## TreeGenes

# A Comprehensive Resource for Forest Tree Genomics



dendrome.ucdavis.edu

Emily Grau
Department of Plant Sciences
University of California, Davis

# TreeGenes Database: History

dendrome.ucdavis.edu

 Began as the Dendrome project (USDA funded initiative) in 1993 to hold forest tree genetic maps and associated markers



# TreeGenes Database: History

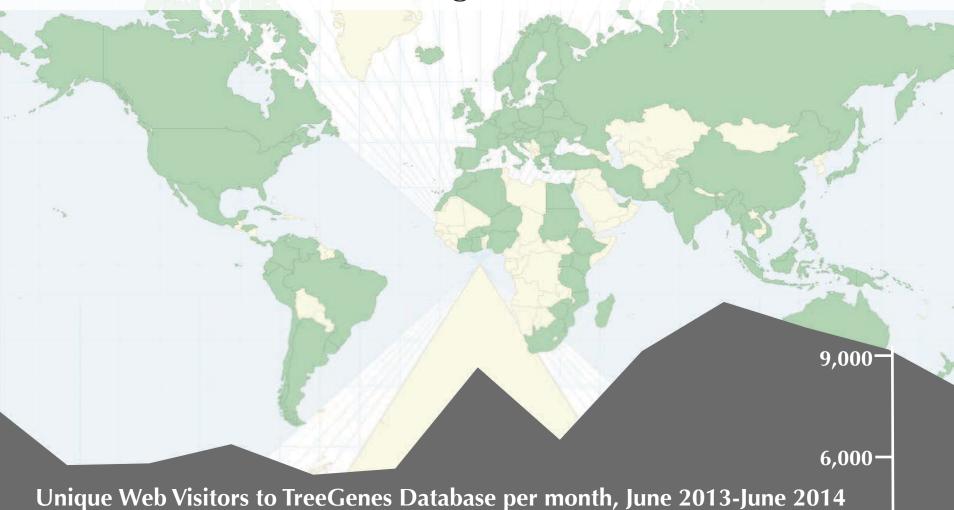
- Began as the Dendrome project (USDA funded initiative) in 1993 to hold forest tree genetic maps and associated markers
- Expanded to other data types
  - Sequence
    - Resequencing, Large-Scale Genotyping, Transcriptomics/ Expression
    - Full Genome Sequences
  - Analysis and Visualization Tools
    - Ability for users to mine the data
  - Resources for the user community
    - Literature, Colleagues



# TreeGenes Database: Users

dendrome.ucdavis.edu

2,060 users from 849 organizations in 94 countries



# TreeGenes Database: Species



- 1,290 species from 101 genera
  - At least one genetic artifact from each species
  - Conifers but is also inclusive of all forest trees
- Full genome sequence: 13 species
- Transcriptome/Expression resources: 3,920,817 sequences from 263 species
- 106 genetic maps from 35 species

dendrome.ucdavis.edu

Automated User submissions

dendrome.ucdavis.edu

### **Automated**

- NCBI (primary repositories)
  - Protein, EST, cDNA, TSA, Unigene databases
  - Introduced to TreeGenes with added value
  - Information should be sent to primary dbs first
- Literature
  - Web of Science, PubMed



</AbstractText>

 - <AbstractText Label="METHODS" NlmCategory="ME Pollination drops were sampled from eight gymnospern

dendrome.ucdavis.edu

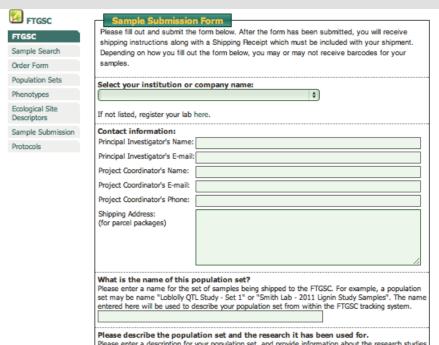
### User submissions

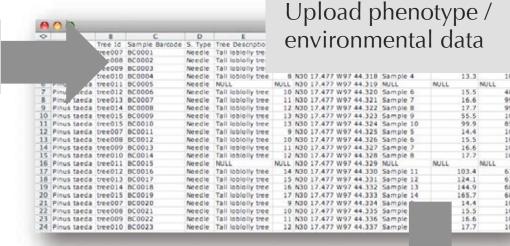
- Internal projects or collaborations (day one)
- Submissions of data post-analysis at publication time

dendrome.ucdavis.edu

User submissions (Software with full front-end and back-end support) Laboratory Information Management System Sequence, Genotype, Phenotype, Environmental Information

