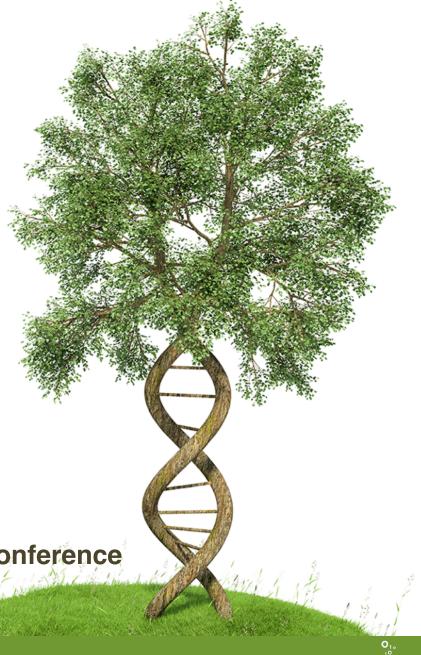
## TreeGenes: Past, Present, and Future Development

Jill Wegrzyn

**Plant and Animal Genome Conference** 

**January 14th 2019** 



# Dendrome/TreeGenes was the original three USDA-ARS databases



## So much time...so many technologies!



## So much time...so many data types!

Colleague Directory Literature Database Markers (SSR, AFLP) Genetic
Maps,
Community
Markers
(SSR, AFLP,
RAPD)
ESTs

Genetic Maps, Community Markers, Community ESTs cDNAs Genetic Maps, Markers, Community ESTs cDNAs Phenotypes Genetic
Maps,
Markers,
Community
ESTs
cDNAs
Phenotypes
SNPs
Genomes

Genetic
Maps,
Markers,
Community
ESTs
cDNAs
Phenotypes
SNPs
Genomes
Environmental

1991: Dendrome (standalone software)

1995: Database Online (ACEDB)

Or Date

2001: Oracle Database **2005:** MySQL

2009: PostGreSQ L **2015:**Tripal
Framewor
k

dendrome.ucdavis.edu

Treegenes. ucdavis.ed

u

Treegenesd b.org

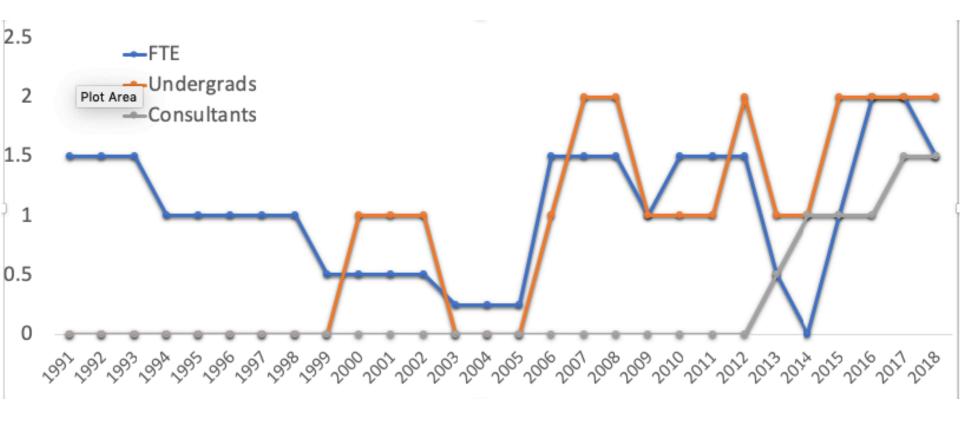


## TreeGenes Audience

- Academic Research Labs
- Non-profit research institutions
- US Forest Service
- Industry
  - Timber companies
- JGI/DOE
  - Data exchange
- Genbank
  - Data exchange



## **TreeGenes Staffing**





## BT Era (Before Tripal)



## **Primary Structure**

#### Web

- Custom Frontends
- Search Fields/ Summary Pages
- Advanced Interface (DiversiTree)

#### **Database**

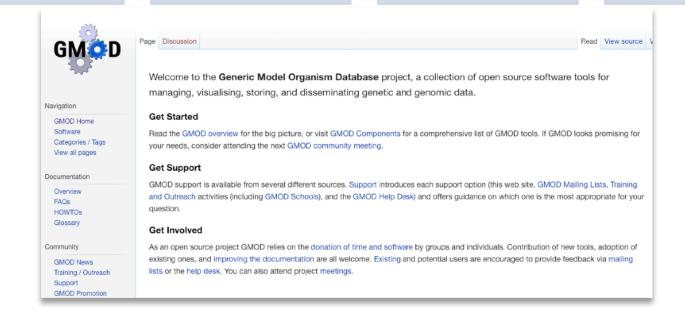
- Custom Schema
- GMOD components:
  - Genetic Maps (CMAP)
  - Genomes (Gbrowse/ Jbrowse)

