

# TreeGenes

PAG 2019

Emily Grau



Welcome to the new TreeGenes Database!  
Powered by Tripal!

#### EXPLORE TREEGENES



##### Browse Genomes

Browse genomes with  
JBrowse



##### CartograTree

CartograTree is a map-based  
web app.



##### Download Data

Download data from our FTP  
site



##### Sequence Search

Search sequences with  
DIAMOND



##### Species

Find your species of interest



##### Submit

Submit your data to  
TreeGenes.

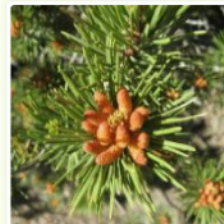
#### FEATURED PROJECTS



B4est



Oak Genome Project



CoAdapTree



10kp

#### MEETINGS

##### PAG XXVII

January 12 to 16 2019  
San Diego, USA

##### 2019 SFTIC

June 3 to 6 2019  
Lexington, USA

- [View more meetings](#)

#### LATEST LITERATURE

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



# TreeGenes

## Data

Genotype, phenotype, environment

Literature

Community

## Curation

## Tools



# TreeGenes Database: Species



- 1,850 species from 153 genera
- Full genome sequence: 34 species
- Transcriptome/Expression resources: 4,761,916 sequences from 317 species
- 101 genetic maps from 37 species
- Population studies
  - Georeferenced trees
  - Extensive genotypic (GBS and array) and phenotypic data



# Data Sources

## Primary + secondary repositories:

*NCBI, Dryad, Phytozome, Web of Science, PLAZA*

Genome assemblies and annotations

Transcriptomes

Genetic markers

Raw data accessions (link out to source)

Literature



# Data Sources

## TreeGenes

### *Data Repository*

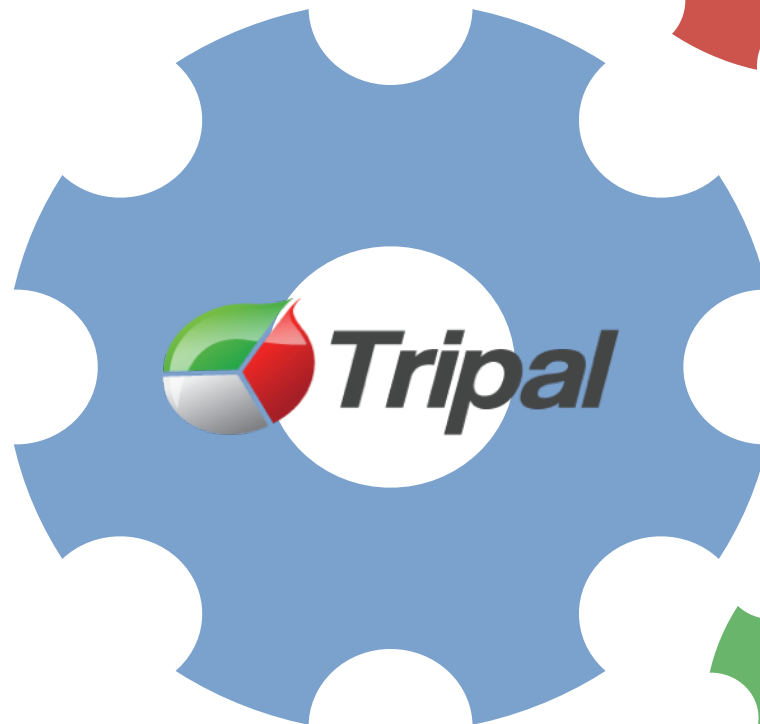
Genetic maps

Association and population study  
datasets

### *TreeGenes Curation*

TreeGenes UniGenes







Graphic interface for command line analysis

Shareable, reusable

Community support

Tripal Galaxy module





Welcome!

[Home](#)

[View](#)
[Edit](#)
[Manage display](#)
[Devel](#)

Welcome to



**TreeGenes**

You're now logged in and ready to go!

### USER MENU

[Edit Profile](#) >

[Log out](#) >

[My account](#) >

[Add new job posting](#) >

[Add new meeting posting](#) >

### MEETINGS

#### PAG XXVII

January 12 to 16 2019  
San Diego, USA

#### 2019 SFTIC

June 3 to 6 2019  
Lexington, USA

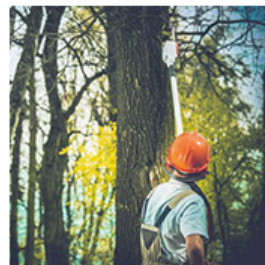
• [View more meetings](#)



Add/Edit Profile



Add/Edit  
Workshop/Conference



Add/Edit Employment  
Listings



# Eucalyptus grandis

Summary

Genome

Transcriptome

Literature

TPPS

**Species:**

Eucalyptus grandis

**Common Name:**

Grand Eucalyptus

**TreeGenes Code:**

Eugr

**Lineages:**

Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta;  
Euphyllophyta; Spermatophyta; Magnoliophyta; Mesangiospermae; eudicotyledons;  
Gunneridae; Pentapetalae; rosids; malvids; Myrtales; Myrtaceae; Myrtoideae;  
Eucalypteae; Eucalyptus