Track barcoded samples from collection through sequencing





Data can be integrated into TreeGenes in real time or at project end

dendrome.ucdavis.edu

User submissions: external

Most submissions from TGG

Submit genetic maps or population study data



Obtain TGDR accession number!



## TreeGenes Database: Data Access

dendrome.ucdavis.edu

### Interfaces

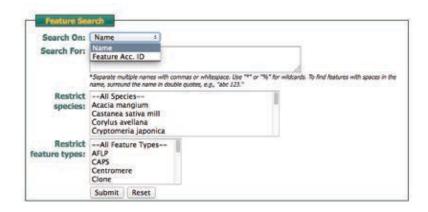
- Existing viewers
- Custom development

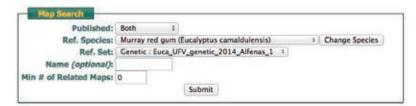
dendrome.ucdavis.edu



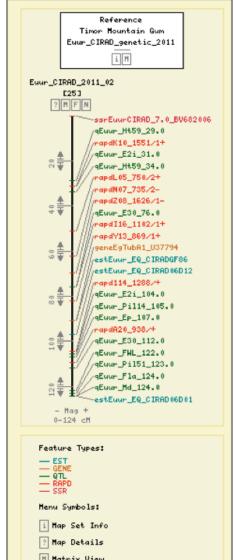


## Comparative mapping





Items I to 11 of 11.								
Map Name	Related	Related Maps per unit	Correspondences	Correspondences per unit	Start	Stop	SSR	
							total	per unit
Euca_UFV_2014_1 (spider)	0	0/unit	0	0/unit	0.00	165.80	12	72.38/K
Euca_UPV_2014_10 (spider)	0	0/unit	0	0/unit	0.00	159.50	8	50.16/K
Euca_UFV_2014_11 (spider)	0	0/unit	Û	0/unit	0.00	64.60	6	92.88/K
Euca_UFV_2014_2 (spider)	0	0/unit	0	0/unit	0.00	65.00	9	0.14/unit
Euca_UFV_2014_3 (spider)	0	0/unit	0	0/unit	0.00	14.80	3	0.20/unit
Euca_UFV_2014_4 (spider)	0	0/unit	0	0/unit	0.00	70.30	8	0.11/unit
Euca_UPV_2014_5 (spider)	0	0/unit	0	0/unit	0.00	53.20	3	56.39/K
Euca_UFV_2014_6 (spider)	0	0/unit	0	0/unit	0.00	144.90	8	55.21/K
Euca_UFV_2014_7 (spider)	0	0/unit	0	Q/unit	0.00	185.60	5	26.94/K
Euca_UPV_2014_6 (spider)	0	0/unit	0	0/unit	0.00	67.20	8	0.12/unit
Euca_UFV_2014_9 (spider)	0	0/unit	0	0/unit	0.00	53.70	5	93.11/K



