

# TreeGenes and CartograTree Tools for Forest Tree Genomics

Emily Grau, Sean Buehler, Nic Herndon, Taylor Falk, Peter Richter, Meg Staton, Stephen Ficklin, Jill Wegrzyn

University of Connecticut, Storrs CT; University of Tennessee, Knoxville TN; Washington State University, Pullman WA



## What is TreeGenes?

**TreeGenes is a database serving the needs of the forest tree research community, bringing uniting data from nearly 1800 species with visualization and analysis tools**

TreeGenes is built using Tripal, a toolkit for creating genetic and genomic websites. Built using Drupal, Tripal is open source project that encourages tool sharing and collaboration. Developers can easily share extension modules that build on the Tripal framework to offer additional functionality. Extension modules used by TreeGenes include:

**Tripal ElasticSearch:** implements ElasticSearch and allows Tripal sites to connect their searches, enabling cross site querying.

**Tripal Galaxy:** integrates access to Galaxy workflows from within a Tripal site.



Poster  
PE0048

## Data

1850 Species  
39 Genomes  
101 Genetic maps  
36,730,176 Genotypes  
926,094 Phenotypes

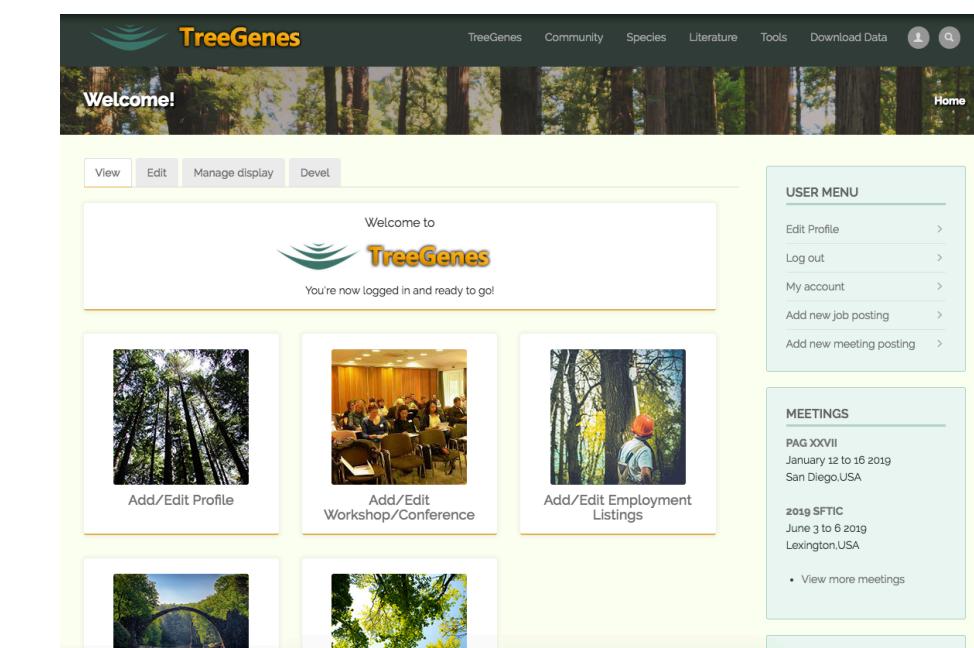
## Data Sources

- Primary repositories** including NCBI and Dryad:
- Full genome assemblies and annotations
  - Transcriptomes
  - Genetic markers, including SNP, EST, SSR, RFLP
  - Links out to raw sequence records including SRA and TSA

