

Genome Database for Rosaceae

A Resource for Genomics, Genetics and Breeding Research

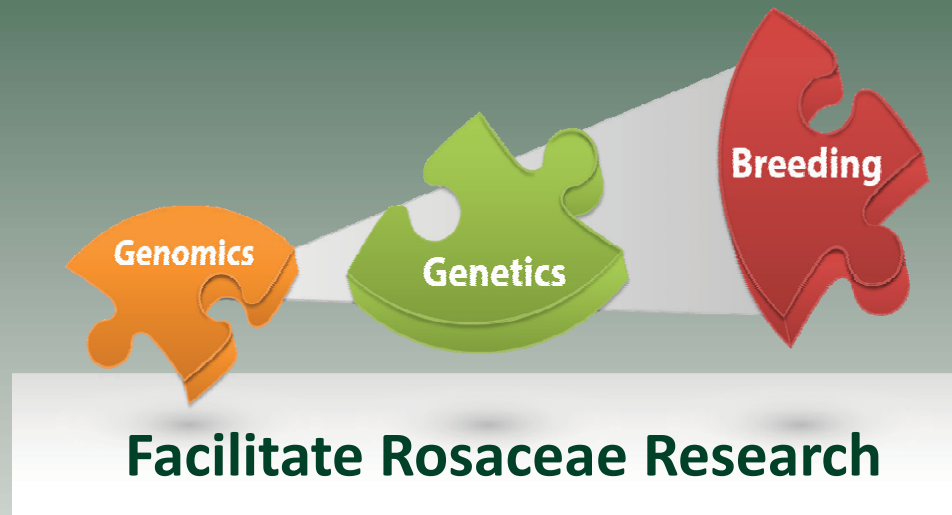
2016 ASHS Annual Conference, Atlanta, GA

Computer Applications in Horticulture

Dorrie Main*, Sook Jung, Chun-Huai Cheng, Taein Lee, Jodi Humann, Stephen Ficklin, Jing Yu, Kristin Scott, Heidi Hough, Cameron Peace, Kate Evans, Mike Kahn, Albert Abbott, Ksenija Gasic, Olmstead, Jim McFerson, Michael Coe

