Cyberinfrastructure for Landscape Genomics: Connecting Biological Databases, Metadata, and Intelligent Analytics

Jill Wegrzyn

Plant and Animal Genome Conference

January 12th 2019

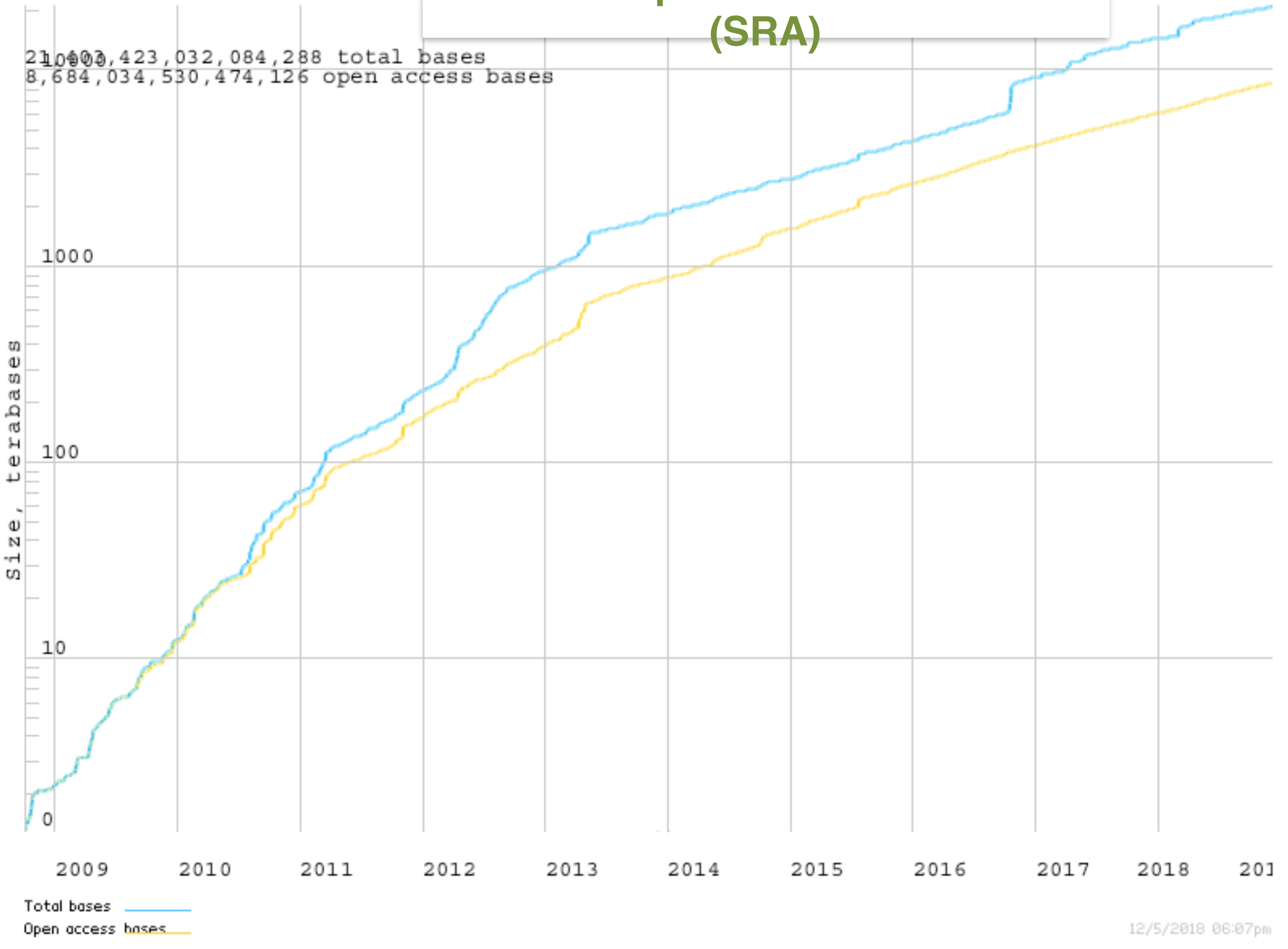
Biologists need more than ‘omics!

