

# Genome Database for Rosaceae

## Updates and New Direction

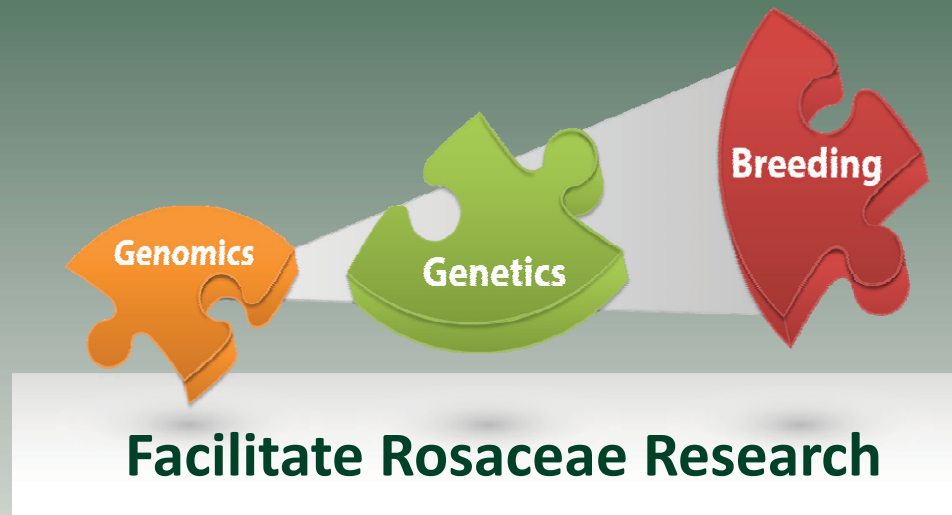
### Fruit and Nut Workshop

Plant and Animal Genome Conference 2016

Dorrie Main\*, Sook Jung, Chun-Huai Cheng, Taein Lee, Stephen Ficklin, Jing Yu, Jodi Humann, Kristin Scott, Michael Byrd, Anna Blenda, Cameron Peace, Kate Evans, Mike Kahn, Albert Abbott, Ksenija Gasic, Mercy Olmstead, Michael Coe

\*[dorrie@wsu.edu](mailto:dorrie@wsu.edu)