***dorrie@wsu.edu**

GDR Vision Part 1



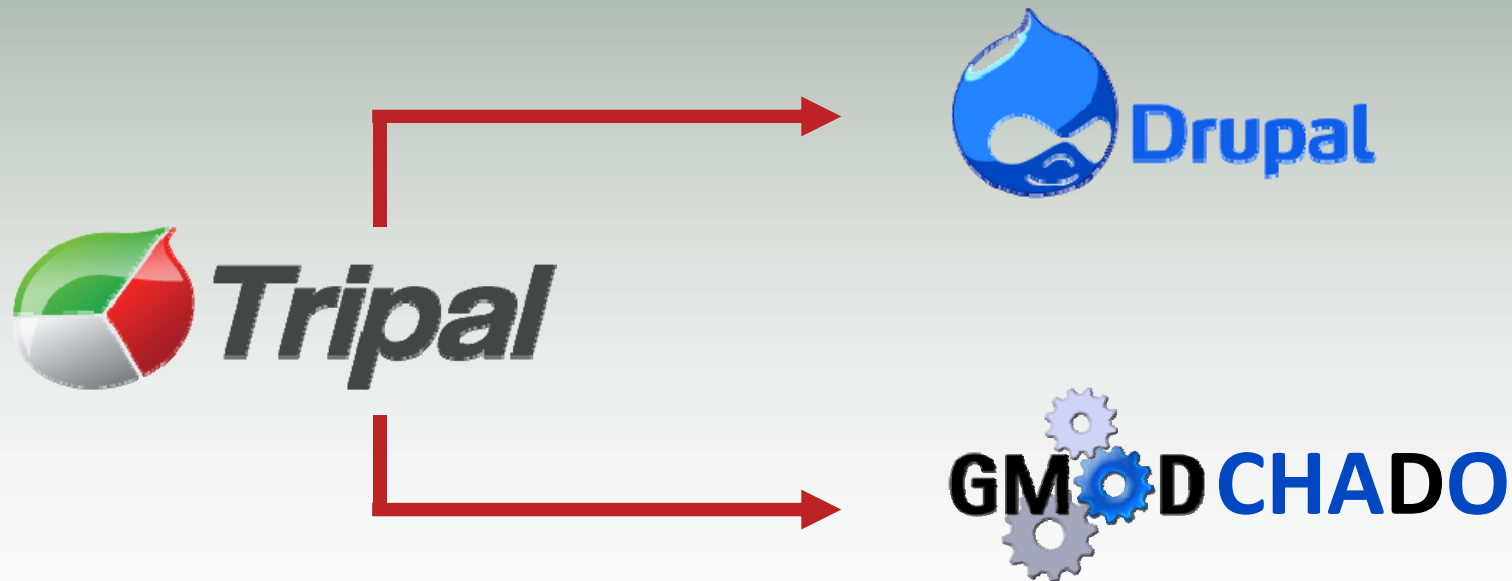
- Discovery
- Translation
- Application

Made possible by

- A highly supportive and engaged community
- Continuous funding (2003-2019) from NSF PGRP, USDA SCRI, USDA NRSP10, Industry and Land Grant Universities
- An excellent and dedicated development and curation team

GDR Vision Part 2

Develop open source software resources to facilitate sharing of tools and data across the bioinformatics community



Over 80 genome databases now using Tripal as a platform



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 is supported by grants from the NSF Plant Genome Program (2003-2008), USDA NIFA Specialty Crop Research Program (2009-2019), USDA NIFA National Research Support Project 10 (2014-2019), and the Washington Tree Fruit Research Commission (2008-2016), Clemson University, University of Florida and Washington State University.

What's new in GDR?

- Genome assembly and annotation of Black raspberry (*Rubus occidentalis* L.) available (04/01/16)
- New QTL, markers and genetic map data available (3/31/16)
- 480K SNP array data for *Malus x domestica* (Bianco et al. 2016) available (3/31/16)
- Species summary information contributed by community members available (3/31/16)
- Genome Assembly of *F. x ananassa* and four wild species v1.0 available (9/10/15)

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

- Mapping between *M. x domestica* genome MDP genes and NCBI genes available (6/1/16)
- Genome assemblies of *F. x ananassa* and four wild species are available in JBrowse (5/27/16)
- GDRcyc databases updated with MetaCyc v19.5 (4/14/16)
- GDR Newsletter April 2016 available (04/13/16)
- NCBI SRA RNASeq and GBS data summary pages available (04/12/16)
- Black raspberry genome sequence and annotation available (04/01/16)
- 8th International Rosaceae Genomics Conference (RGC8) on June 21-24, 2016 in Angers, France
- Announcements archive

Previous GDR Design

GENOME DATABASE FOR ROSACEAE



Resources for Rosaceae Research Discovery and Crop Improvement

News and Events

- Newly designed GDR launched (8/01/16)
- Rubus RNASeq reference transcriptome available (8/01/16)
- Germplasm image search available (8/01/16)
- View synteny between apple, peach, strawberry and pear (7/25/16)

Major Genera Quick Start



Fragaria



Malus



Prunus



Pyrus



Rosa



Rubus

Tools Quick Start

Genomics

[View Genomes](#)
[Find Sequences](#)
[Search Genes](#)
[BLAST Sequences](#)

Genetics


[Browse Maps](#)
[Search Markers](#)
[Find QTLs](#)
[Compare Maps](#)

Breeding

[Search Phenotype](#)
[Search Genotype](#)
[Search Haplotype](#)
[Decision Tools](#)

Developed by Mainlab Bioinformatics at Washington State University | © 2016
Funded by a partnership of USDA, NSF, Industry and US Land Grant Universities



[contact us](#) | [report a problem](#) | [make a suggestion](#) | [follow us](#) 

New GDR Design



GDR

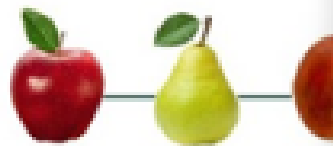
Species

Data ▾

Search ▾

Tools ▾

GENOME



Resources for Rosa

Major Genera



Fragaria (all species)

Fragaria x ananassa (strawberry)

Malus (all species)

Malus x domestica

Prunus (all species)

Prunus armeniaca (apricot)

Prunus avium and cerasus (sweet & tart cherry)

Prunus dulcis (almond)

Prunus persica (peach)

Prunus serotina (black cherry)

Pyrus (all species)

Pyrus communis (European pear)

Rosa (all species)

Rubus (all species)