#### Scope

- Genetic Maps
- Sequence
- Phenotypes
- Genotypes
- Conifers/Forest Trees

#### **Funding**

- No dedicated funding
  - Forest Service
  - USDA CAP
  - NSF Projects
  - Inbetween? plugging along (i.e. Maintenance)





## AT Era (After Tripal)

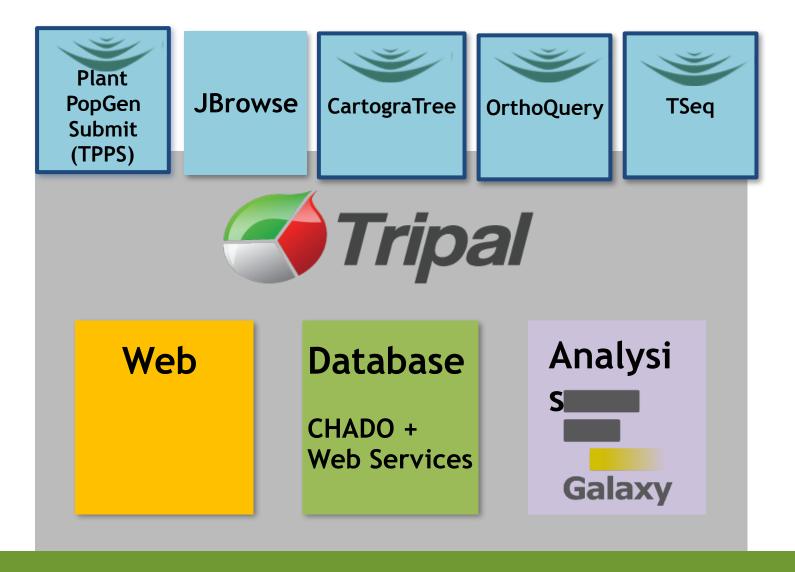


## **Current Funding**

- NSF PGRP (2015-2019)
  - Conversion to Tripal
  - Ontologies
  - CartograTree
- NSF DIBBS (2014-2018)
  - Tripal Galaxy
  - Module development
  - CartograTree
- USDA AFRI PineSNP
  - Partial project support



## Tripal v3 Framework in TreeGenes



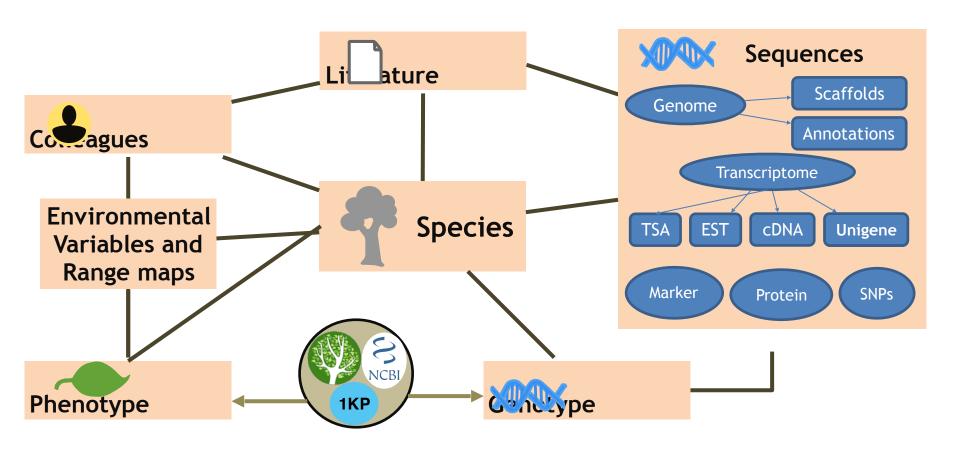
## TreeGenes Database: Species

treegenesdb.org



- 1,850 species from 112 genera
  - At least one genetic artifact from each species
- Full genome sequence: 35 species
- Transcriptome/Expression resources: 6,920,817 sequences from 322 species
- 101 genetic maps from 37 species
- Population studies
  - Georeferenced trees
  - Extensive genotypic (GBS and array) and phenotypic data