**Haploid Chromosome  
Number:**

12

**GBIF:**

<https://www.gbif.org/species/3177028>

**Bioproject Records:**

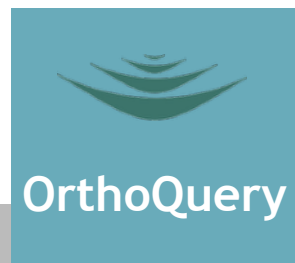
50

**Biosample Records:**

130



# Tools



# 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, transcripts (BLASTN) while protein or translated protein searches are executed with BLASTp (BLAST) at a similar sensitivity. Both packages accept similar input parameters. You can provide FASTA formatted sequences for searching. You can select from a list of databases provided in FASTA format and formatted for you.

### SEQUENCE QUERY TYPE

☐ Protein

☒ Nucleotide (coding or whole genome)

Database Search Type

### PROVIDE/LOAD QUERY SEQUENCE

ccgaagtccttgctaatgatttccggtcatcccatgctttctcaaggatccttcatgcattatgtt

tttcatgtatgtgtctcacccggcgaagtatccgtataaaccaattatccaagcattccctcgatt

Choose File

No file chosen


UPLOAD

### TARGET/DATABASE SELECTION

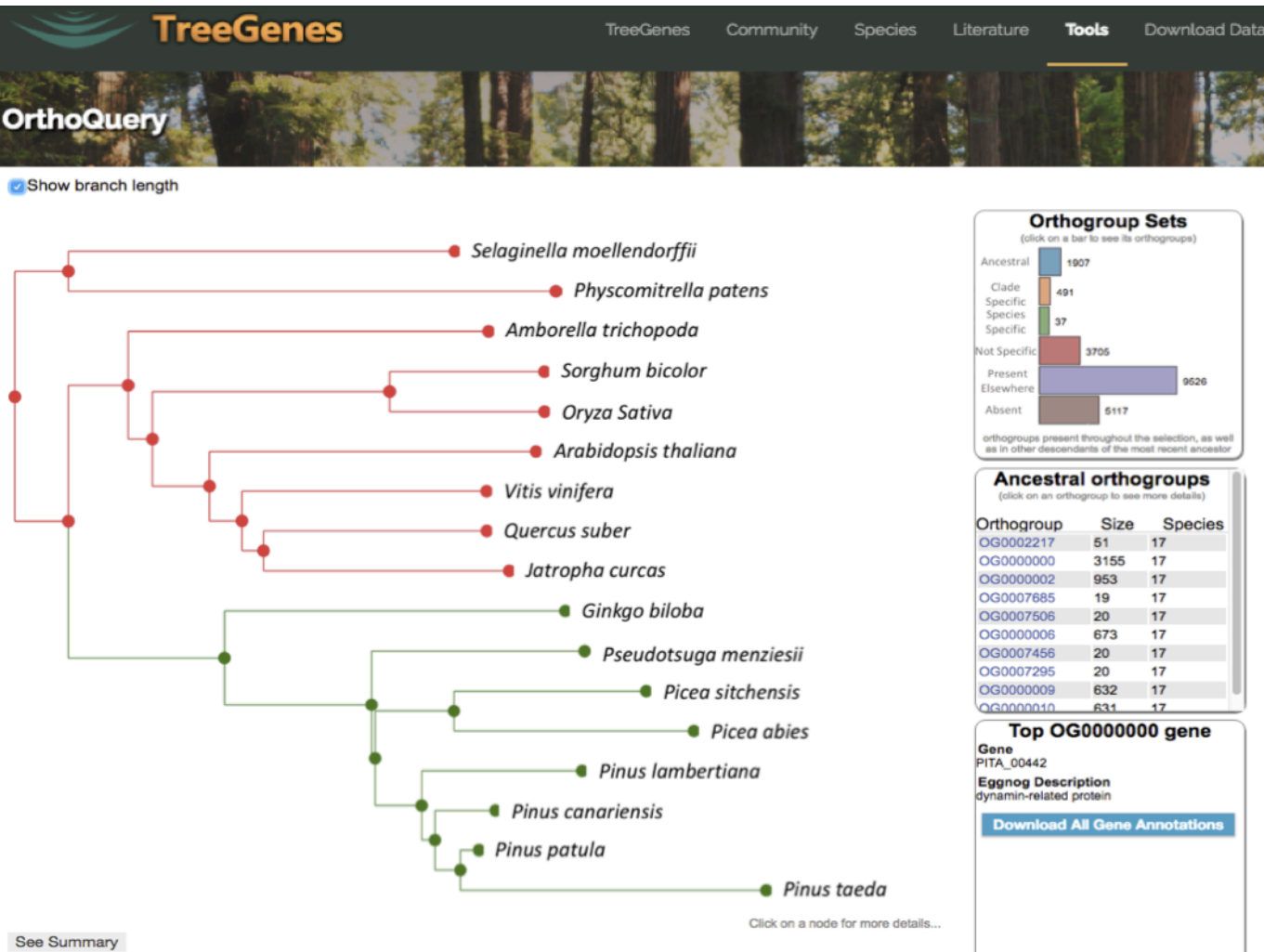
Protein

Protein (UniGene)

