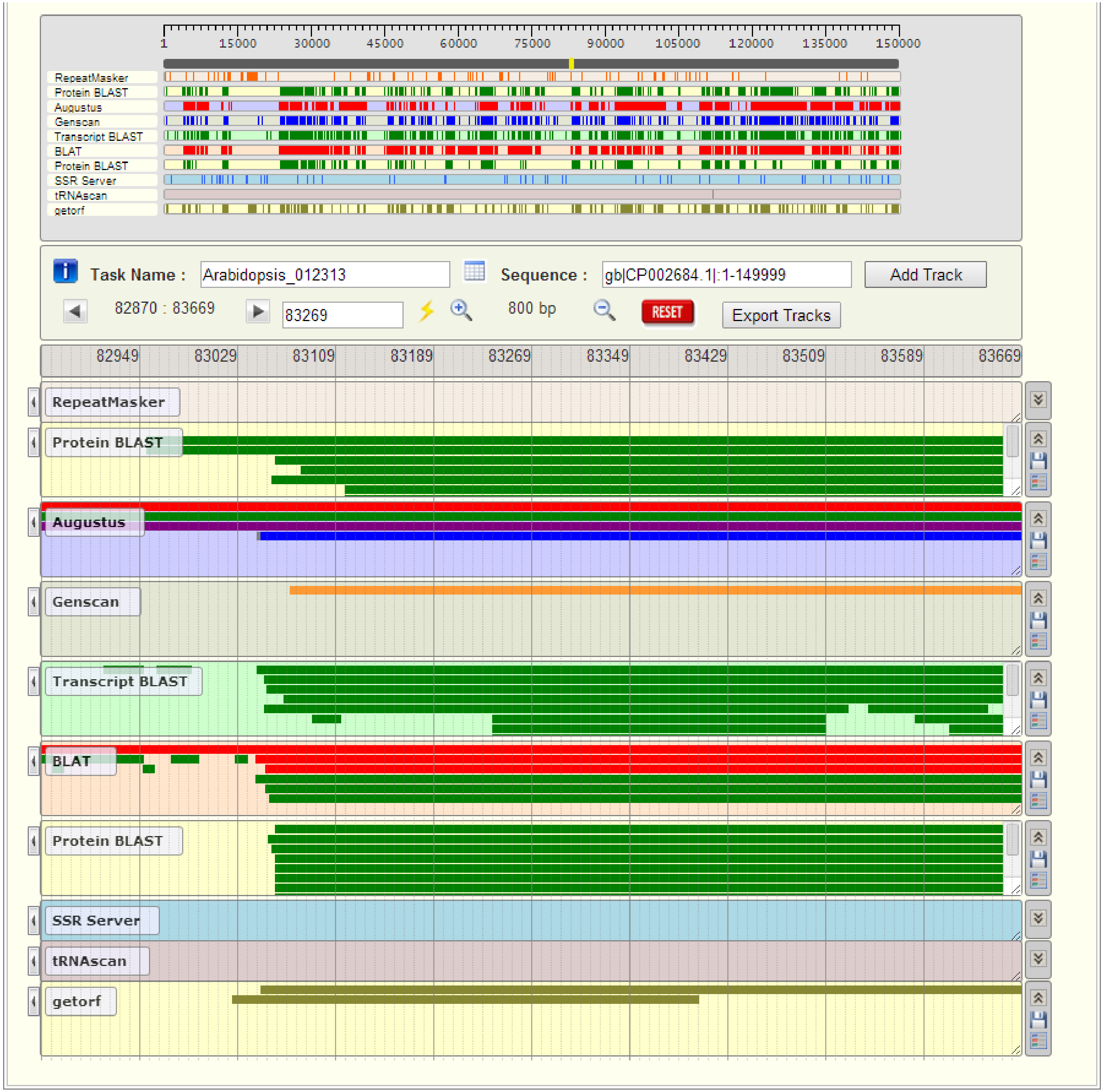
☑ RepeatMasker

scaffold15920_...

scaffold43_tes...

Tasks are given custom names and added added to the task queue

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



Task Name : Arabidopsis_012313

Sequence : gb|CP002684.1|:1-149999

82870 : 83669

83269

800 bp

Specific Objectives

1. Expand online community databases currently housing high quality genomic, genetic and breeding data for Rosacaeae, 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)

Description	Yr1	Yr2	Yr3	Yr4	Yr5
Salaries	303,631	315,165	326,834	338,969	351,591
Travel	20,000	20,000	20,000	20,000	20,000
Supplies	35,000	35,000	35,000	35,000	35,000
Hardware	40,000		40,000		
Total	**398,631**	**370,165**	**381,834**	**433,969**	**406,591**

- 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)

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

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 ☺

WASHINGTON STATE UNIVERSITY