Rubus occidentalis

others

Prunus all species

Data

- [Overview](#)
- [Description](#)
- [Genes](#)
- [Genetic Maps](#)
- [Genomes](#)
- [Markers](#)
- [Publications](#)
- [SNP array](#)
- [Trait Loci](#)
- [Transcripts](#)

Tools

- [BLAST](#)
- [BLAST Batch](#)
- [CMap](#)
- [GBrowse](#)
- [JBrowse](#)
- [PathwayCyc](#)
- [Sequence Retrieval](#)
- [Synteny Viewer](#)

Overview

Genus	<i>Prunus</i>
Species	<i>all species</i>
Common Name	Prunus
Abbreviation	<i>Prunus (all species)</i>
Ploidy	Diploid
Chromosome Number	2n=2x=16
Genome Size	varies by species
Available Markers	103,342
Available Maps	86
Available QTLs	1,235
Available MTLs	56
Available Phenotype Data	98,747 measurements for 255 descriptors
Available Genotype Data	8,044 SNPs; 558 SSRs
Genome Assemblies	2





Data Submission

Please contact us to arrange transfer of data to GDR from accepted peer reviewed work. Below are Excel template files we recommend you contact us before filling in the template so that we can work closely with you to ensure this process is successful. For data such as transcriptome data or genome data, please contact us and we will work with you to get the data accessible in GDR.

- [Gene data submission form](#) (refer to [Gene Naming Guideline](#) and the list of gene class symbols in Rosaceae)
- [Diversity data template](#) (genotype data only)
- [Map \(genetic\) data template](#) (for quick display in CMap, use CMap template)
- [Map \(QTL\) data template](#)
- [Genotypic/phenotypic data from breeding or cultivar evaluation projects](#)
- [Breeding data templates](#) (with examples from RosBreed Project)
- [List of files for genome data submission](#)

Data Overview

+ Genes : 241590 genes and 331639 mRNAs

+ Genomes : 10 genome assemblies

+ Germplasm : 17386 germplasm

+ Maps : 167 maps

+ Markers : 2345749 markers

+ Phenotypes : 845467 phenotypic measurements

+ Publications : 6820 publications

+ QTL : 3189 QTL and 149 MTL for 414 agronomic traits

+ SNP Genotype : 11748 markers 3122 germplasm

+ SSR Genotype : 5013 markers 17839 genotypic measurements

+ Species : 1899 species

+ Transcripts

- Markers : 2345749 markers

+ Fragaria : 96910

+ Malus : 2141345

+ Prunus : 103342

+ Pyrus : 1592

+ Rosa : 2560

+ Rubus : 7830



GDR

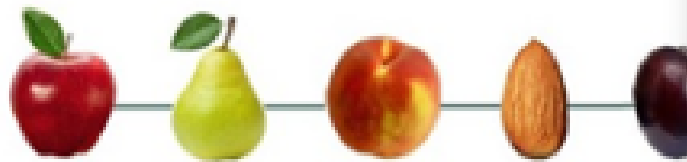
Species ▾

Data ▾

Search ▾

Tools ▾

GENOME DATABASE



Resources for Rosaceae Research

Data Contributors

Data Overview

Data Submission

Disclaimer

Download Data

GDR Marker Types

Publication Datasets

TGG Data Submission

NCBI RNASeq and DNA Datasets



Fragaria vesca - NCBI SRA Datasets

RNA Datasets (as of March 2016)

Species	Study Accession	Cultivar	Reference
<i>Fragaria vesca</i>	ERP004230	YellowWonder, Hawaii	Jambagi and Dunwell. 2015
<i>Fragaria vesca</i>	SRP047520	Yellow Wonder 5AF7	Xia et al. 2015
<i>Fragaria vesca</i>	SRP035308	5AF7	Hollender et al. 2014
<i>Fragaria vesca</i>	SRP019247	Yellow Wonder 5AF7'	Xia et al. 2013
<i>Fragaria vesca</i>	SRP018410	Yellow Wonder 5AF7	Darwish et al. 2015

DNA Datasets (as of March 2016)

Species	Study Accession	Cultivar	Reference
<i>Fragaria vesca</i> subsp. <i>bracteata</i>	SRP022950	<i>bracteata</i>	Tennessen et al. 2013
<i>Fragaria vesca</i> f. <i>alba</i>	SRP044273	Yellow_Wonder CFRA480	Bassil et al. 2015
<i>Fragaria vesca</i> subsp. <i>americana</i>		Pawtuckaway CFRA1948	
<i>Fragaria vesca</i> f. <i>semperflorens</i>		Baron Solemacher CFRA47	