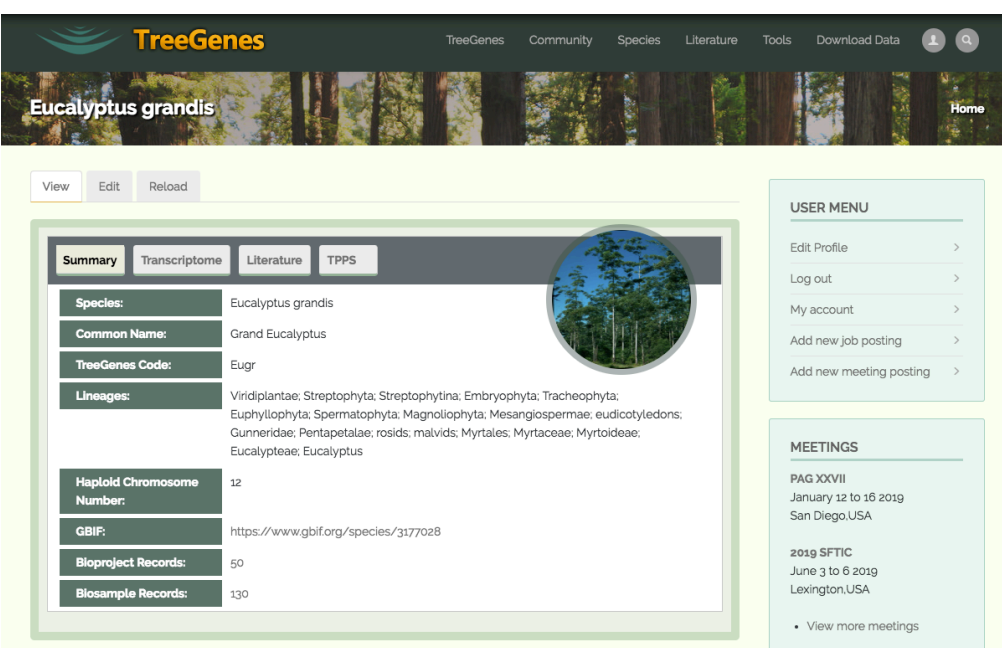
The **Tripal Plant Popgen Submit (TPPS)** pipeline accepts genetic maps and data from association and population studies, with an emphasis on capturing metadata

TreeGenes curated UniGenes

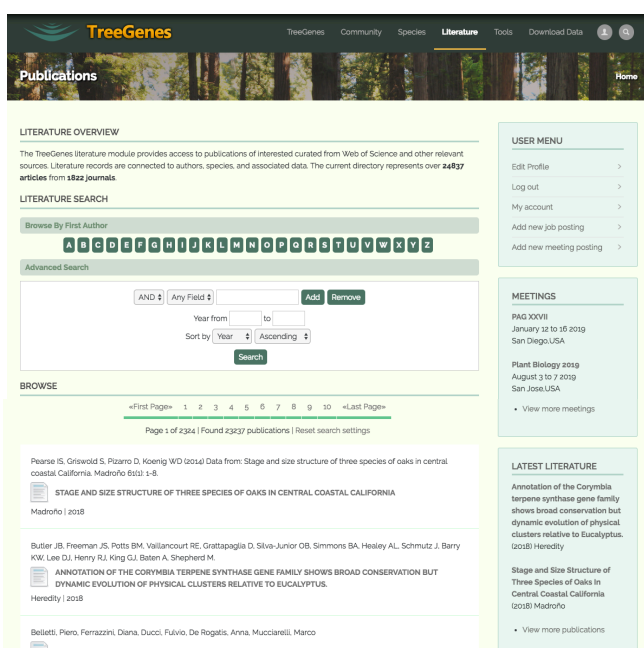
Community data, including a colleague directory, job and meeting postings



TreeGenes user accounts now offer additional functionality: Submit datasets directly to TreeGenes, view and edit past submissions. View and download results from a Diamond sequence search or Galaxy analysis