Protein (TSA)

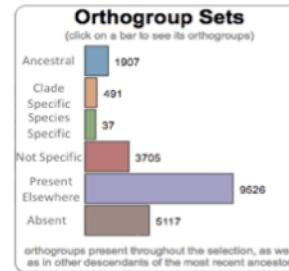


# OrthoQuery



Poster  
PO0021

Computer  
Demo  
C25



**Ancestral orthogroups**  
(click on an orthogroup to see more details)

Orthogroup	Size	Species
OG0002217	51	17
OG0000000	3155	17
OG0000002	953	17
OG0007685	19	17
OG0007506	20	17
OG0000006	673	17
OG0007456	20	17
OG0007295	20	17
OG0000009	632	17
OG0000010	631	17

**Top OG0000000 gene**

Gene  
PITA\_00442

**EggNog Description**  
dynamin-related protein

[Download All Gene Annotations](#)





# Tripal Plant PopGen Submit

## Population and association study submission pipeline

### Tripal Plant PopGen Submit pipeline

✓ Author and Species Information   ✓ Experimental Conditions   ✓ Tree Accession   ⚠ Submit Data

#### FRAXINUS PENNSYLVANICA:

##### ▼ PHENOTYPE INFORMATION:

Upload a file and/or fill in form fields below to provide us with metadata about your phenotypes.

☐ I would like to upload a phenotype metadata file

ADD PHENOTYPE

REMOVE PHENOTYPE



##### Phenotype 1:

Phenotype 1 Name: \*

Phenotype "name" is the human-readable name of the phenotype, where "attribute" is the thing that the phenotype is describing. Phenotype "name" should match the data in the "Phenotype Name/Identifier" column that you select in your Phenotype file below.

Phenotype 1 Attribute: \*

Some examples of attributes include: "amount" "width" "mass density" "area" "height" "age" "broken" "time" "color"



# TPPS

## TPPS Details by Accession

[Home](#)[Return to TPPS List](#)

### TPPS/TGDR DETAILS FOR TGDR001

Attribute	Details
Accession	TGDR001
Title	Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood ( <i>Populus trichocarpa</i> , Salicaceae) secondary xylem.
Species	<i>Populus trichocarpa</i>
Study Type	GxP
File Downloads	<ul style="list-style-type: none"><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	<a href="#">View in CartograTree</a>
Tree Count	1376
Phenotype Count	4032
Unique Phenotypes	3

#### USER MENU

- [Edit Profile](#) >
- [Log out](#) >
- [My account](#) >
- [Add new job posting](#) >
- [Add new meeting posting](#) >

#### MEETINGS

**PAG XXVII**  
January 12 to 16 2019  
San Diego, USA

**2019 SFTIC**  
June 3 to 6 2019  
Lexington, USA

- [View more meetings](#)



# TPPS

## TPPS Details by Accession

[Home](#)[Return to TPPS List](#)

### TPPS/TGDR DETAILS FOR TGDR001

Attribute	Details
Accession	TGDR001
Title	Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood ( <i>Populus trichocarpa</i> , Salicaceae) secondary xylem.
Species	<i>Populus trichocarpa</i>
Study Type	GxP
File Downloads	<ul style="list-style-type: none"><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	<a href="#">View in CartograTree</a>
Tree Count	1378
Phenotype Count	4032
Unique Phenotypes	3

### USER MENU

- [Edit Profile](#) >
- [Log out](#) >
- [My account](#) >
- [Add new job posting](#) >
- [Add new meeting posting](#) >