## Relational Data in TreeGenes





## TreeGenes Database: Users



treegenesdb.org 3,100 unique visitors from 116 countries .000 750 500 700 250 Unique Web Visitors to TreeGenes (by month), July-December 2018 200 Sep-18

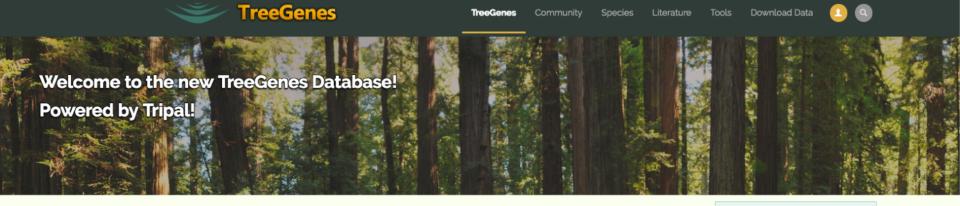
Oct-18

Nov-18

Dec

Jul-18

Aug-18



#### **EXPLORE TREEGENES**

**JBrowse** 



**Browse Genomes** Browse genomes with



CartograTree

CartograTree is a map-based web app.



Download Data

Download data from our FTP



Submit

Submit your data to TreeGenes.

#### MEETINGS

PAG XXVII January 12 to 16 2019 San Diego, USA

2019 SFTIC June 3 to 6 2019 Lexington, USA

· View more meetings



Sequence Search Search sequences with DIAMOND



Species Find your species of interest

#### FEATURED PROJECTS



B<sub>4</sub>est



Oak Genome Project



CoAdapTree



10kp

#### LATEST LITERATURE

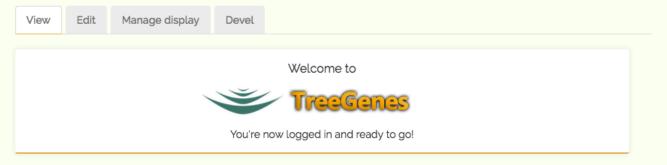
Genetic variation and signatures of natural selection in populations of European beech (Fagus sylvatica L.) along precipitation gradients. (2018) Tree Genetics & Genomes

Community Resources: Colleague directory, mailing list, meetings, employment listings, and literature











Add/Edit Profile



Add/Edit Workshop/Conference



Add/Edit Employment Listings

## Edit Profile Log out

Add	new job	posting	

#### Add new meeting posting

#### MEETINGS

My account

#### PAG XXVII

January 12 to 16 2019 San Diego,USA

#### 2019 SFTIC

June 3 to 6 2019 Lexington,USA

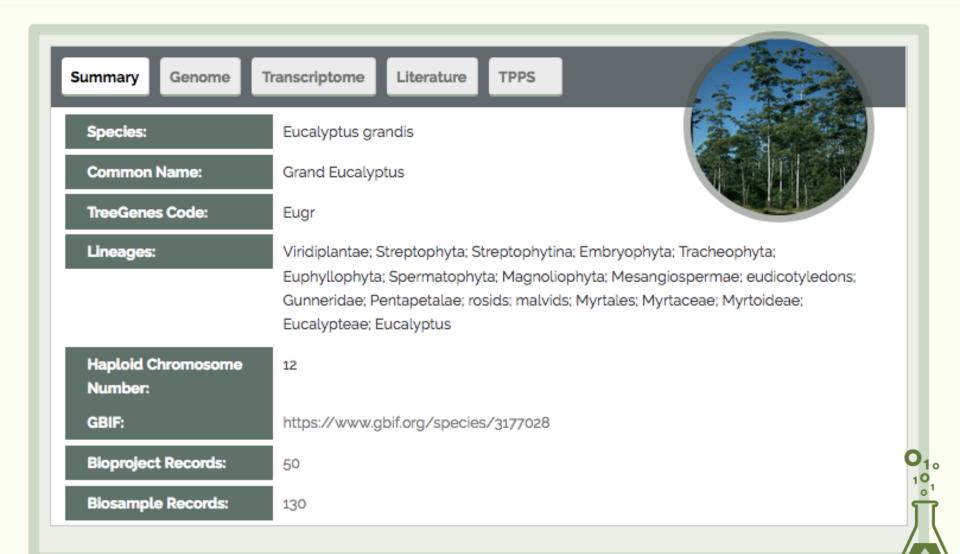
· View more meetings





Free account to: edit profile, add listings, submit data!