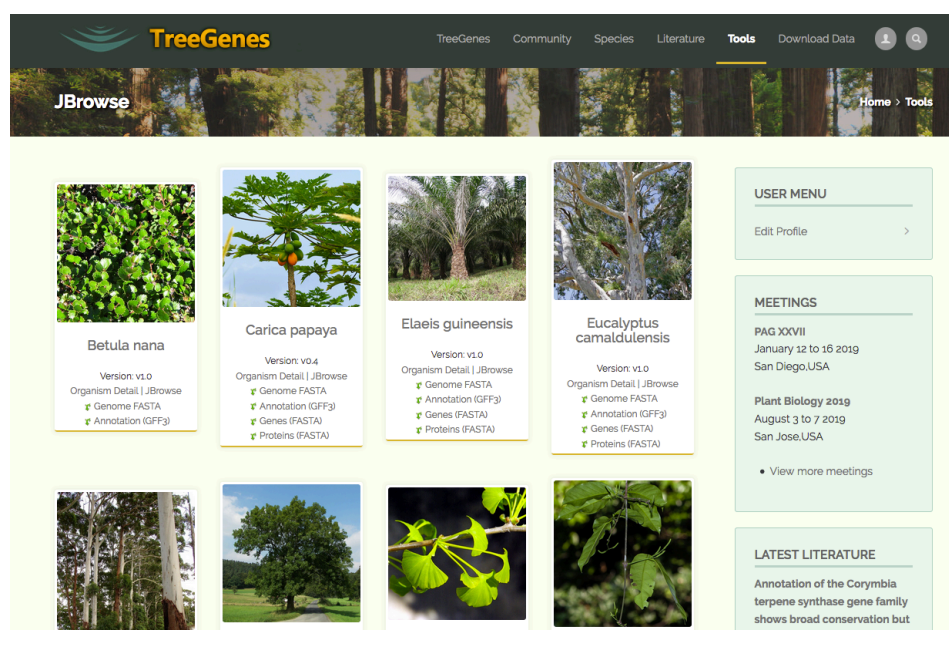


Browse species or search by taxonomic rank. Each species detail view includes a summary of available literature, genotypic, and phenotypic data as well as range maps and species images.