### MEETINGS

**PAG XXVII**  
January 12 to 16 2019  
San Diego, USA

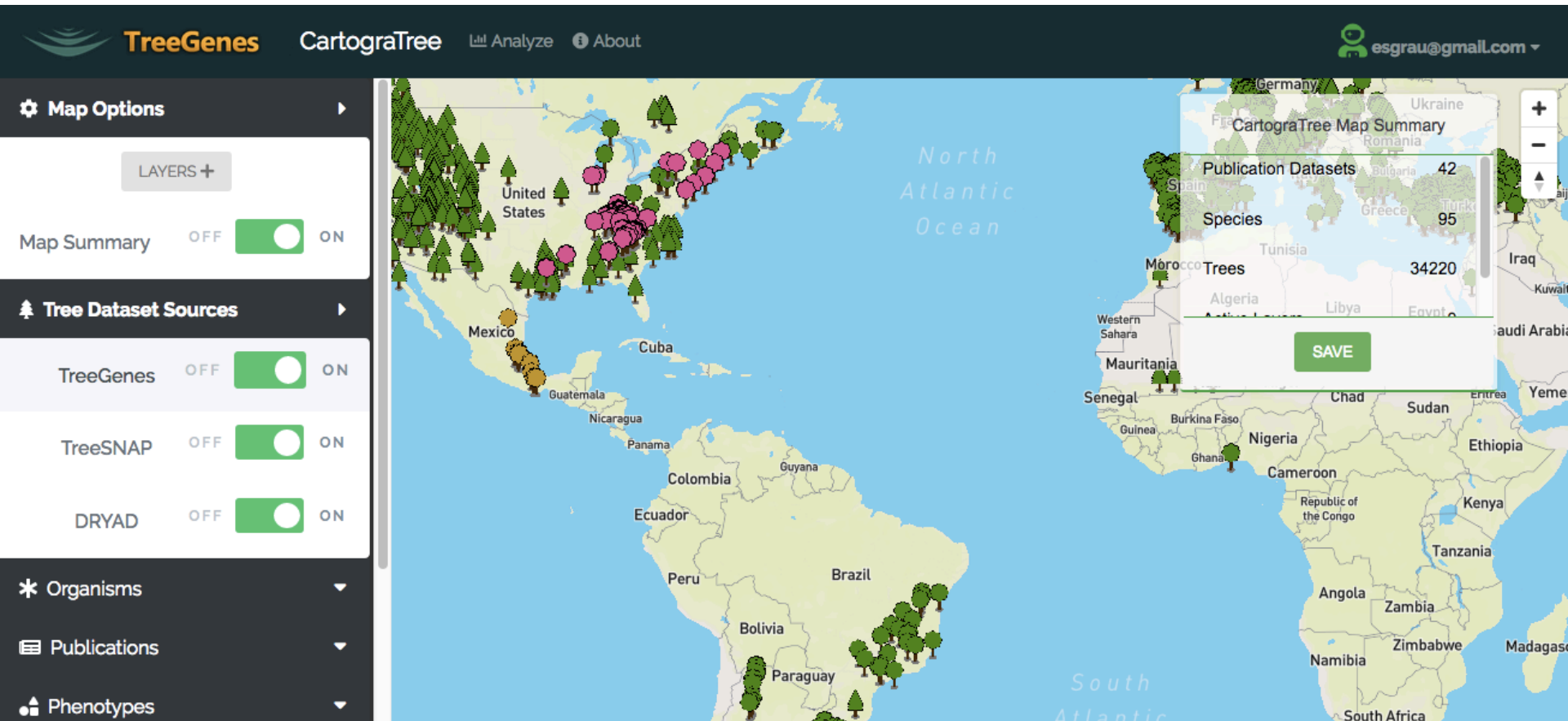
**2019 SFTIC**  
June 3 to 6 2019  
Lexington, USA

- [View more meetings](#)