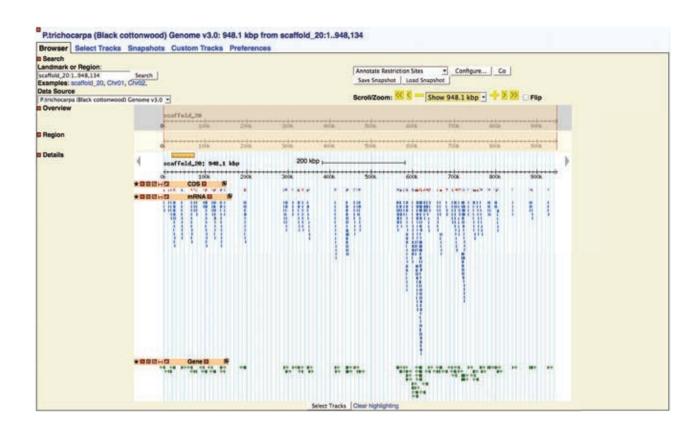
dendrome.ucdavis.edu







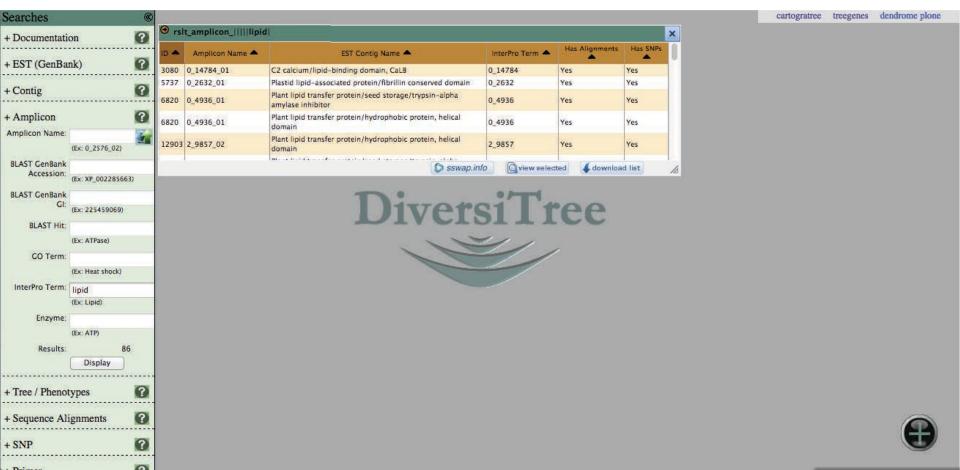
## Genome browsing & annotation



dendrome.ucdavis.edu



 Bulk retrieval of resequencing data, genotypes, and phenotypes

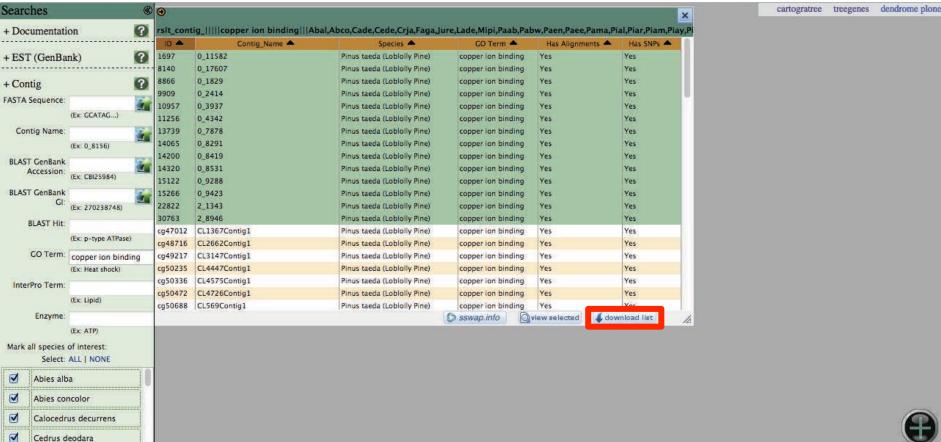


dendrome.ucdavis.edu

### **DiversiTree** Download results



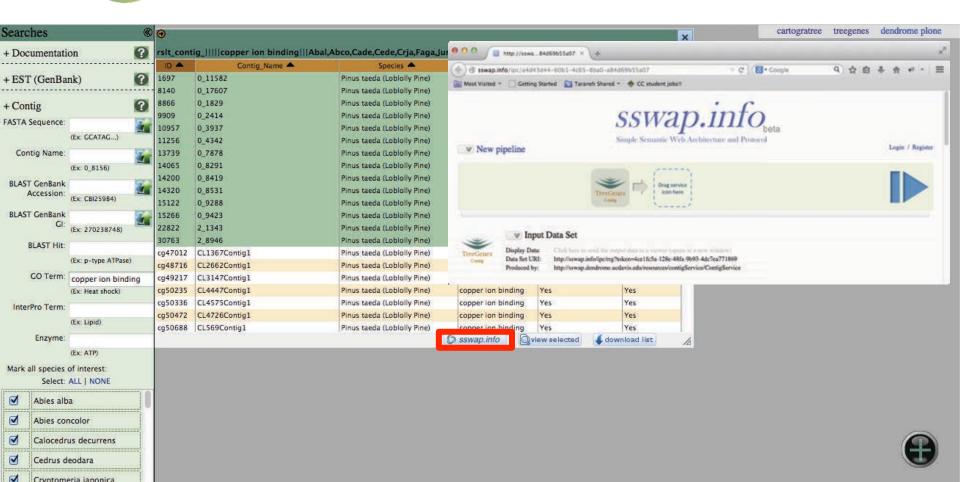
Cryptomeria japonica



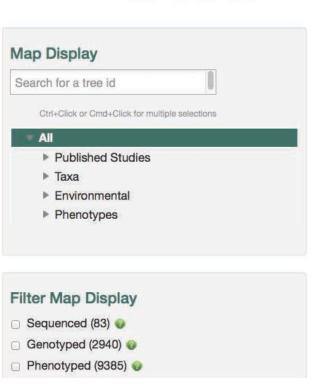


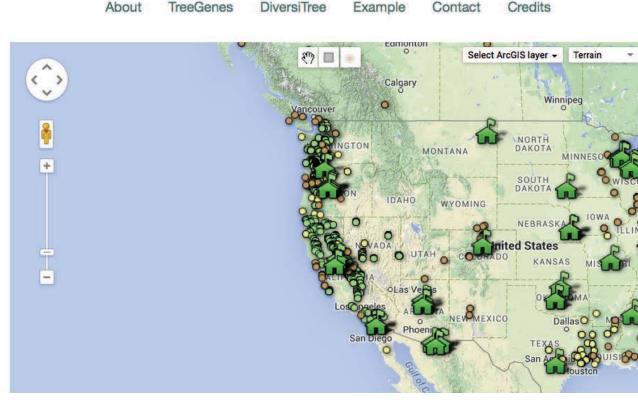
dendrome.ucdavis.edu

**DiversiTree** Download results or pipe to CartograTree via SSWAP (Simple Semantic Web Architecture Protocol)



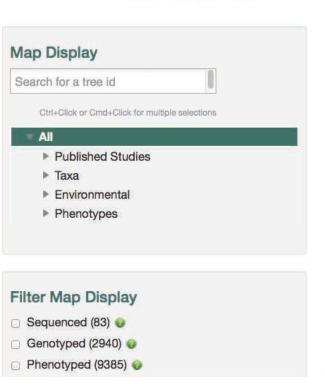


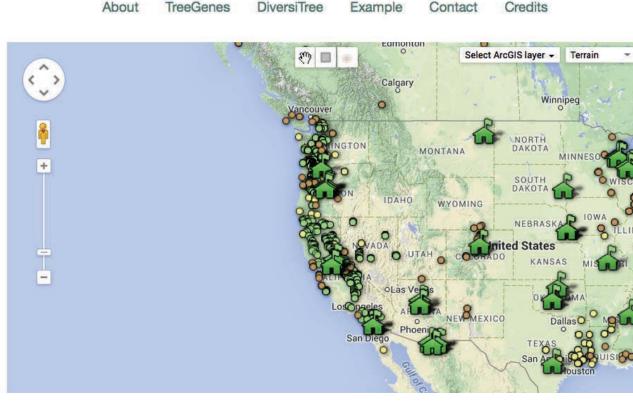




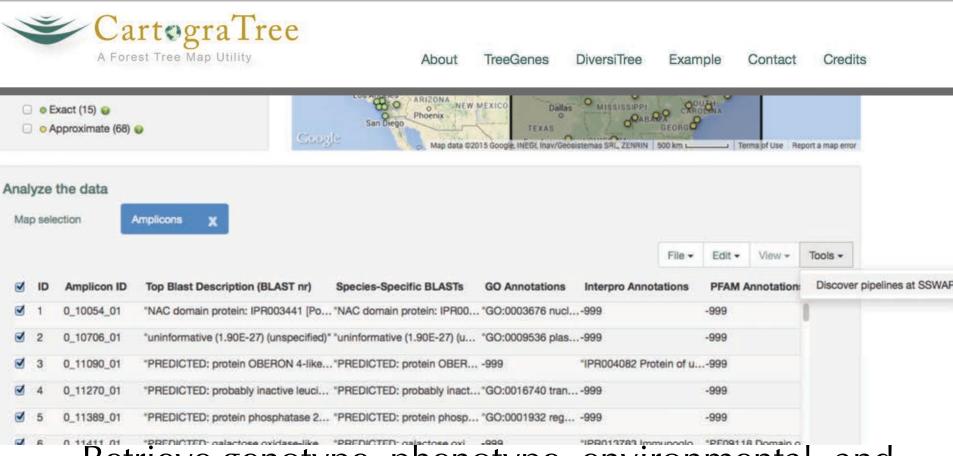
- Providing context to geo-referenced data