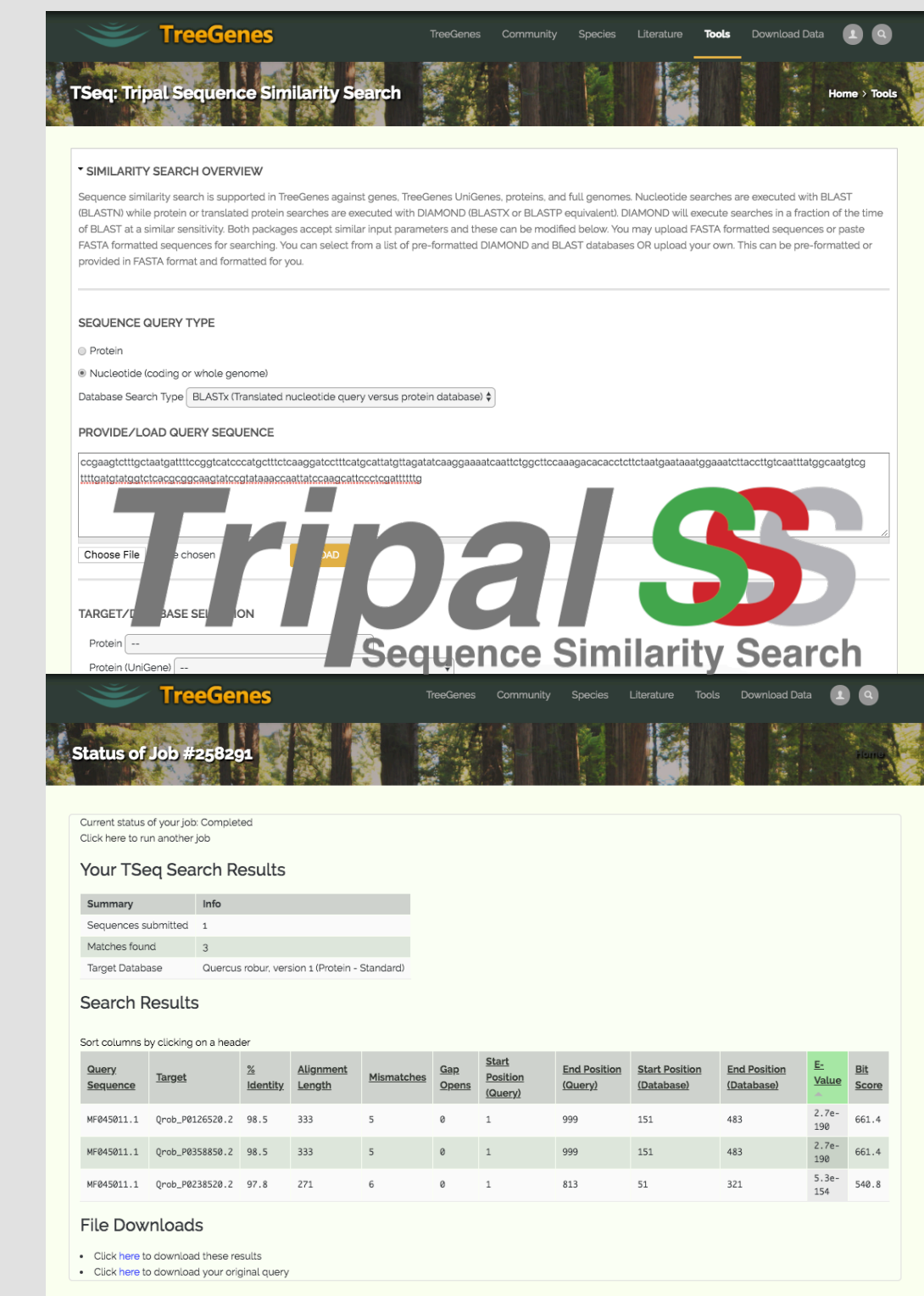


The literature module holds over 25,000 curated articles.

TreeGenes uses cmap-js to display genetic maps.



## TSeq



TSeq is a sequence alignment tool that utilizes a combination of BLAST (nucleotide) and Diamond (protein/x) search functions for faster sequence similarity searching. Execution of the analysis can be performed on local or remote servers.