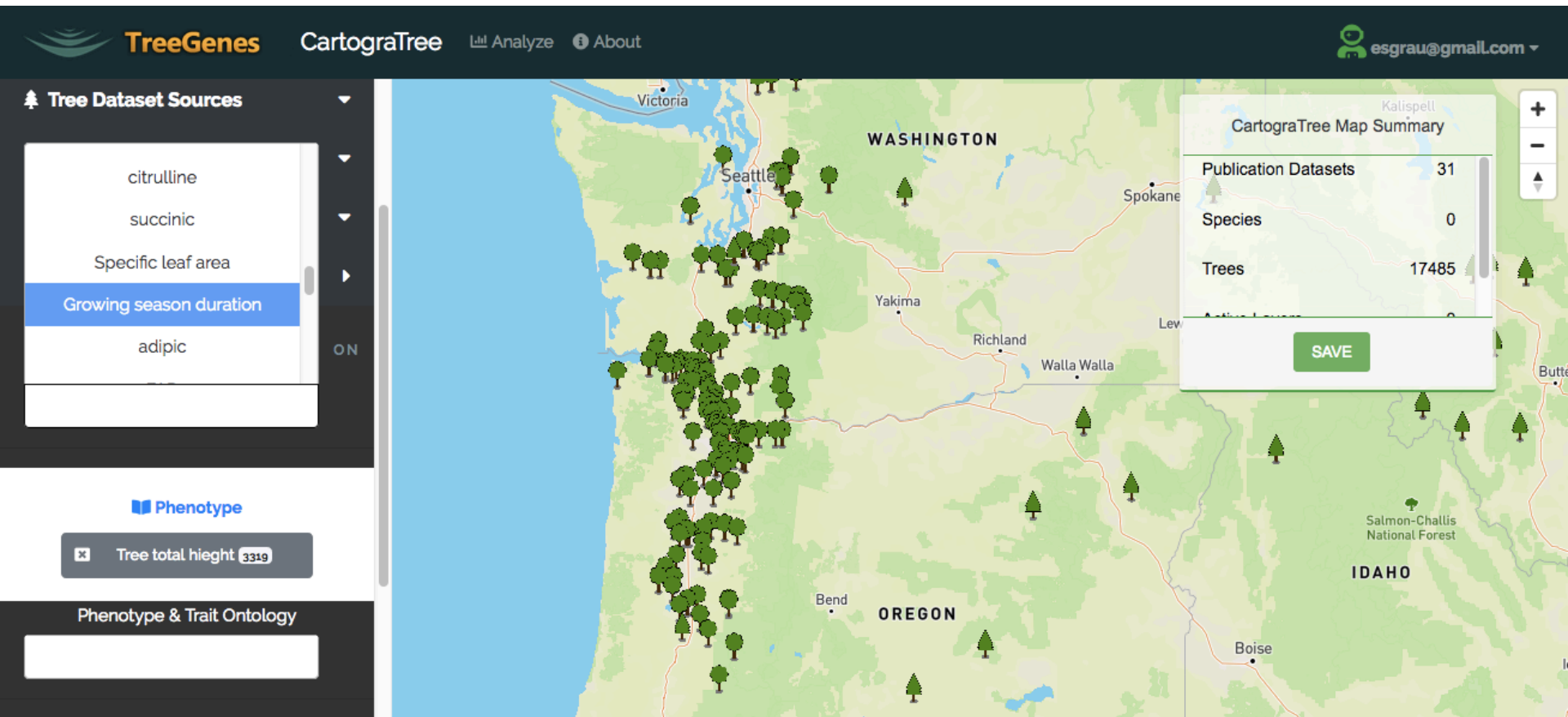




# CartograTree



# CartograTree



# CartograTree

The screenshot displays the CartograTree web application interface. The top navigation bar includes the TreeGenes logo, the site name "CartograTree", and links for "Analyze" and "About". A user profile icon and email address "esgrau@gmail.com" are visible on the right.

The main interface is divided into several sections:

- Map Options:** Includes a "LAYERS +" button and a "Map Summary" toggle switch currently set to "ON".
- Tree Dataset Sources:** A section for selecting data sources.
- Organisms:** A section for selecting organisms, with a "Genus" dropdown menu.
- Family:** A section for selecting the family, currently showing "Pinus" with 15 results.
- Map:** A map of California showing the distribution of tree species. A pop-up window displays details for a selected tree:

**Pinus lambertiana (gymnosperm)**  
Family: Pinaceae  
ID: IFG0292  
Latitude: 39.559 | Longitude: -120.404  
Coordinate type: exact

**Environmental Data:**

Soil Type	Soil Descr
Acrisols	Soils with subsurface accumulation of low activity clays and low base saturation

**CartograTree Map Summary:**

Publication Datasets	Species	Trees
7	1	5083

**Major Soil Types:**

- Acrisols
- Alisols
- Andosols
- Anthrosols
- Arenosols
- Calcisols
- Cambisols
- Chernozems
- Ferralsols



# CartograTree

**TreeGenes** CartograTree **Analyze** About

esgrau@gmail.com

**Map Options**

LAYERS +

Map Summary OFF ☒ ON

**Tree Dataset Sources**

**Organisms**

Genus

Family

Genus

Pinus 15

Latitude: 39.559 | Longitude: -120.404  
Coordinate type: exact

**Pinus lambertiana (gymnosperm)**

Family: Pinaceae

ID: IFG0292 DEL ☒ ADD

**Environmental Data:**

Soil Type Acrisols

Soil Descr Soils with subsurface accumulation of low activity clays and low base saturation

1 of 4

**CartograTree Map Summary**

Publication Datasets	7
Species	1
Trees	5083

SAVE

**Major Soil Types**

- Acrisols
- Alisols
- Andosols
- Anthrosols
- Arenosols
- Calcisols
- Cambisols
- Chernozems
- Ferralsols



# CartograTree



Select State of Map

Additional Options

Summary and Confirm

## Current Map State

# Trees	# Species	# Layers	# Publications
13683	1	1	0

## Load Saved Searches

Date	Title	# Trees	# Species	# Layers	# Publications
------	-------	---------	-----------	----------	----------------

LOAD

NEXT

CLOSE



Select State of Map

Additional Options

Summary and Confirm

## Analysis type

- ✓ Landscape GxE
- Associative Genetics PxG
- Population Structure G

## Publications Selected

ID	Title	Author	Year	# Trees	Study type	Status
----	-------	--------	------	---------	------------	--------