Early drivers of ‘omics: Genomics

Next Generation Sequencers

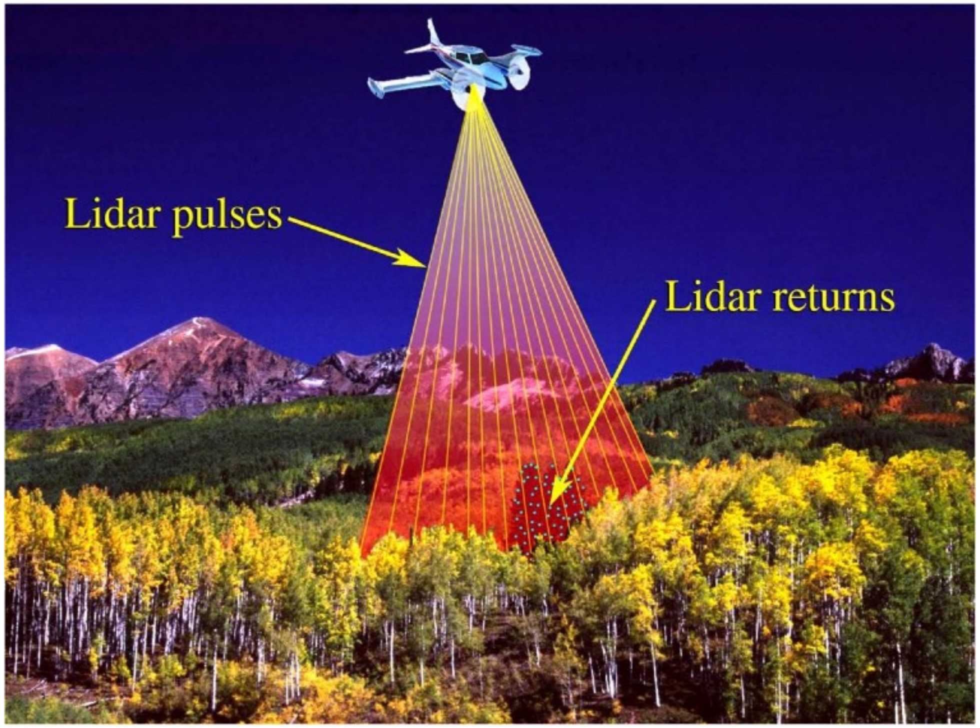
High Throughput Sequencing

NCBI Sequence Read Archive (SRA)



Phenomics

Environmental-omics?
Remote Sensing - LIDAR



Active and passive methods

Monitor the impacts climate change, manage natural resources, and assess research plots

- Shoreline changes
- Ocean temperature
- Soil composition/Sediment transport
- Forest canopy
- Species composition
  - Biodiversity/invasive species

Citizen Science: non-expert contributions

budburst
a project of the Chicago Botanic Garden

My Account

About Budburst | Get Started | Plants | Projects | Data

Photo courtesy of Jane E. Ogilvie.

- Learn How to Observe
- Observe Your Plant
- Report Your Observation
- Download Data

Get Started With Budburst!

Budburst is a national network of citizen scientists monitoring plants as the seasons change. To join, follow these steps: Learn how to observe, Make an observation, and Report your observation.