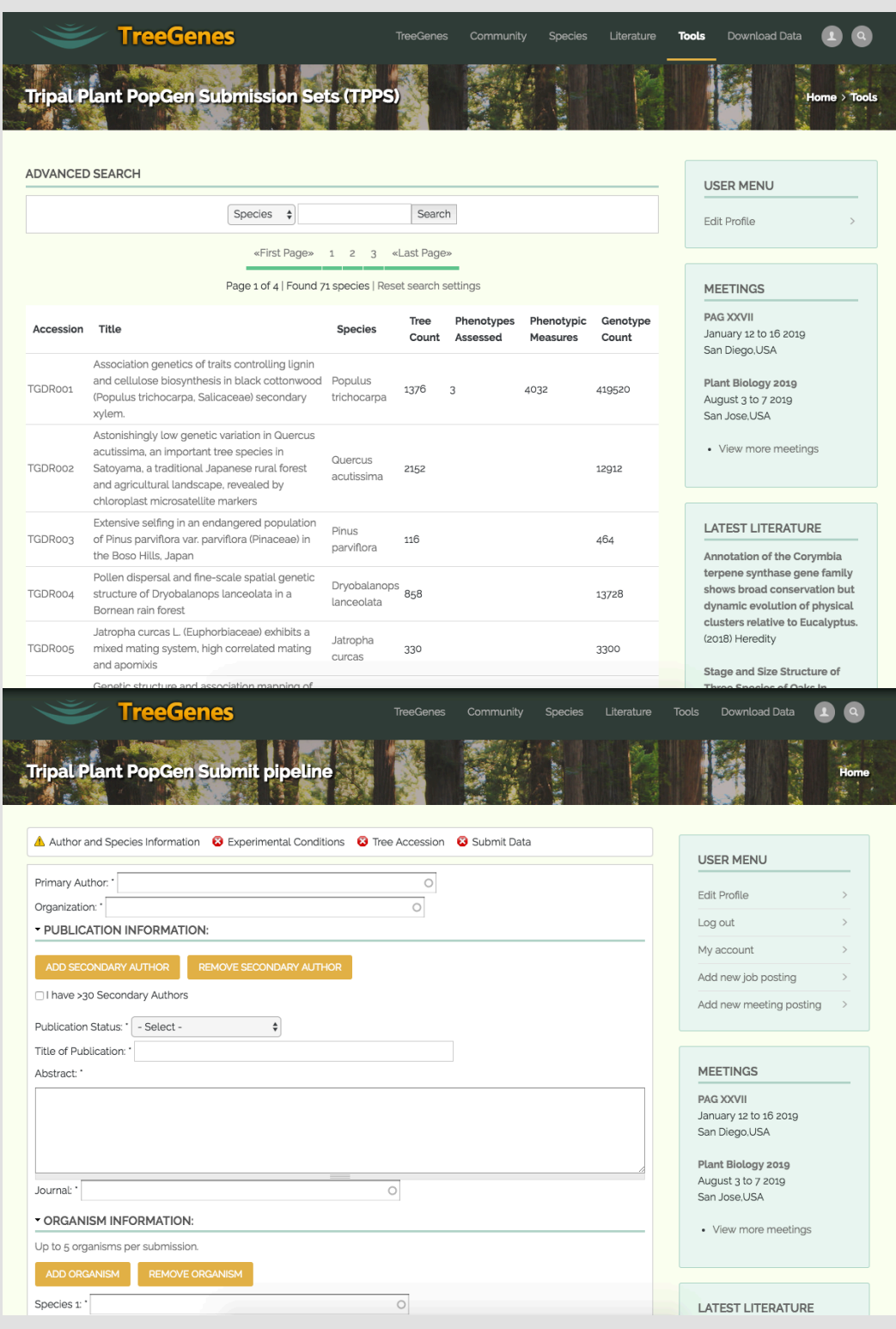
## Tripal Plant Popgen Submit

The Tripal Popgen Pipeline Submit (TPPS) module is a data intake pipeline module optimized for metadata capture of association and population studies. The redesign captures additional experimental parameters while also easing the process for users.

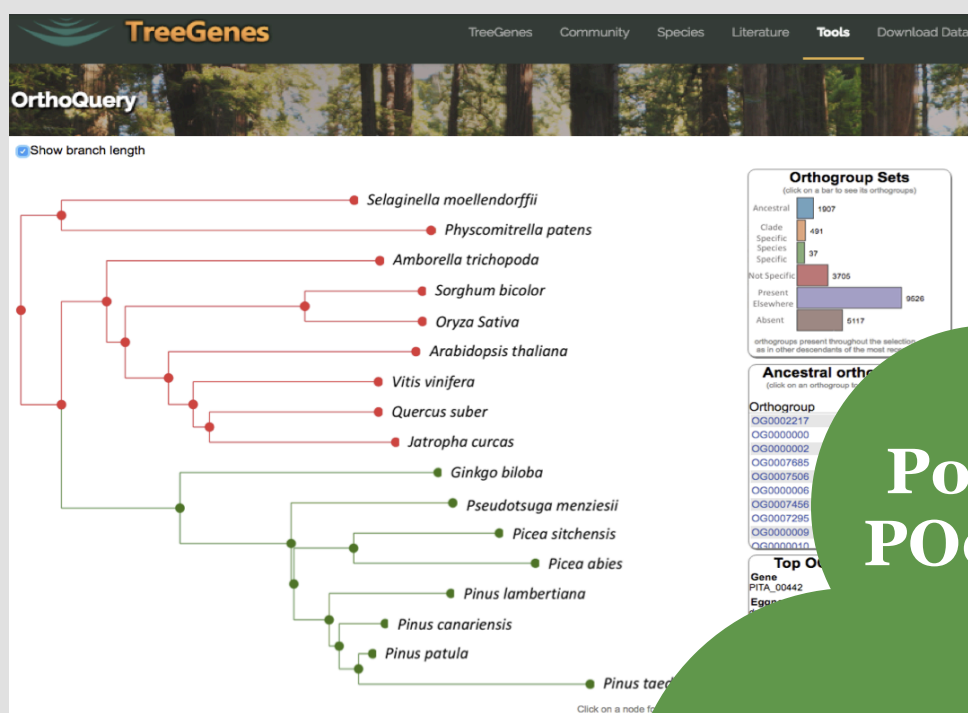


Captured parameters include:

- Sample location data
- NCBI accession identifiers for markers and raw sequence data
- Plant Ontology, Trait Ontology, and Crop Ontology term assignment for phenotypes
- Study metadata including study type experimental parameters,; complies with Minimum Information About a Plant Phenotyping Experiment (MIAPPE) standards