## Environmental variables

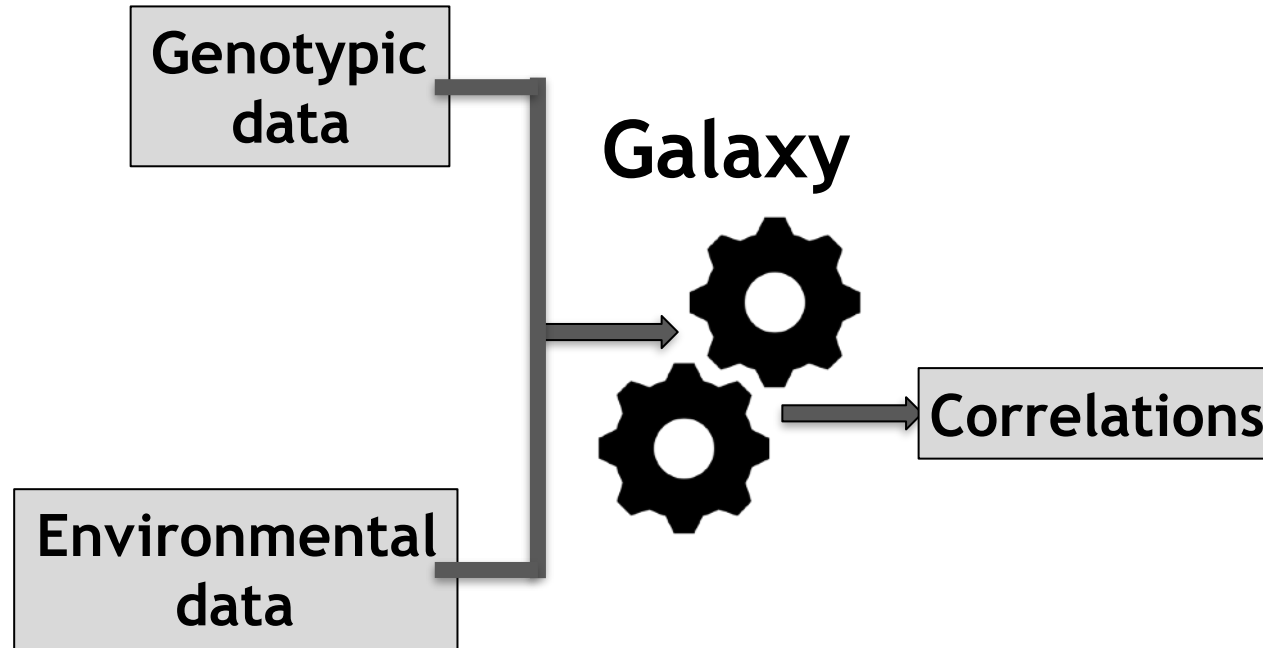
Layer name	Source	Environmental values
Major Soil Groups	<a href="http://www.arcgis.com/home/Item.html?id=1d16ed2a0aa24ab39e5ee6c491965883">http://www.arcgis.com/home/Item.html?id=1d16ed2a0aa24ab39e5ee6c491965883</a>	<input checked="" type="checkbox"/> value <input checked="" type="checkbox"/> grp_descr