### **Eucalyptus grandis**



### TreeGenes: Genome Access



#### Betula nana

Version: v1.0 Organism Detail | JBrowse

- T Annotation (GFF3)



#### Carica papaya

Version: v0.4 Organism Detail | JBrowse

- T Annotation (GFF3)
- Froteins (FASTA)



#### Elaeis guineensis

Version: v10 Organism Detail | JBrowse

- # Annotation (GFF3)
- Proteins (FASTA)



#### Eucalyptus camaldulensis

Version: v10 Organism Detail | JBrowse

- Annotation (GFF3)
- Proteins (FASTA)



#### Eucalyptus grandis

Version: v1.0 Organism Detail | JBrowse

- Annotation (GFF3)
- F Proteins (FASTA)



#### Fraxinus excelsior

Version: v0.5 Organism Detail | JBrowse

- Annotation (GFF3)
- Froteins (FASTA)



#### Ginkgo biloba

Version: v10 Organism Detail | JBrowse

- Genome FASTA
- Annotation (GFF3)
- Proteins (FASTA)



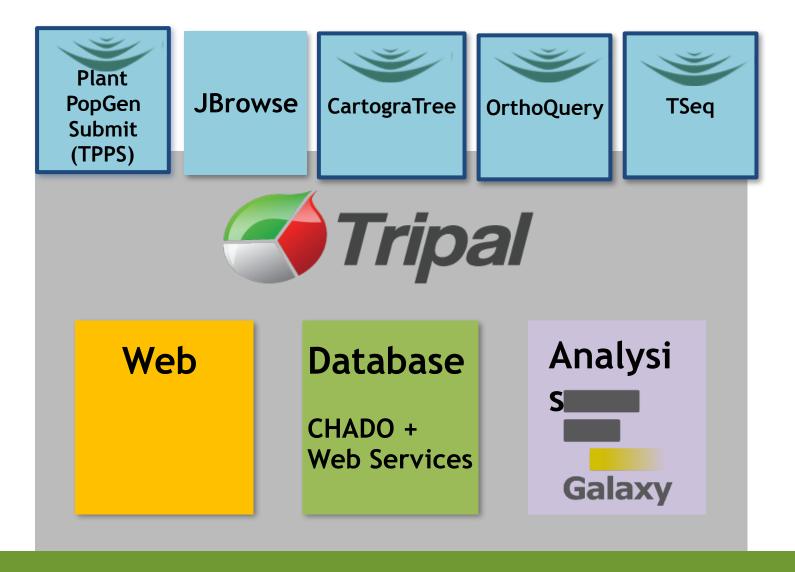
#### Gnetum montanum

Version: v10 Organism Detail | JBrowse

- Genome FASTA
- Annotation (GFF3)
- Proteins (FASTA)



## Tripal v3 Framework in TreeGenes



## **TSeq**

### Diamond search

BLASTp/x

Faster performance

### **BLAST**

**BLASTn** 

Analysis on remote (non-web) server





TSeq: Tripal Sequence Similarity Search

### ▼ SIMILARITY SEARCH OVERVIEW Sequence similarity search is supported in TreeGenes against genes, (BLASTN) while protein or translated protein searches are executed w of BLAST at a similar sensitivity. Both packages accept similar input pa FASTA formatted sequences for searching. You can select from a list of provided in FASTA format and formatted for you. SEQUENCE QUERY TYPE Protein Nucleotide (coding or whole genome) Database Search Type | BLASTx (Translated nucleotide guery versus PROVIDE/LOAD QUERY SEQUENCE ccgaagtctttgctaatgattttccggtcatcccatgctttctcaaggatcctttcatgcattatgt ttttgatgtatggtctcacgcggcaagtatccgtataaaccaattatccaagcattccctcgati Choose File No file chosen **UPLOAD** TARGET/DATABASE SELECTION Protein ---Protein (UniGene) --

Protein (TSA) --

## TreeGenes UniGenes to OrthoQuery

### **TreeGenes: OrthoQuery Module**



### 3 Primary Uses Cases for the OrthoQuery Module



Input: Small Set of Protein Sequences

Question: Which gene family does it belong to?