GDR

Species ▾

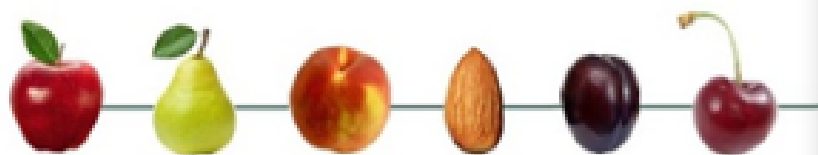
Data ▾

Search ▾

Tools ▾

General

GENOME DATABASE



Resources for Rosaceae Research Discoveries

Search Genes and Transcripts

Search Genotype

Search Germplasm Images

Search Haplotype Block

Search Maps

Search Markers

Search Publications

Search QTL

Search Sequences



GDR

Species ▾

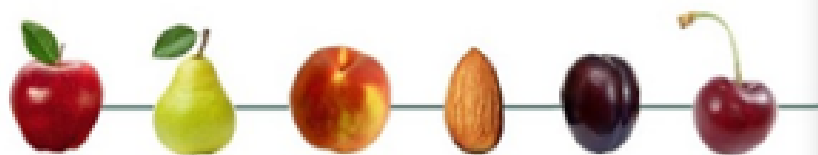
Data ▾

Search ▾

Tools ▾

General

GENOME DATABASE



Resources for Rosaceae Research Discovery

Search Genes and Transcripts

Search Genotype

Search Germplasm Images

Search Haplotype Block

Search Maps

Search Markers

Search Publications

Search QTL

Search Sequences



Search Genes and Transcripts

Search genes and transcripts by species, dataset, genome location, name and/or keyword. For keyword, enter any protein name of h term, or InterPro term. [Short video tutorial](#) | [Text tutorial](#) | [Email us with problems and suggestions](#)

Genus

Species

Dataset ?

- Any
- Curated Genes**
 - GDR Gene Database
 - NCBI Rosaceae gene and mRNA sequences
- Predicted Genes**

Genome Location

between

and

Gene/Transcript Name

Choose File

No file chosen

Keyword

(eg. polygalacturonase, resistance, EC:1.4.1.3, cell cycle,

Search

Reset

Description of Sequence Datasets in GDR

- **Curated Genes:**

- GDR Gene Database: A single non-redundant list of Rosaceae genes with gene symbols. Majority of some user-contributed data. Multiple gene sequences, from different sources, are associated with the
- NCBI Rosaceae gene and mRNA sequences: All gene and mRNA sequences parsed out from NCBI species of *Prunus*, *Malus*, *Fragaria*, *Pyrus* and *Rubus* are anchored to the *P. persica* genome v1.0, *M. communis* genome v1.0 and *Rubus occidentalis* genome v1.0, respectively, using blat with criteria of

- **Predicted Genes:** Genes and mRNAs from whole genome assemblies. Additional annotation of these predicted genes with homology to genes of closely related or plant model species and assignment of InterPro protein domain database (KEGG) pathway and ortholog terms.

- **RefTrans:** RefTrans combines published RNA-Seq and EST data sets to create a reference transcriptome identified by homology to known proteins.

- **Unigene:** EST contigs for the Rosaceae family and each genus, constructed from the publicly available Rosaceae data. The construction for the GDR occurs in four steps: (i) sequence filtering and trimming to obtain high-quality sequences, (ii) removing sequence redundancy, (iii) building unigene sets from the combined contigs and singlets and (iv) sequence annotation.

- **Other**

- Towards an improved apple reference transcriptome using RNA-seq (Bai et al. 2014):
- RosCos unigene sets
- RosaR80 v1.0

Dataset ?