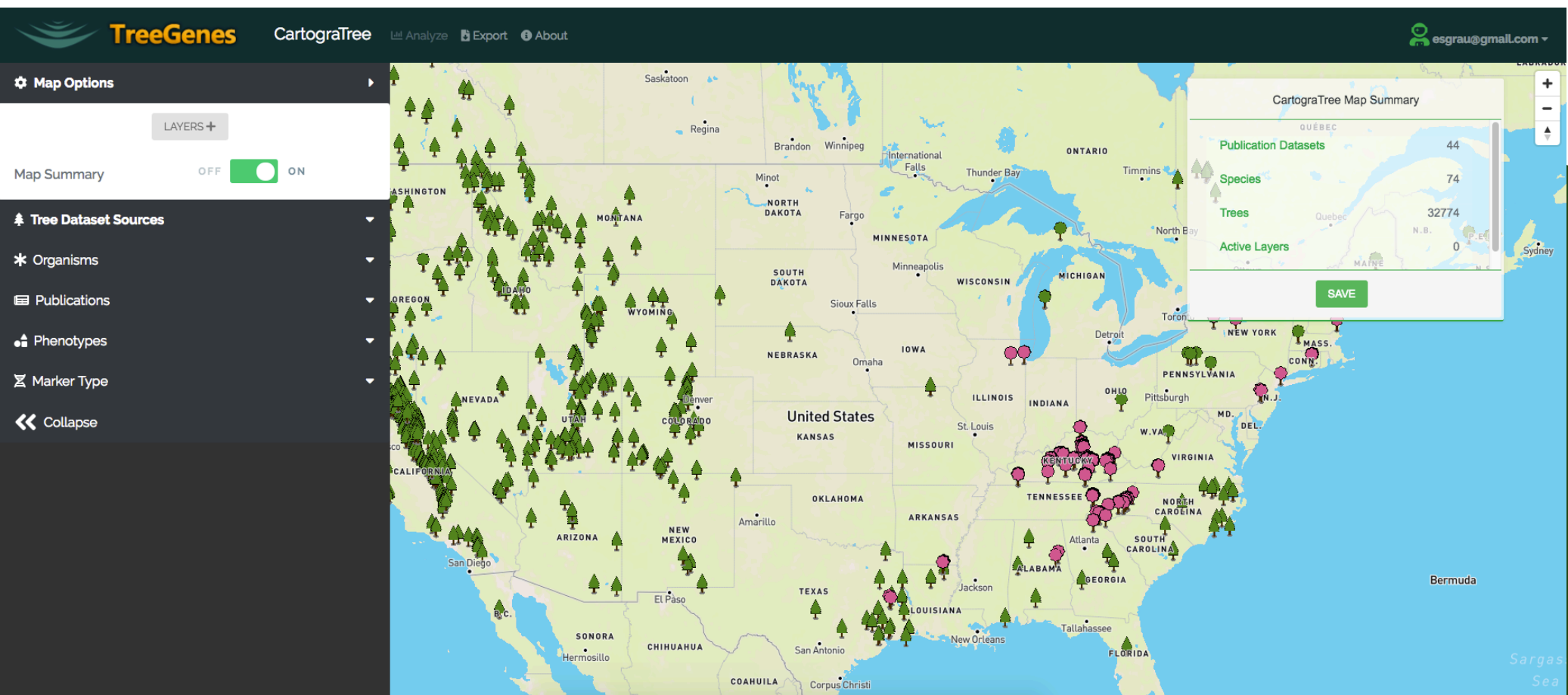
## OrthoQuery



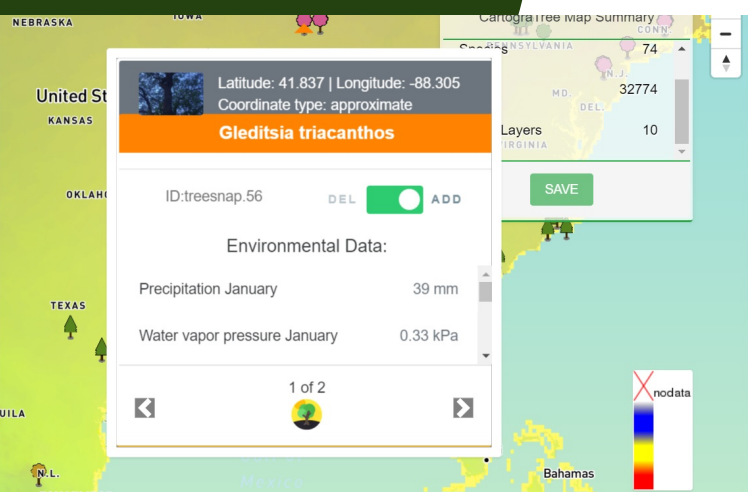
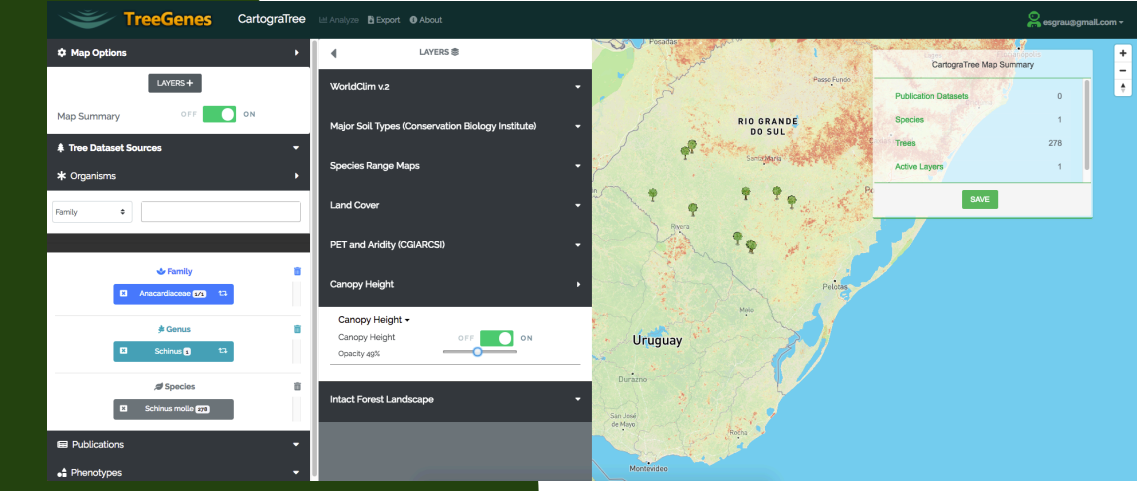
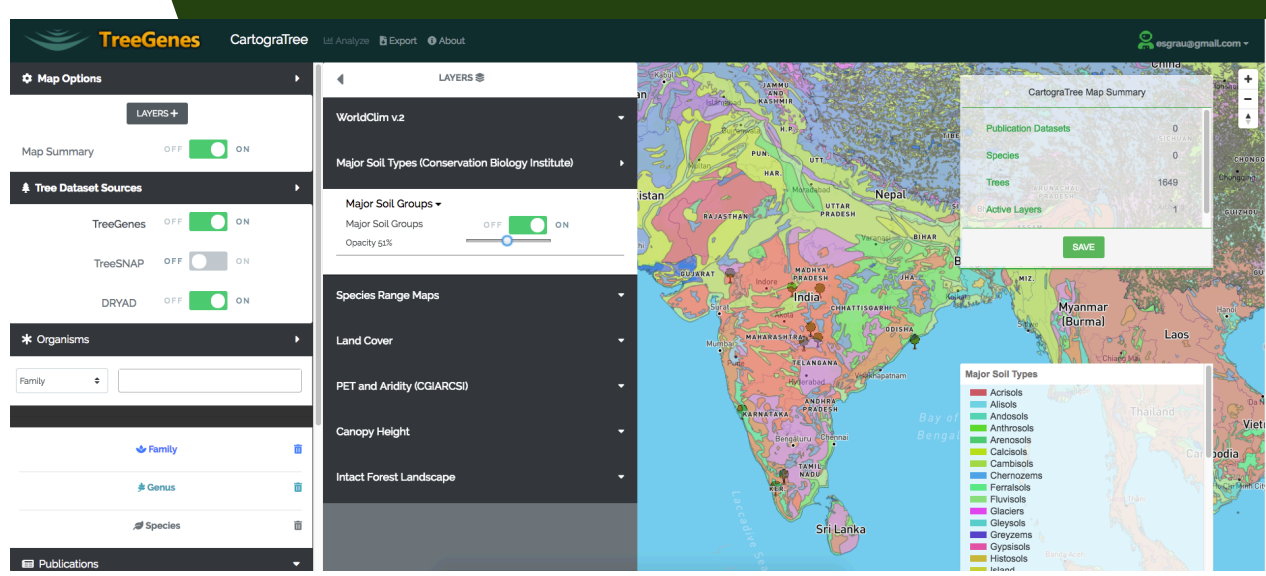
Poster  
PO0021  
Computer  
demo  
C25

OrthoQuery provides a semi-automated analytical framework to enable comparisons among curated proteins and interactive visualizations in context of the resulting species tree. Leverages Orthofinder optimized with Diamond as the tool for protein level comparisons, and Tripal framework with Galaxy integration to support visualization and query.