#### Process:

- Map existing orthogroups to the protein sequence(s)
- Provide visualization
- Assign functional annotation



Question: Orthologous genes, expansion/contraction of gene families

#### Process:

- Construct Proteome
- Discover new orthogroups with existing tools
- Provide visualization & statistical results

Input: Proteome

Question: Orthologous genes, expansion/contraction of gene families

#### Process:

- Discover new orthogroups with existing tools
- Provide visualization & present statistical results

## **Tripal Plant PopGen Submit (TPPS)**

Population Study

- Publication
- Species

Study Design

- Landscape
- Common Garden
  - Greenhouse
  - Growth Chamber
- Breeding (Plot)

Phenotype, Genotype, Environmen t

Georeference d

Raw Data



## Minimal Information About a Plant Phenotyping Experiment (MIAPPE)



### **MIAPPE**

### Minimum Information About a Plant Phenotyping E

MIAPPE is an open, community driven project to harmonize data from plant phenotypin both a conceptual checklist of metadata required to adequately describe a plant pheno validate, store and disseminate MIAPPE-compliant data. We welcome contributions fro phenotyping data.

Development of MIAPPE is an open process, so if you would like to do more than just comment, and to participate in the ongoing development, please let us know via miappe-feedback@ebi.ac.uk.

### MIAPPE v1.1

MIAPPE version 1.1 was officially released on 10th January 2019, following consideration of responses to two requests for comments. Major developments over v1.0 include:

## Ontology Integration:

- Plant Ontology (PO)
- Crop Ontology (CO)
- Trait Ontology (TO)
- Environmental comprises Ontology (ENVO)

## **TreeGenes Ontology Integration**



#### TREEGENES PHENOTYPE AND STRUCTURE ONTOLOGY DIRECTORY

TreeGenes is actively curating submitted traits and structures and assigning these traits to the following ontologies: Plant Ontology (structure), Trait Ontology (traits), Crop Ontology (woody trait ontology), and PATO (descriptive ontology). When possible, assignments are made to existing terms that are curated by the <u>Planteome</u> and <u>Crop Ontology</u> projects. For terms that are not yet assigned to Assignment of individual measurements and structures to an ontological framework provides improved recovery of data downstream for all users. This also enables comprehensive analytical pipelines that can efficiently integrate across a multitude of independent studies. To submit your phenotype and/or genotype study, please use the <u>Tripal Plant Population Genetics</u>

Submit Pipeline.

We currently have 306 unique phenotypes described on 22 unique structures representing a total of 724265 phenotypic measures from 41451 samples representing 18 species.

#### ONTOLOGY TERM SEARCH

SEARCH



Plant Ontology (PO)



Trait Ontology (TO)



Crop Ontology (CO)



## **Tripal Plant PopGen Submit (TPPS)**

Phenotype Metadata File: Pleas your phenotypes: *	se upload a file conta	nining columns with the name	e, attribute, description, and units of each of
phenotype metadata.xlsx	REMOVE		
File Upload empty field: NA			
By default, TPPS will treat cells here. • DEFINE DATA	with the value "NA" a	as empty. If you used a differe	TreeGenes Data
Please define which columns h	old the required data	: Phenotype name	Repository Accession
name	attribute	units	(TGDR#####) -> DOI
Phenotype Name/Identifier \$	Attribute	<b>♦</b> Units	Description
phenotype 1	age	years	quantitative
phenotype 2	age	years	quantitative
phenotype 3	age	years	quantitative
Phenotype alle below.  Phenotype 2 Attribute: *  Some examples of attributes include: 'an 'composition', etc.  Phenotype 2 Description: *  Please provide a short description of Phenotype 2 Units: *		y", "area", "height", "age", "broken", "time",	+ -
Humidity regime		Click here to view trees on map	Map data @2018 Google, IN:Terms of Use