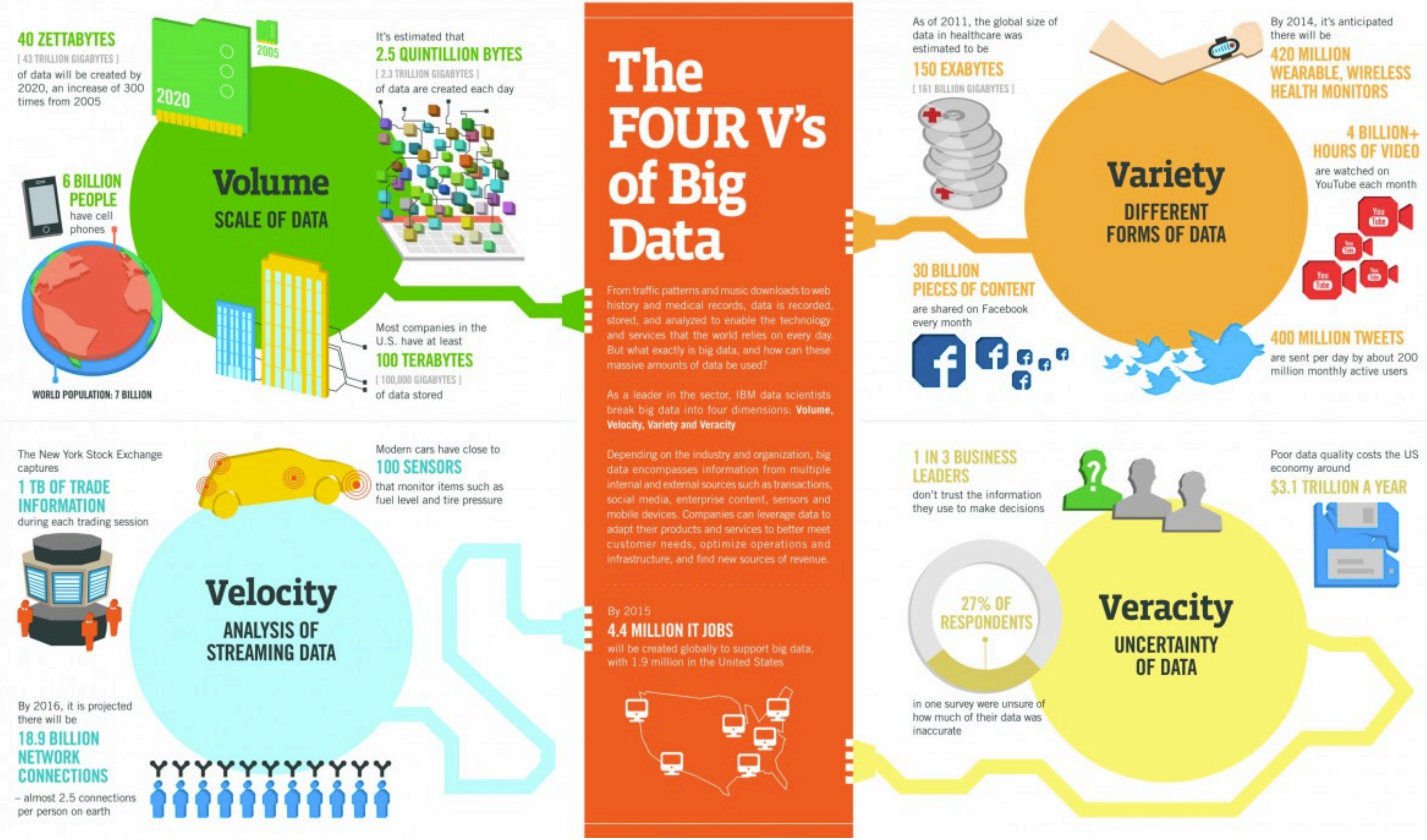
Learn How to

Watching plants and re
of a plant during the gr
including leafing, flowe
observers benefit from

Citizen Science and Climate Change: Mapping the Range Expansions of Native and Exotic Plants with the Mobile App Leafsnap FREE

W John Kress, Carlos Garcia-Robledo, João V B Soares, David Jacobs, Katharine Wilson, Ida C Lopez, Peter N Belhumeur