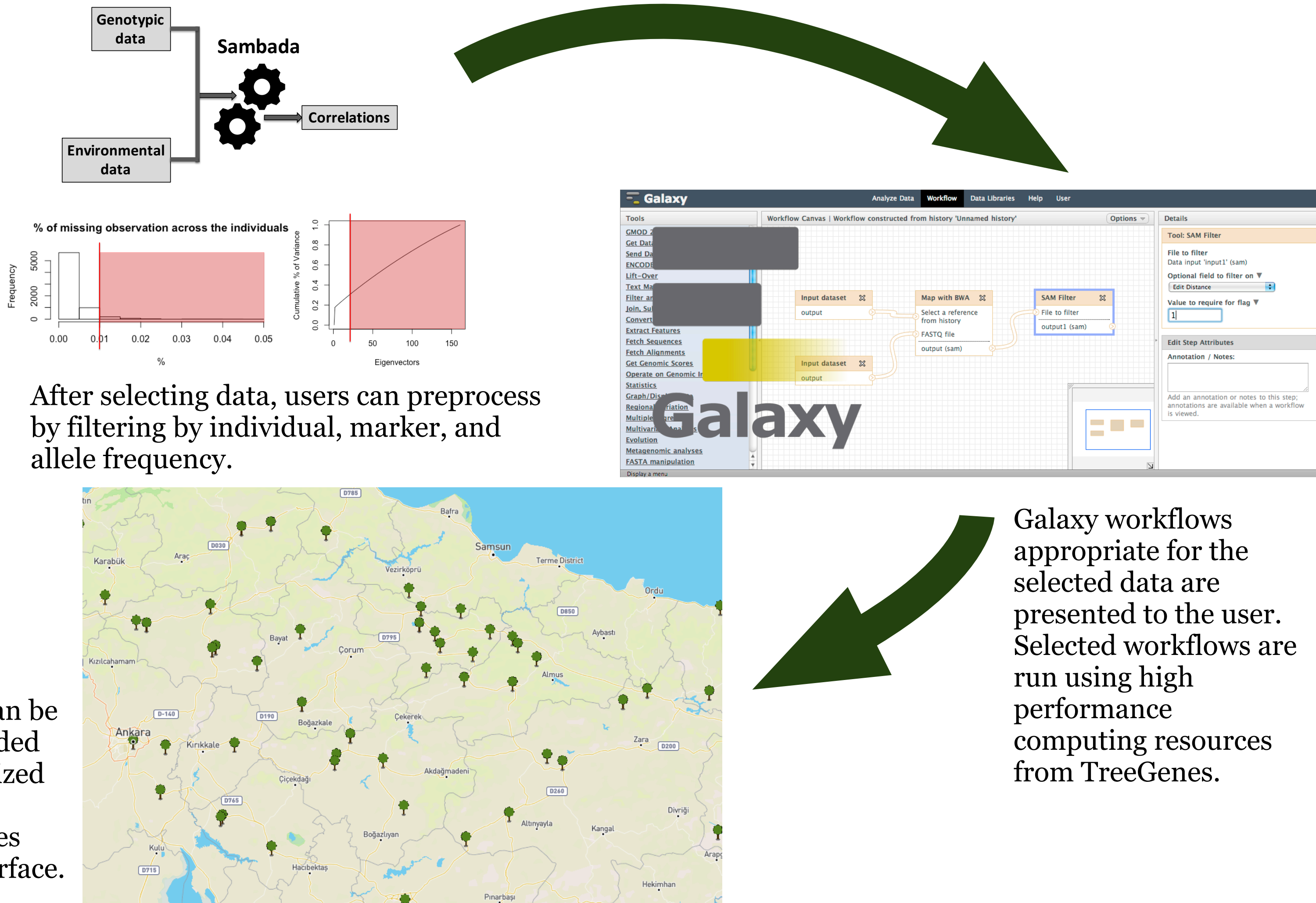
## CartograTree



CartograTree is a map based web application that brings together genetic, phenotypic, and environmental data from TreeGenes, Dryad, WorldClim, CGIAR-CSI, and others to enable analysis.



Analysis results can be downloaded or visualized via the TreeGenes user interface.



Galaxy workflows appropriate for the selected data are presented to the user. Selected workflows are run using high performance computing resources from TreeGenes.

treegenesdb.org

treegenesdb@gmail.com

@treegenes

Linked in treegenes database

UConn

Tripal

Galaxy

GMOD

USDA

NSF