# GDR Vision Part 1

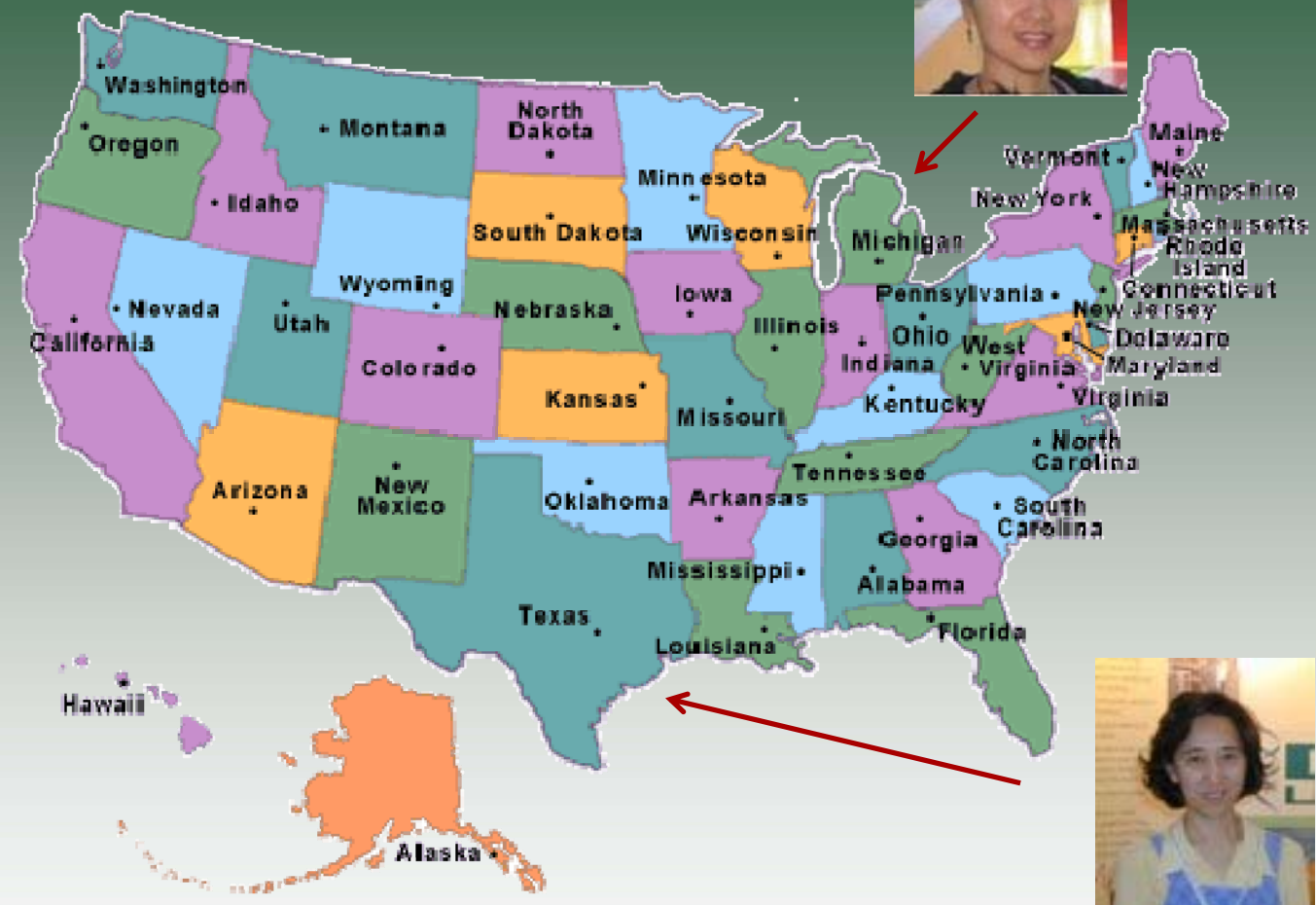


- Discovery
- Translation
- Application

## Made possible by

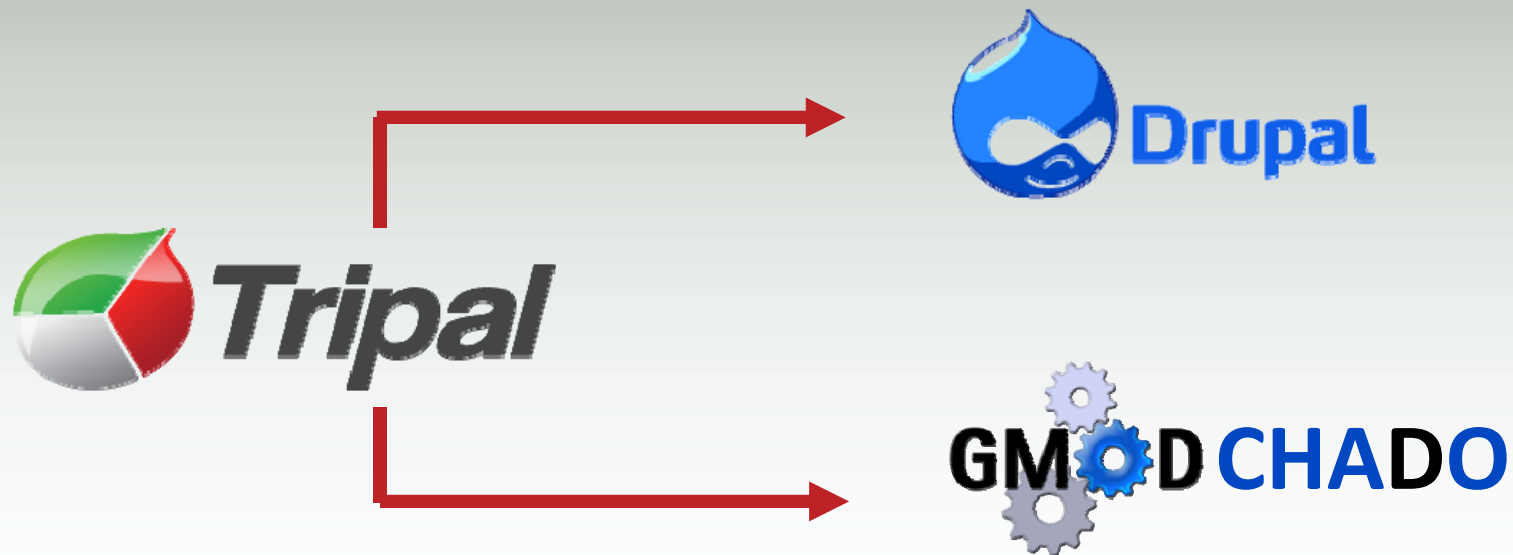
- A highly supportive and engaged community
- Continuous funding (2003-2019) from NSF PGRP, USDA SCRI, Industry and Land Grant Universities
- An excellent development and curation team