Challenges of Integratomics: 'Big Data' explained by Data Science

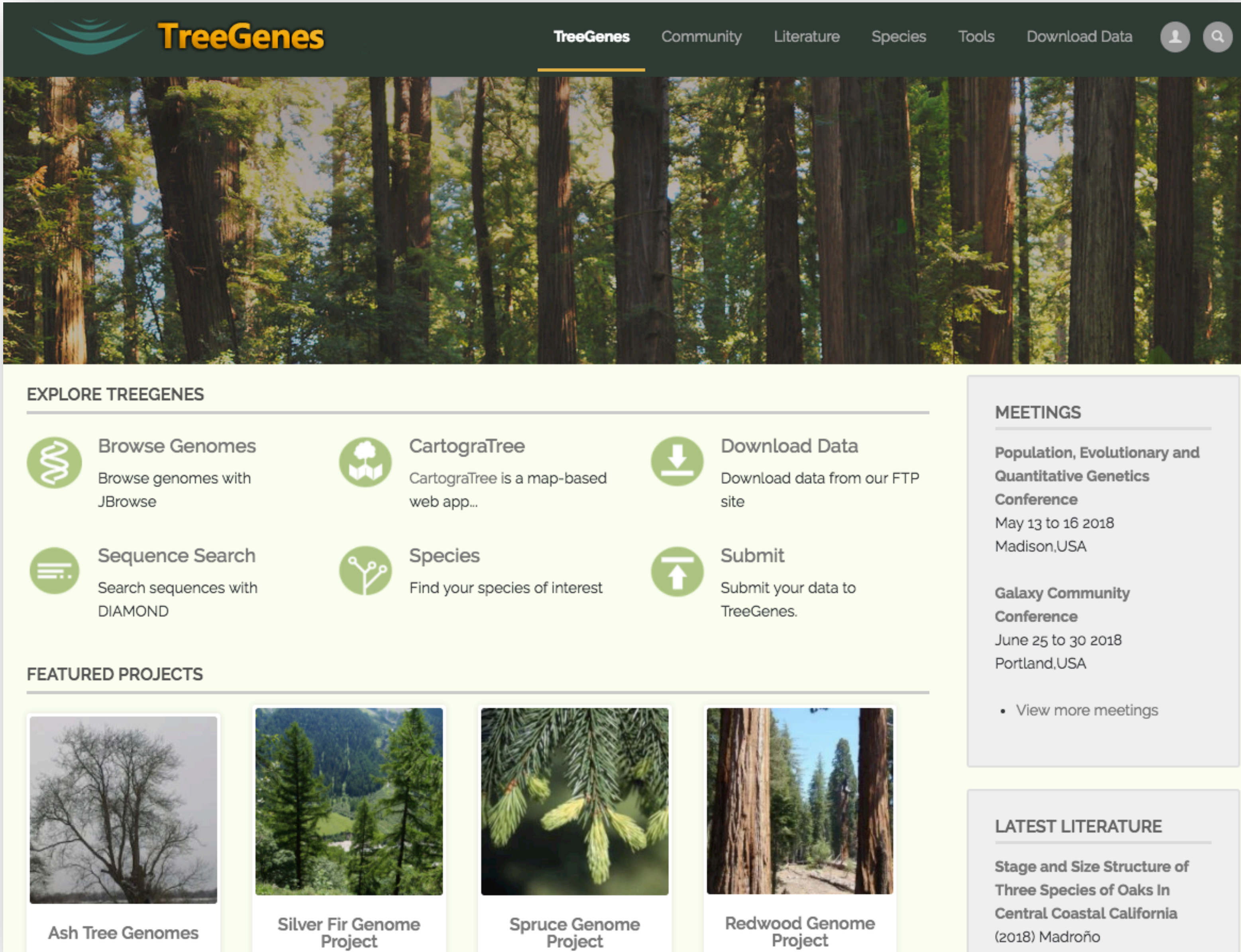


Challenges of Integratomics

Forest Tree Research Community Survey

Improving infrastructure for tree genomic and phenomic data

TreeGenes Database



TreeGenes Database: Species

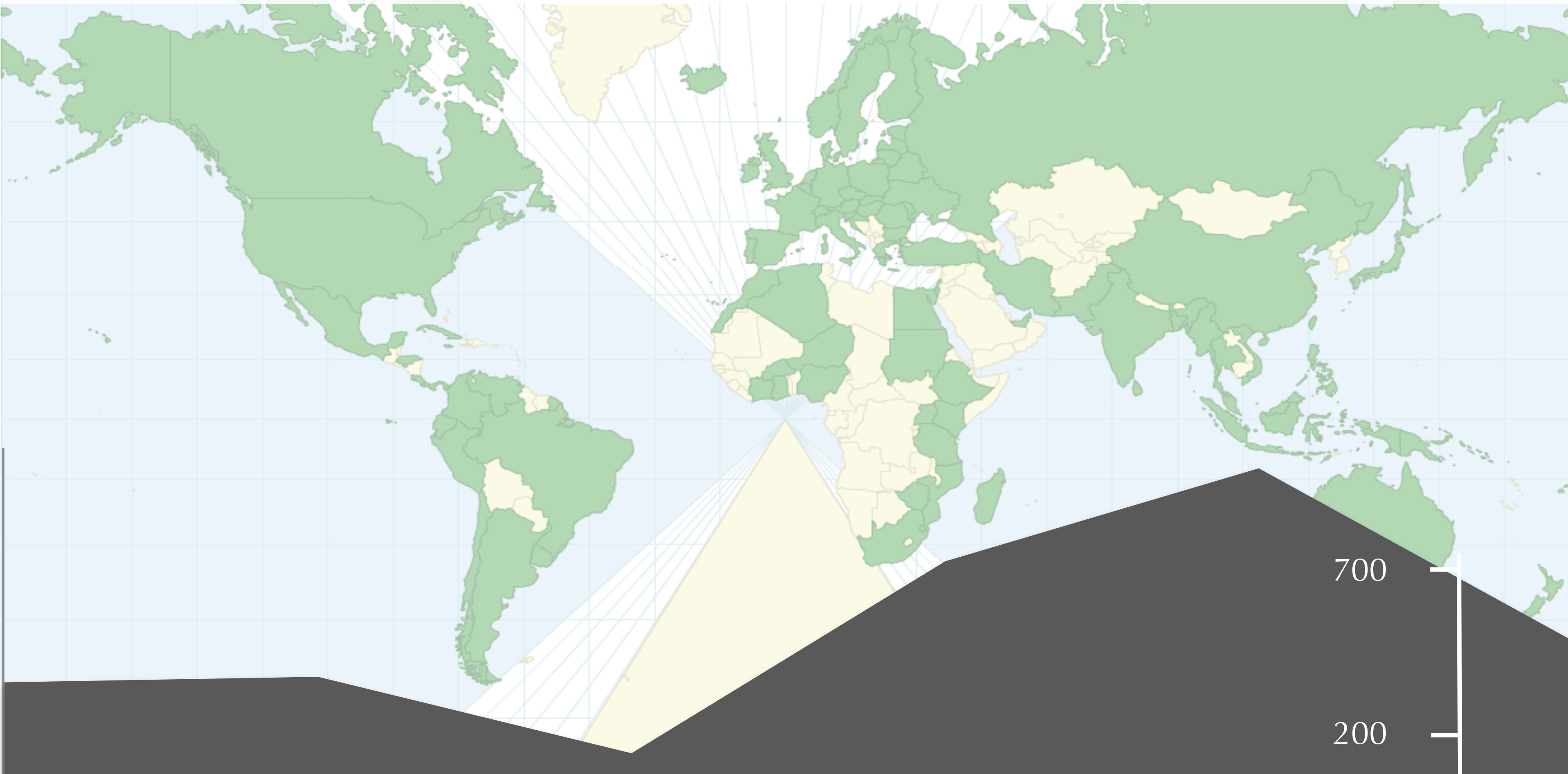
treegenesdb.org

- 1,701 species from 112 genera
  - At least one genetic artifact from each species
- Full genome sequence: 25 species
- Transcriptome/Expression resources: 6,920,817 sequences from 322 species
- 108 genetic maps from 37 species
- Population studies
  - Georeferenced trees
  - Extensive genotypic (GBS and array) and phenotypic data