BACK

NEXT



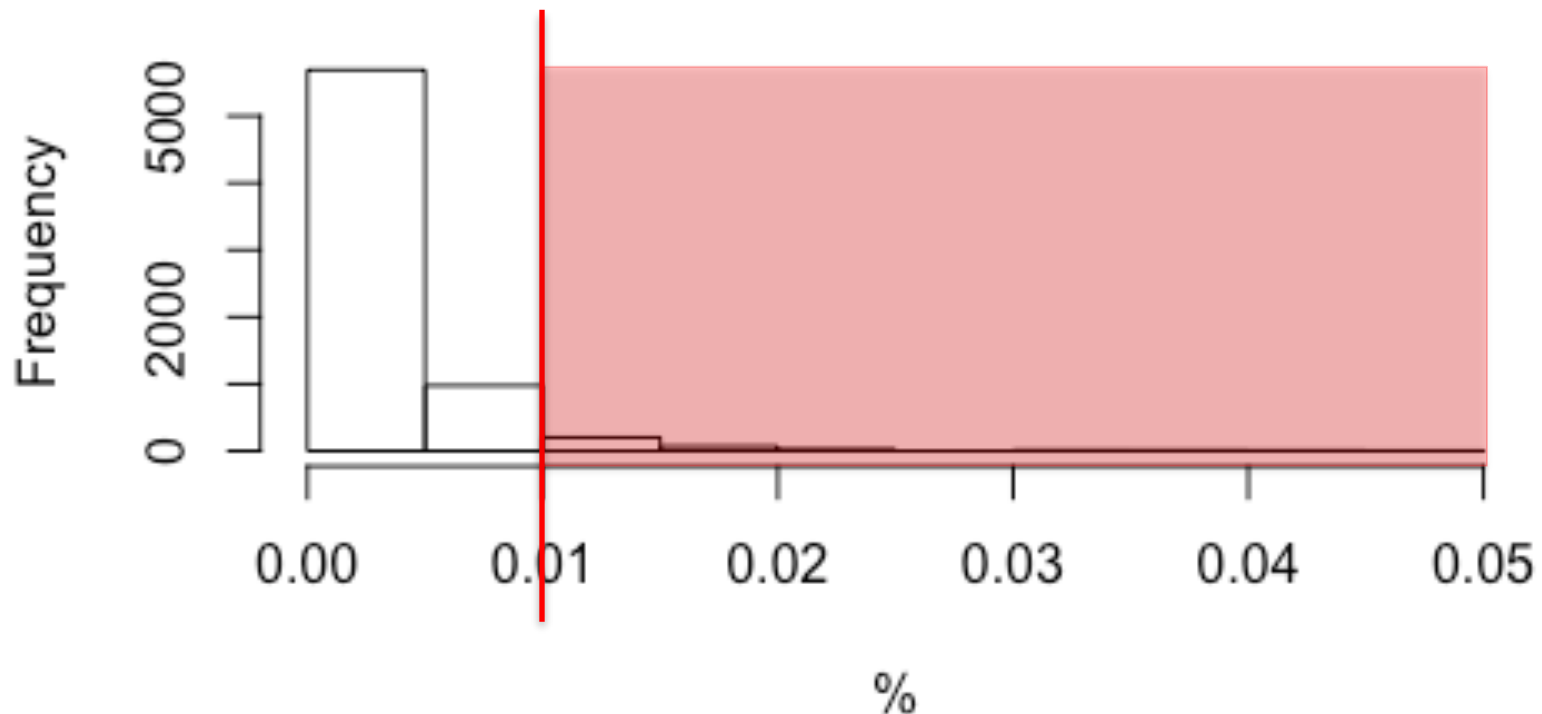
# CartograTree: Analysis



# CartograTree

## Filter genetic data: missing SNPs

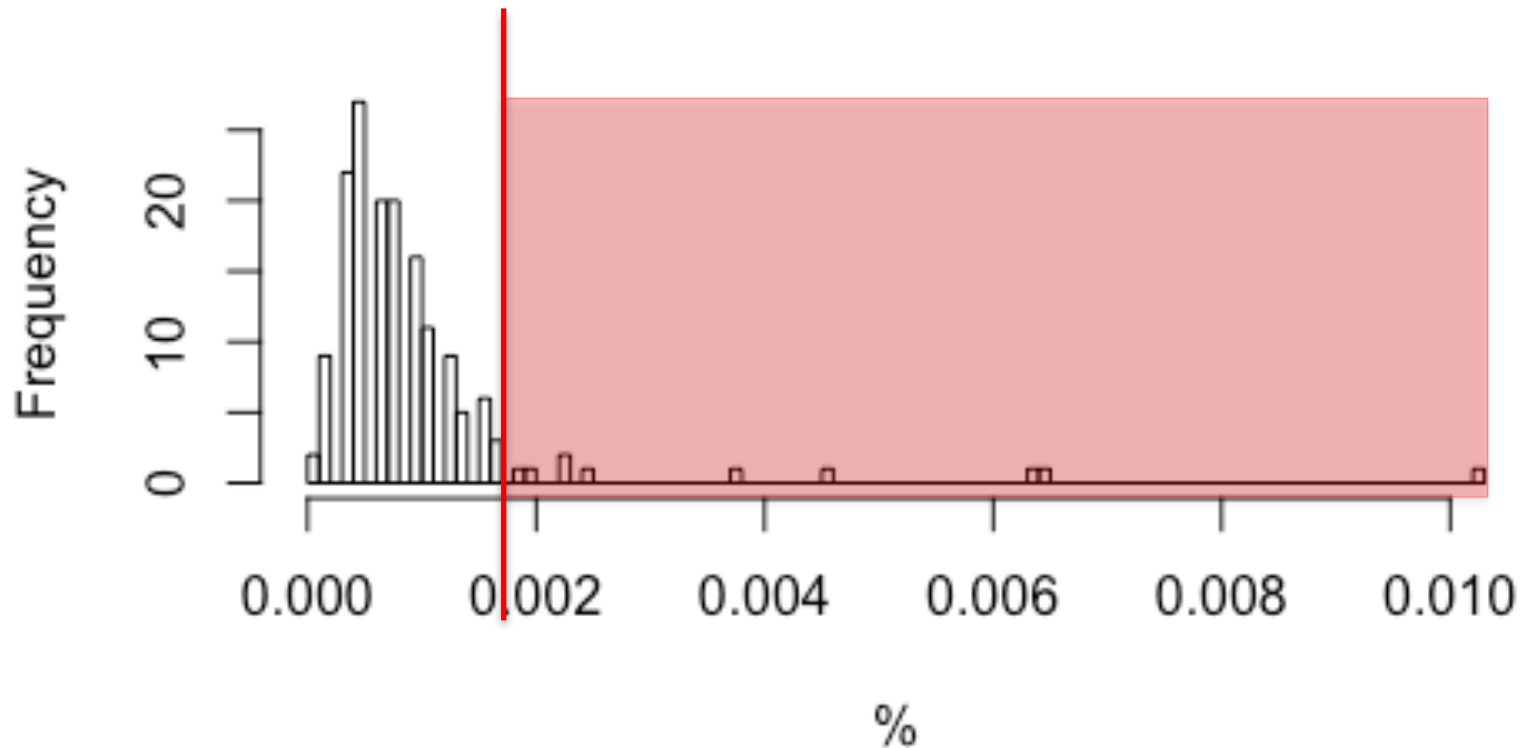
**% of missing observation across the individuals**