Predicted Genes

- Fragaria vesca Whole Genome v1.0 (build 8) Assembly & Annotation
- Malus x domestica Whole Genome v1.0 Assembly & Annotation
- Malus x domestica Whole Genome v1.0p Assembly & Annotation
- Prunus persica Whole Genome Assembly v2.0 & Annotation v2.1 (v2.0.a1)**

Genome Location Any **between** **and**

Gene/Transcript Name contains No file chosen

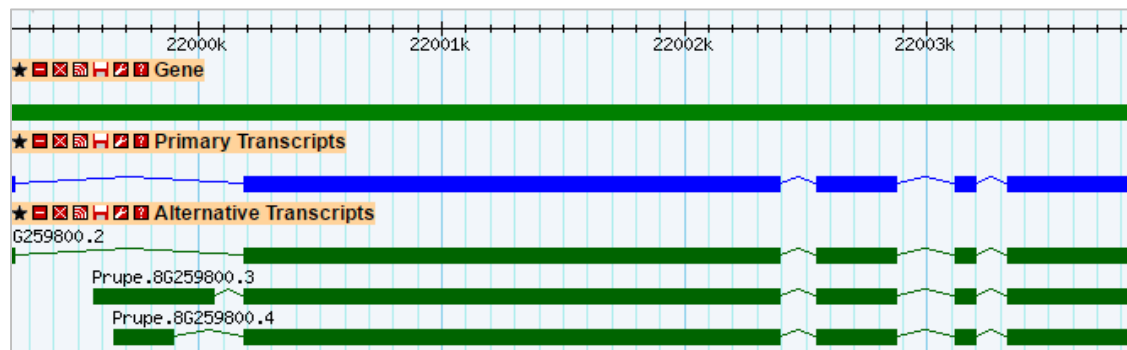
Keyword contains **resistance** (eg. polygalacturonase, resistance, EC:1.4.1.3,

1488 records were returned

[Download Table](#) | [Fasta](#)

#	Name	Organism	Type	Source	Location
1	Prupe.8G259800.4	Prunus persica	mRNA	Prunus persica Whole Genome Assembly v2.0 & Annotation v2.1 (v2.0.a1)	Pp08:21999649..22007026
2	Prupe.6G311600.1	Prunus persica	mRNA	Prunus persica Whole Genome Assembly v2.0 & Annotation v2.1 (v2.0.a1)	Pp06:27915405..27915821
3	Prupe.3G123800.1	Prunus persica	mRNA	Prunus persica Whole Genome Assembly v2.0 & Annotation v2.1 (v2.0.a1)	Pp03:11104095..11110214

#	Name
1	Prupe.8G259800.4
2	Prupe.6G311600.1
3	Prupe.3G123800.1



Prupe.8G259800.4, Prupe.8G259800.4_v2.0.a1 (mRNA) Prunus persica

Overview

Alignments

Analyses

Annotated Terms

Homology

InterPro

Relationships

Sequences

InterPro

Analysis Name: InterProScan analysis of Peach Genome v2.0.a1

Date Performed: 2015-01-27

IPR Term	IPR Description	Source	Source Term	Source Description	Alignment
IPR003593	AAA+ ATPase domain	SMART	SM00382	AAA	coord: 670..851 e-value: 1.7E-12 score: 57.5 coord: 1292..1476 e-value: 1.3E-7 score: 41.3
None	No IPR available	PANTHER	PTHR24223	FAMILY NOT NAMED	coord: 10..1505
IPR011527	ABC transporter type 1, transmembrane domain	SUPERFAMILY	90123	ABC transporter transmembrane region	coord: 946..1246
IPR003439	ABC transporter-like	PROSITE	PS50893	ABC_TRANSPORTER_2	coord: 643..866 score: 23.716
IPR027417	P-loop containing nucleoside triphosphate hydrolase	SUPERFAMILY	52540	P-loop containing nucleoside triphosphate hydrolases	coord: 640..865

From the gene details page, find out all the associated genome location, functional annotation, sequence, alignments, analyses etc for that gene.