FAIR
Findable
Accessible

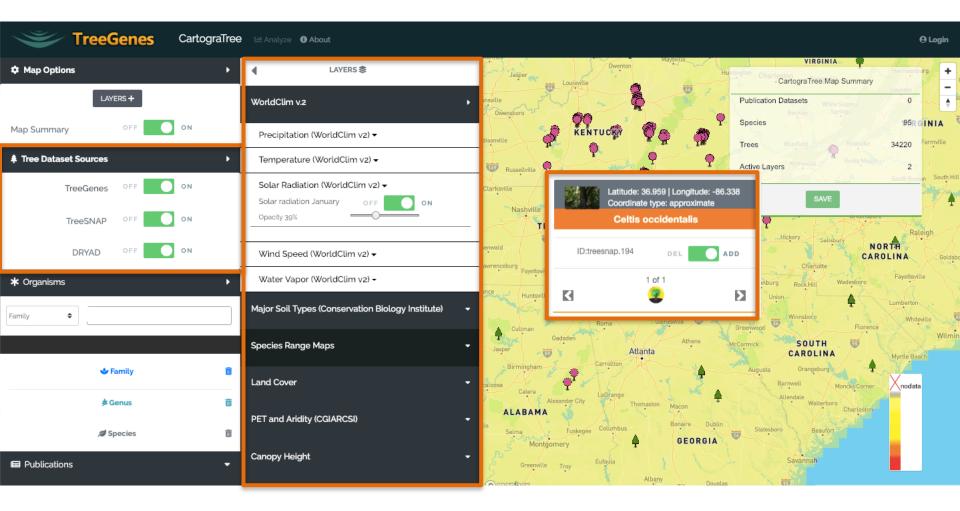


## TreeGenes to CartograTree

TPPS/TGDR DETAILS FOR TGDR001		
Attribute	Details	
Accession	TGDR001	
Title	Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (Populus trichocarpa, Salicaceae) secondary xylem.	
Species	Populus trichocarpa	
Study Type	GxP	
File Downloads	<ul> <li>✓ ASSOCIATION RESULTS FILE 1 FILE</li> <li>✓ GENOTYPES SNP</li> <li>✓ GPS COORDINATES</li> <li>✓ HAPLOTYPE DATA FILE</li> <li>✓ PHENOTYPES</li> <li>✓ PHENOTYPES DEFINITIONS</li> </ul>	
CartograTree	View in CartograTree	
Tree Count	1376	
Phenotype Count	4032	
Unique Phenotypes	3	