# CartograTree

Filter genetic data: missing individuals

**% of missing observation across the SNPs**

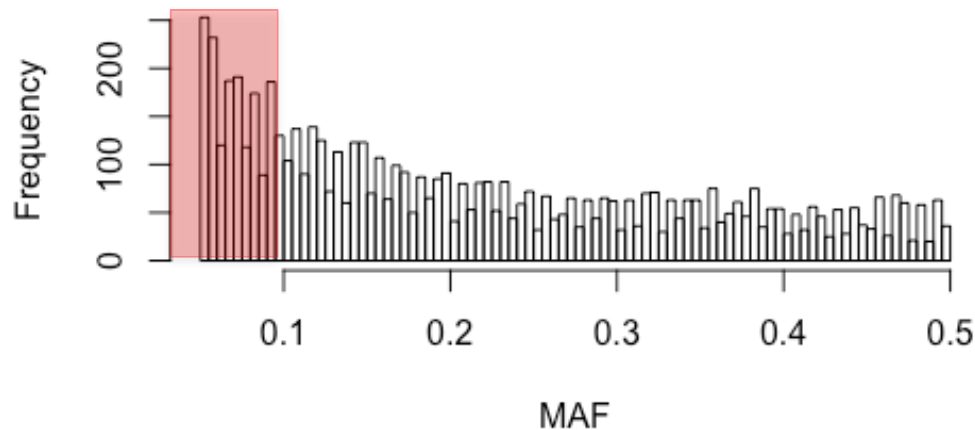




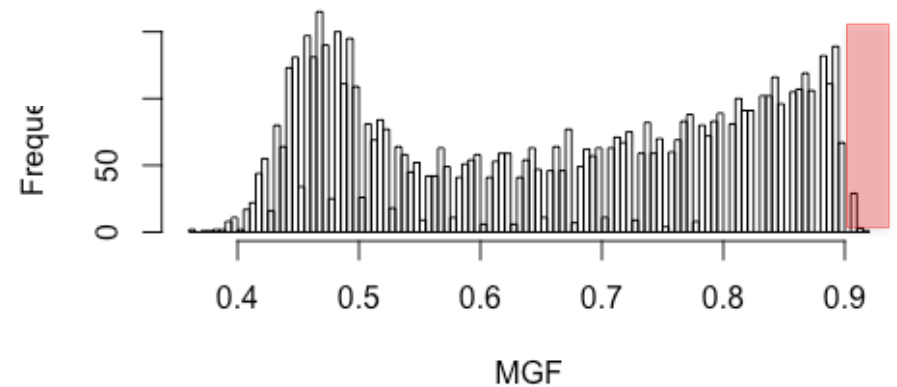
# CartograTree

## Filter genetic data: allele frequency

Minor Allele Frequency across all individuals



Major Genotype Frequency across all individuals



Tools

search tools

Inputs

[Get Data](#)

[Send Data](#)

[Collection Operations](#)

[Lift-Over](#)

[Text Manipulation](#)

[Filter and Sort](#)

[Join, Subtract and Group](#)

[Convert Formats](#)

[Fetch Alignments/Sequences](#)

[Operate on Genomic Intervals](#)

[Phenotype Association](#)

[NGS: QC and manipulation](#)

[AMTools](#)

[Comparative Genomics](#)

[Statistics](#)

[Graph/Display Data](#)

[Association Mapping](#)

[Sequence Search](#)

[Annotation](#)

[testing](#)

[Statonlab Tools](#)

[Aurora Galaxy Tools](#)

[usearch](#)

[vsearch](#)

[Data Manager Tools](#)

[Workflows](#)

Workflow Canvas | Orthofinder

Input dataset collection

output

VSearch clustering

Select your input FASTA file

- msaout (fasta)
- consout (fasta)
- centroids (fasta)
- alnout (fasta)
- notmatched (fasta)
- matched (fasta)
- blast6out (tabular)
- fastapairs (fasta)
- uc\_outfile (tabular)

OrthoFinder OnlyGroups

Select input fasta proteomes

- wdblast
- wdfasta
- genetrees
- orthogroups1 (txt)
- orthogroups2 (csv)
- specs\_overlap (csv)
- unassigned\_genes (csv)
- stat\_overall (csv)
- stat\_specs (csv)
- SpeciesIDs (txt)
- SequenceIDs (txt)

Details

Edit Workflow Attributes

Name:

Orthofinder

Tags:

Apply tags to make it easy to search for and find items with the same tag.

Annotation / Notes:

*Describe or add notes to workflow*  
Add an annotation or notes to a workflow; annotations are available when a workflow is viewed.

# Team

## TreeGenes

Jill Wegrzyn

Nic Herndon

Sean Buehler

Peter Richter

Risharde Ramnath

Ronald Santos



**Tripal** Stephen Ficklin



**HWG** Meg Staton



**GDR** Dorrie Main

[treegenesdb.org](http://treegenesdb.org)

[treegenesdb@gmail.com](mailto:treegenesdb@gmail.com)