- Originated from Tree Biology Working Group through iPlant







- Data from TreeGenes, WorldClim, Ameriflux, TRY-db
- Google fusion tables & Google maps



- Retrieve genotype, phenotype, environmental, and sequence data
- Further analysis (TASSEL, MUSCLE) via SSWAP

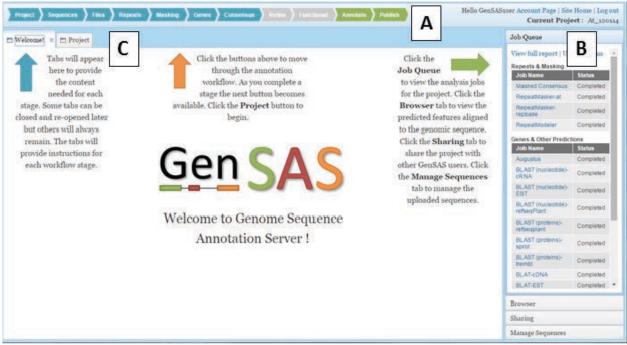
dendrome.ucdavis.edu



## **Genome Sequence Annotation Server**

Can handle large, complex genomes





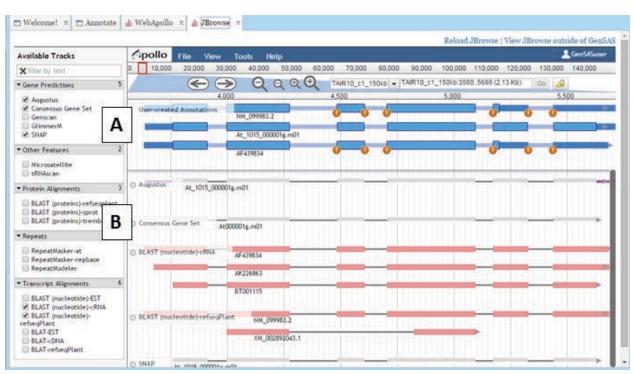
dendrome.ucdavis.edu



Save work, upload modifications for approval

P1153





dendrome.ucdavis.edu















dendrome.ucdavis.edu









- Tripal
  - Frontend & backend open source database solution
  - CHADO: database schema from GMOD
  - Drupal: open source web development platform
  - TreeGenes will transition into using Tripal to ease data transfer

dendrome.ucdavis.edu









- Galaxy
  - Workflow & data analysis platform
  - Build multi-step analysis pipeline
  - Tripal Galaxy will develop modules for analysis with Galaxy

dendrome.ucdavis.edu











- Work with other databases
- Improve data integration, data transfer
- Pull datasets easily from other datasets
   & sources on the web for analysis

# TreeGenes Database: Team

### dendrome.ucdavis.edu

tg-help@ucdavis.edu

### **Project Leads**

David Neale Jill Wegrzyn

University of Connecticut

## **Lead Database Administrator**

**Emily Grau** 

### **Development Team**

**Jacob Zieve** Hans Vasquez-Gross

## **Advising**

Damian Gessler

**Andrew Brown** 

Semantic Options/University of Arizona