# The Team

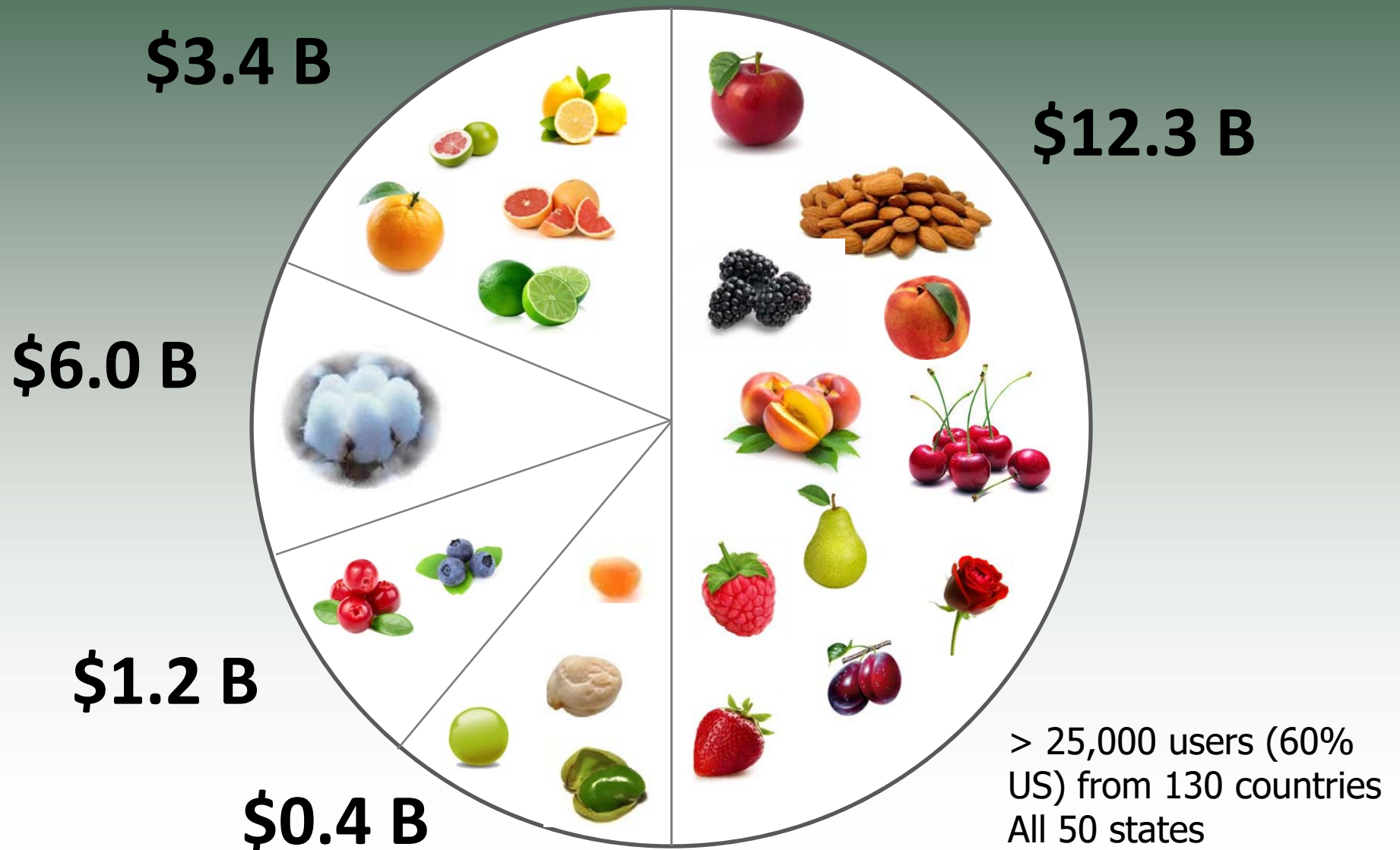


# GDR Vision Part 2

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



# WSU Crop GGB Databases



# New Data in GDR in 2015

Data Type	2015	2016	Increase
Genes	233,191	330,235	97,044
Genomes	6	11	5
Genetic maps	123	160	37
Genotypes	28,298	15,513,927	15,485,631
Germplasm	13,500	16,345	2,845
Markers	2,193,827	2,311,771	117,944
Phenotypes	734,343	783,639	49,296
Publications	6,319	6,703	384
QTLs	2255	2318	63