GDR

Species ▾

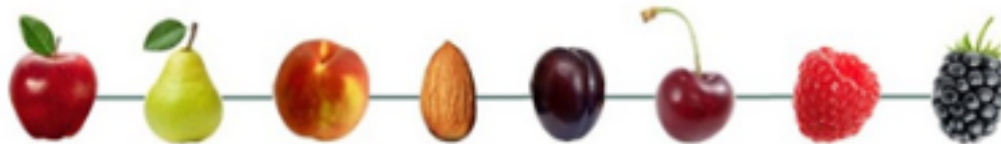
Data ▾

Search ▾

Tools ▾

General ▾

GENOME DATABASE FOR RO



Resources for Rosaceae Research Discovery and Cr

BLAST

Batch BLAST

Breeders Toolbox

CMap

GBrowse

GDRcyc

JBrowse

Primer3

Sequence Retrieval

Synteny Viewer

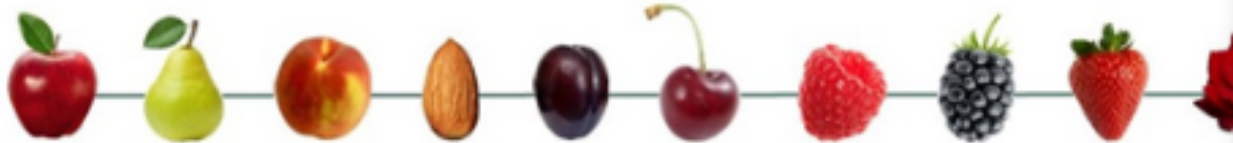
New synteny analysis between
apple, peach, strawberry and pear



GDR

[Species](#) ▾[Data](#) ▾[Search](#) ▾[Tools](#) ▾[General](#) ▾[Help](#) ▾

GENOME DATABASE FOR ROSACEAE



Resources for Rosaceae Research Discovery and Crop Improvement

[About](#)[Newsletter](#)[News Archives](#)[People](#)[Presentations](#)[Steering Committee](#)[Usage](#)[Work Completed](#)[Work in Progress](#)

In the last year GDR was accessed by 19,051 users from 145 countries who viewed 222,671 pages



GENOME DATABASE FOR ROSACEAE

[Contact](#)[GDR FAQ](#)[GDR Tutorials](#)[Gene Naming Guideline](#)[Genome Naming Guideline](#)

GDR Tutorials

Tutorial Topic	Video	Text
Site Overview	Watch	Read
Species Page	Watch	Read
Searches:		
Genes and Transcripts	Watch	Read
Genotypes	Watch	SSR SNP
Germplasm Images	Watch	Read
Haplotype Blocks	Watch	Read
Markers	Watch	Marker Nearby Marker
Publications	Watch	Read
QTLs	Watch	Read
Sequences	Watch	Read
Tools:		
Breeders Tool Box	Watch	Read

[Conferences](#)[Employment](#)[Mailing Lists](#)[RosIGI](#)[USRosExec](#)[Community Projects](#)

Mailing Lists

GDR Communication Mailing List

- [Subscribe to the GDR Mailing List](#)
- [View/Search the GDR Mailing List Archives](#)
- [Subscribe to the Rosaceae QTL Mailing List](#)
- [View/Search the Rosaceae QTL Mailing List Archives](#)
- [Subscribe to the Rosaceae Genetic Diversity Mailing List](#)
- [View/Search the Rosaceae Genetic Diversity Mailing List Archives](#)

International Rosaceae Genomics Community Communication

- [Subscribe to the International Rosaceae Genomics Community Mailing List](#)
- [View/Search the International Rosaceae Genomics Community Mailing List Archives](#)

US Rosaceae Genomics Community Communication

- [Subscribe to the US Rosaceae Genomics Community Mailing List](#)
- [View/Search the US Rosaceae Genomics Community Mailing List Archives](#)

US Rosaceae Genomics, Genetics, and Breeding Executive Committee (US RosEXEC) Communication

- [View/Search the USRosEXEC Mailing List Archives](#)
- [View/Search the USRosEXEC Research Strategies and Priorities Mailing List Archives](#)

Major Work in Progress

- Designing and implementing a comprehensive breeding information management system (TriPalBIMS)
 - See the GDR website for more details
- Evaluating the FieldBook App for collection of phenotype data (Jesse Poland Program).
 - See Ksenija Gasic talk, Wed 2.00 PM at the Fruit Breeding Oral Session
- Creating cross database querying and analysis capability between tree databases – adding environmental data



Will start to offer Webinar training this fall

Acknowledgements

- USDA SCRI, USDA NRSP10, NSF Plant Genome Program, USDA-ARS, SAAEDS, Washington Tree Fruit Research Commission, Cotton Incorporated, USA Dry Pea and Lentil Commission, Northern Pulse Growers, Mars Inc.
- Mainlab bioinformatics team and various projects CoPIs/PIs
- Rosaceae, Citrus, Blueberry, Legume, Cacao, Cotton and Bioinformatics Communities
- US Land Grant University researchers and extension agents