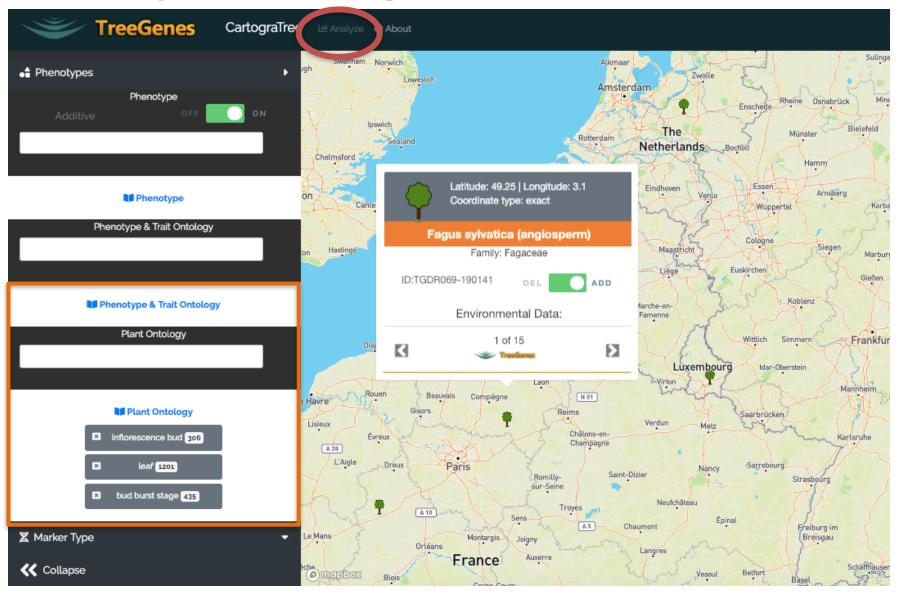


# CartograTree: Integrating environmental layers with georeferenced trees

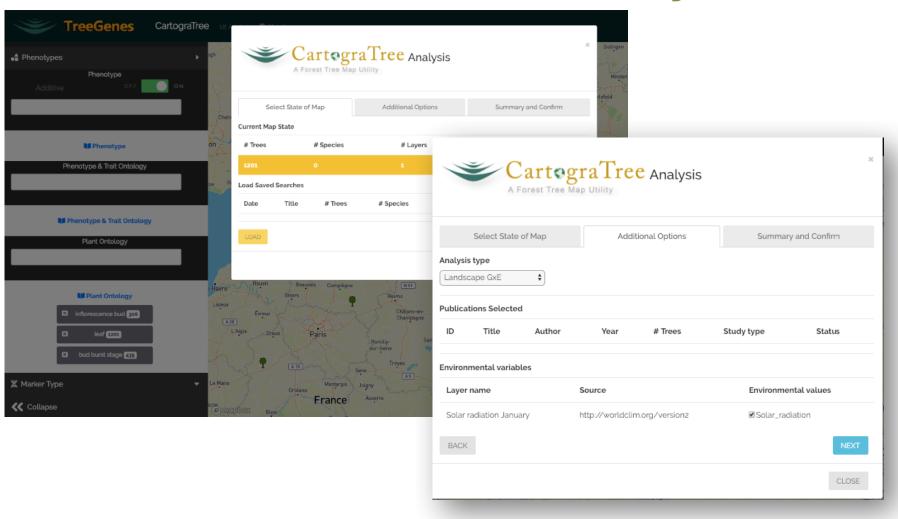




# CartograTree: Integrating environmental layers with georeferenced trees

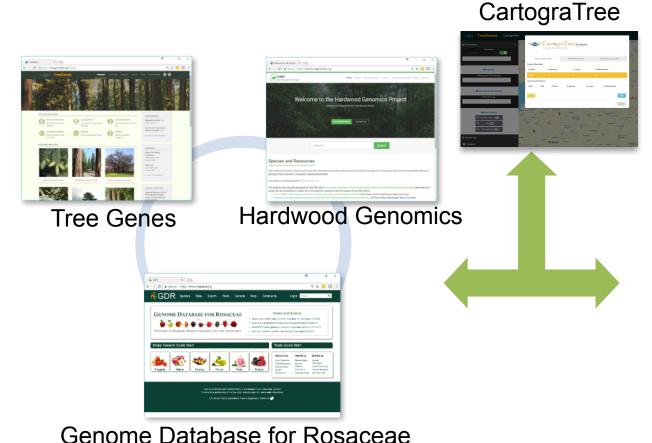


# CartograTree: Save searches locally and select for meta-analysis





## **Bringing Analytical Capacity to the Data**

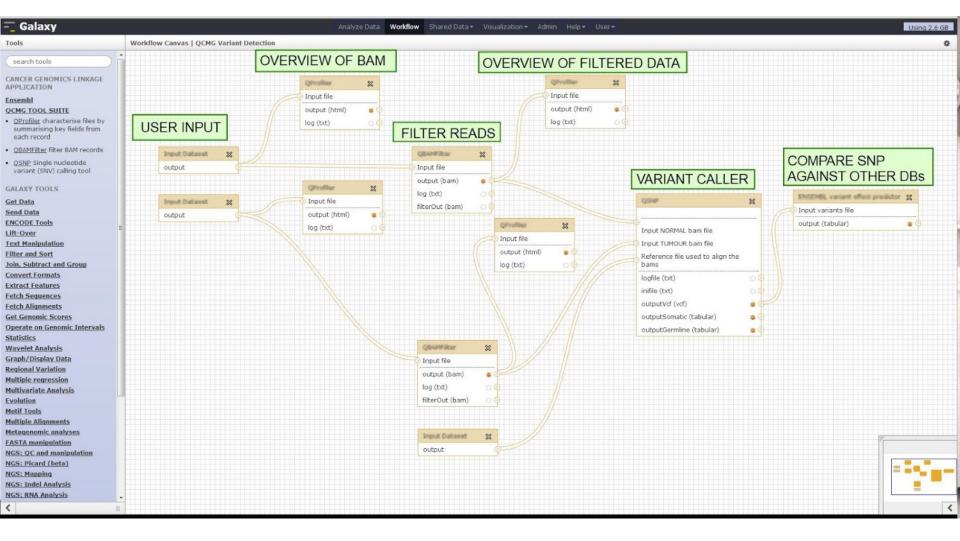






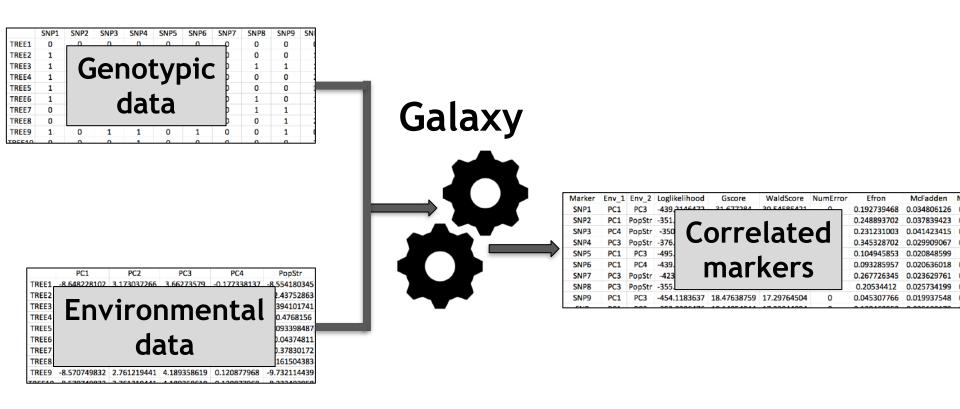


## Galaxy: Open Source Web-based Platform for Bioinformatic Analysis



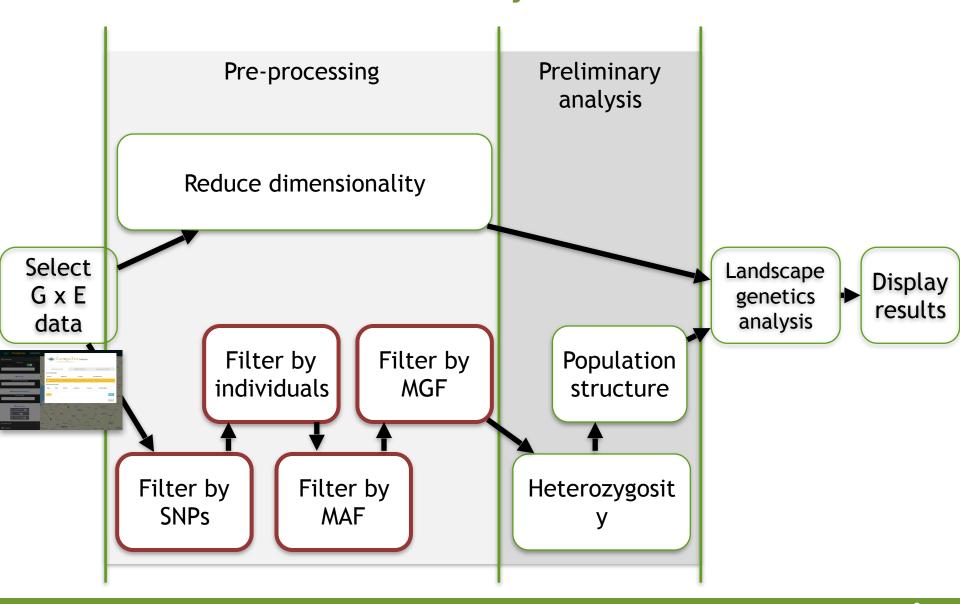


## **Workflows for Landscape Genomics: Integrating across diverse datasets**

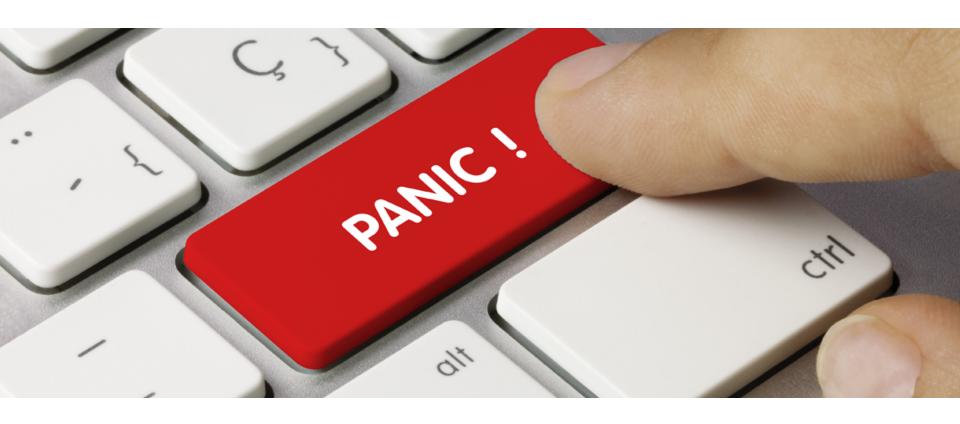




### **Workflows: Executed in Galaxy with metadata**



## Sustainability?





## Sustainability

### **Funding**

- CartograTree
- Modules + APIs
- Tripal Core

### Collaborative

- Consultant
- DMP Coordination

### Industry/NGO/Gov

- Industry partners?
- GTTN
- Forest Service



## Global Timber Tracking Network



## Plant Computational Genomics University of Connecticut

- Nic Herndon
- Emily Grau
- Sean Buehler
- Ronald Santos
- Risharde Ramnath
- Peter Richter

### Washington State University

- Stephen Ficklin
- Shawna Spoor
- Doreen Main
- Sook Jung







- Margaret Staton
- Ming Chen
- Bradford Condon