# New Genome Sequence Data

- *Prunus persica* genome V2.0.a1 (IPGC 2015)
- *Malus x domestica* genome V3.0.a1 (Velasco et al., 2015)
- *Fragaria vesca* V1.1.a2 genome (Darwish et al., 2015)
- *Fragaria vesca* V2.0.a1 genome (Tennessen et al., 2014)
- *Fragaria x ananassa* v1.0 genome (Hirakawa et al., 2014)
- *Fragaria iinumae* v1.0 (Hirakawa et al., 2014)
- *Fragaria nipponica* v1.0 (Hirakawa et al., 2014)
- *Fragaria nubicola* v1.0 (Hirakawa et al., 2014)
- *Fragaria orientalis* v1.0 (Hirakawa et al., 2014)

# New Functionality in 2015

- Search interface for haplotype data
- New Interface to access SNP data
- Species pages reorganized
- Efforts towards data standardization
  - Standard gene nomenclature in the Rosaceae
  - QTL metadata
  - Rosaceae Trait Ontology

Tree Genetics & Genomes (2015) 11:108  
DOI 10.1007/s11295-015-0931-5



OPINION PAPER

## **A standard nomenclature for gene designation in the Rosaceae**

Sook Jung<sup>1</sup> · Carole Bassett<sup>2</sup> · Douglas G. Bielenberg<sup>3</sup> · Chun-Huai Cheng<sup>1</sup> ·  
Chris Dardick<sup>2</sup> · Dorrie Main<sup>1</sup> · Lee Meisel<sup>4</sup> · Janet Slovin<sup>5</sup> · Michela Troglio<sup>6</sup> ·  
Robert J. Schaffer<sup>7,8</sup>



# New Species Pages

## Malus x domestica

### Overview

Genus	Malus
Species	<i>x domestica</i>
Abbreviation	M. x domestica
Scientific Name	Malus x domestica
Common Name	Apple
Geographic Origin	cultivated
Haploid Chromosome Number	17
Ploidy	2x, 3x
Growth Habit	tree
Propagation Method	grafting for clones
Usage	fruit
Resistance to biotic stress	Dessert, cider varieties
Germplasm	[view all 5445]
Sequence	[view all 3703185]

### Resources

#### Data

- Overview
- Publications
- Genetic Maps
- Genes
- Markers
- Sequences
- Trait Loci
- Whole Genomes
- SNP arrays
- Germplasm

#### Tools

- GBrowse
- JBrowse
- GBrowse\_Syn
- CMap
- AppleCyc
- NCBI Blast
- Batch Blast
- Sequence Retrieval Tool
- GO Viewer
- KEGG Pathway Viewer

# Gene Naming Guideline Page

## Gene Naming Guideline

[View](#)[Moderate](#)[Edit](#)[Devel](#)[Open Access Publication](#) | [Detailed Guideline](#) | [Gene Class Symbol](#) | [Species Prefix](#)

## Gene Naming Guidelines

Genes in GDR are composed of two types:

1. Genes with gene symbol, assigned by individual investigators, usually with known or predicted function or phenotype.
  - These genes with gene symbols in GDR are obtained by parsing the NCBI nr database and submission from individual investigators.
  - Prior to naming your genes, please refer to the list of [gene class symbols in Rosaceae](#) and gene naming guidelines below and **submit the gene/gene class data**.
2. Genes predicted from the whole genome assembly

### 1. Naming Guideline for Genes with known or predicted function or phenotype

**(for gene data submission and gene nomenclature in manuscript)**

The guideline has been put together by [Rosaceae Gene Name Standardization Subcommittee of RosEXEC/RosIGI](#).

Gene Symbol is composed of:

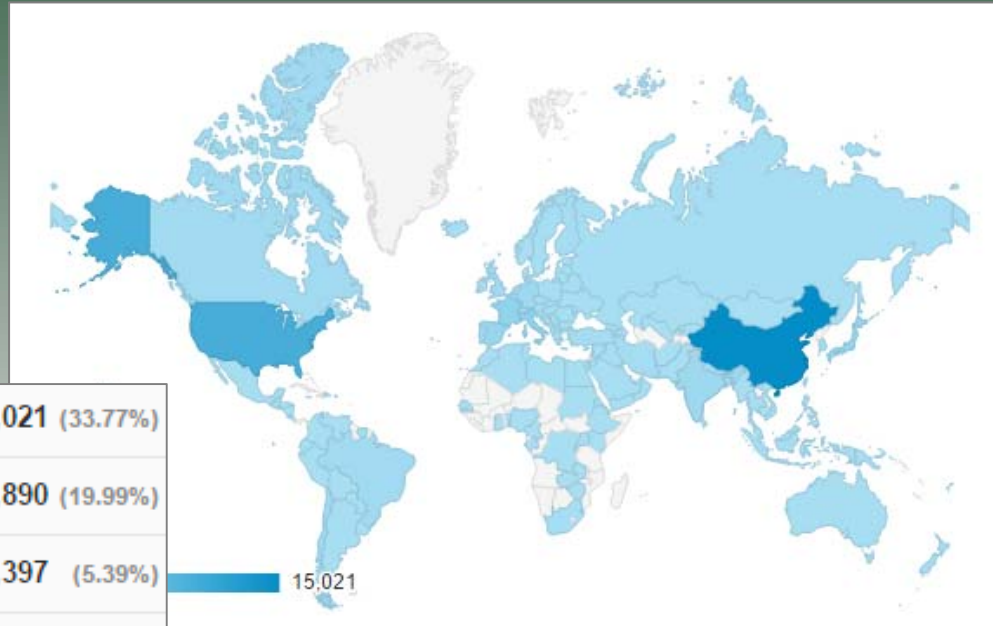
[species prefix][3-letter code class symbol].[numeric suffix for a gene][-\_][numeric suffix for an allele or a splice variant]

Where

- [species prefix]: *publication purposes* - three letter prefix for major species and 5 letter prefix for others

# Usage of GDR in 2015

- 18,244 visitors
- 44,479 visits
- 199,936 pages
- 144 countries



1.	 China	15,021 (33.77%)
2.	 United States	8,890 (19.99%)
3.	 Italy	2,397 (5.39%)
4.	 France	2,152 (4.84%)
5.	 Japan	1,803 (4.05%)
6.	 Spain	1,552 (3.49%)
7.	 India	1,065 (2.39%)
8.	 Germany	914 (2.05%)
9.	 Brazil	827 (1.86%)
10.	 Canada	735 (1.65%)

# Work in Progress

- Converting GDR from Tripal 1 to Tripal 2 and Drupal 7)
- Designing and implementing a comprehensive breeding information management system (TripalBIMS)
- Evaluating the FieldBook App for collection of phenotype data (Jesse Poland Program)  
Rosaceae breeders may request provision of a Samsung tablet with the App installed for testing in their program
- Adding more GGB data and working on data standards and ontologies



# FieldBook App Evaluation

“Very easy to use. Fully customizable right in the field within app. Allows simultaneous collection of the same data with very little training needed, e.g. anyone can assist in data collection since the traits can be designed to be exactly the same and avoid improvisation. Data are exportable as database and excel files and fit into statistical programs. Picture and Audio “traits” are very useful.”

Dr. Ksenija Gasic  
Peach Breeder  
Clemson University

# Some Requests (Please)

- Cite us (Jing et al., 2014 – on website)
- Join our mailing list to receive biannual newsletters and community communication
- Please tell us if you find a problem with the site, have new functionality you would like
- Come see the GDR demo on Tuesday at 12.25 PM in the California room
- Come see what we are planning for the BIMS on Sunday at 5.20 PM in the California room
- Come participate in the RosEXEC/RosIGI open meeting Sunday 8-10,10-12 at theStratford Room

# 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, Legume, Cotton and Bioinformatics Communities
- USDA NIFA SCRI, USDA DOE, NSF Plant Genome Program, USDA-ARS, SAAEDS, Mars Inc, Washington Tree Fruit Research Commission, Cotton Incorporated, USA Dry Pea and Lentil Commission, Northern Pulse Growers,
- US Land Grant University researchers and extension agents