TreeGenes Database: Users

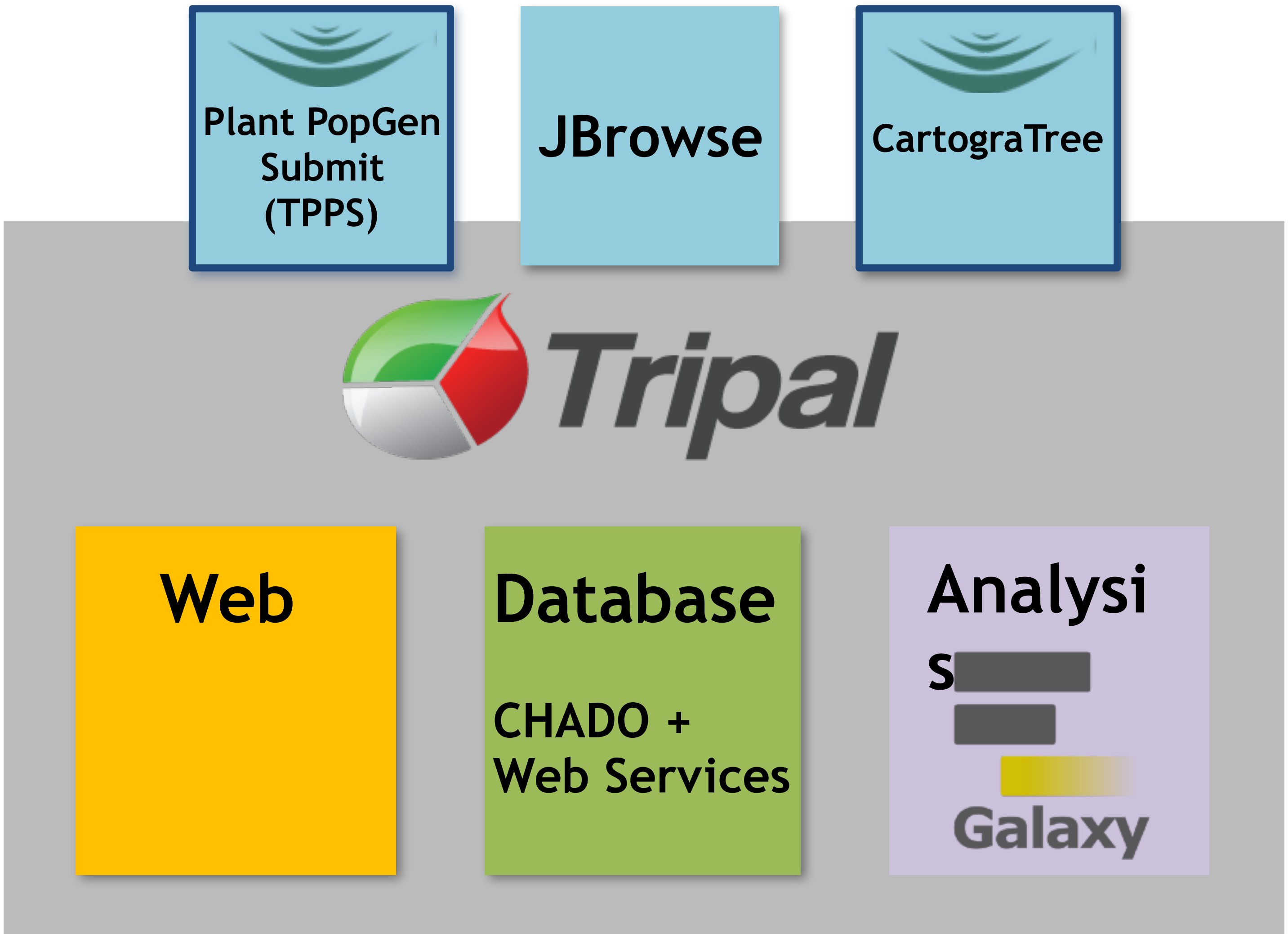
treegenesdb.org

3,100 unique visitors from 116 countries

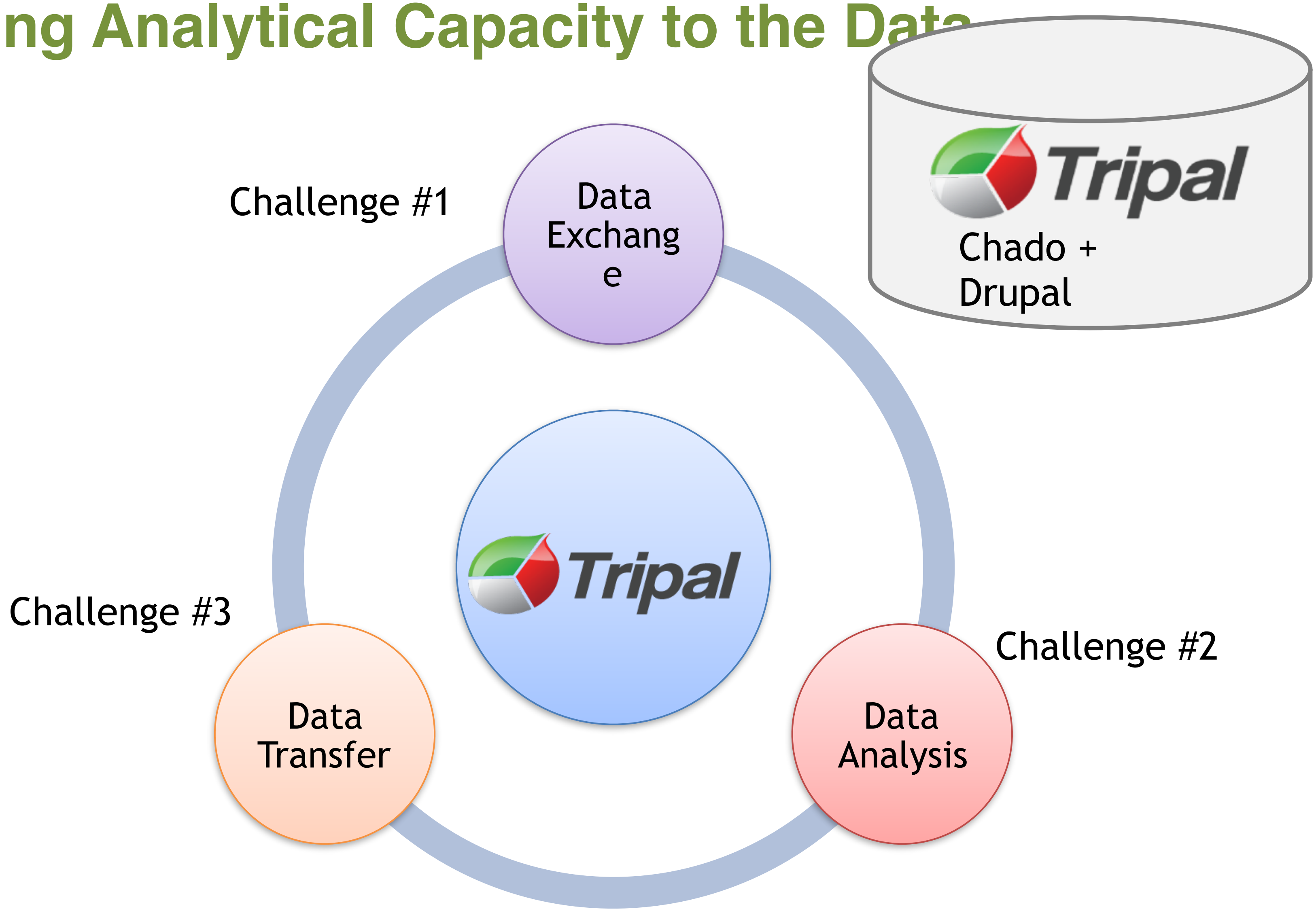


Unique Web Visitors to TreeGenes (by month), July-December 2018

Tripal Framework in TreeGenes



Tripal Gateway Project:
Bringing Analytical Capacity to the Data



TreeGenes hosts georeferenced plants that can be integrated with environmental metrics

Scientists require this integration (GxPxE)

- What genotypes contribute to traits related to timber production?
- What genotypes are most adapted to specific elevations/climates for reforestation?
- What genotypes are most resistant to invasive and native pests and pathogens?
- What individuals are best suited for migration within their range in the face of a changing climate?

Tripal Plant PopGen Submit (TPPS)

Metadata collection remains sparse and incomplete

- Long-term accessions (and storage)
- Integration with existing ontological frameworks
- Standards related to data collection
- Integration with primary repositories
- Focus on capturing georeferenced data

Tripal Plant PopGen Submit (TPPS)

Population Study

- Publication
- Species

Study Design

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

Phenotype, Genotype, Environment

- Georeferenced

Raw Data

Minimal Information About a Plant Phenotyping Experiment (MIAPPE)

miappe

About | Links | Publications | People | Data | Projects

MIAPPE

Minimum Information About a Plant Phenotyping Experiment

MIAPPE is an open, community driven project to harmonize data from plant phenotyping experiments. MIAPPE comprises both a conceptual checklist of metadata required to adequately describe a plant phenotyping experiment, and software to validate, store and disseminate MIAPPE-compliant data. We welcome contributions from anyone interested in plant 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.

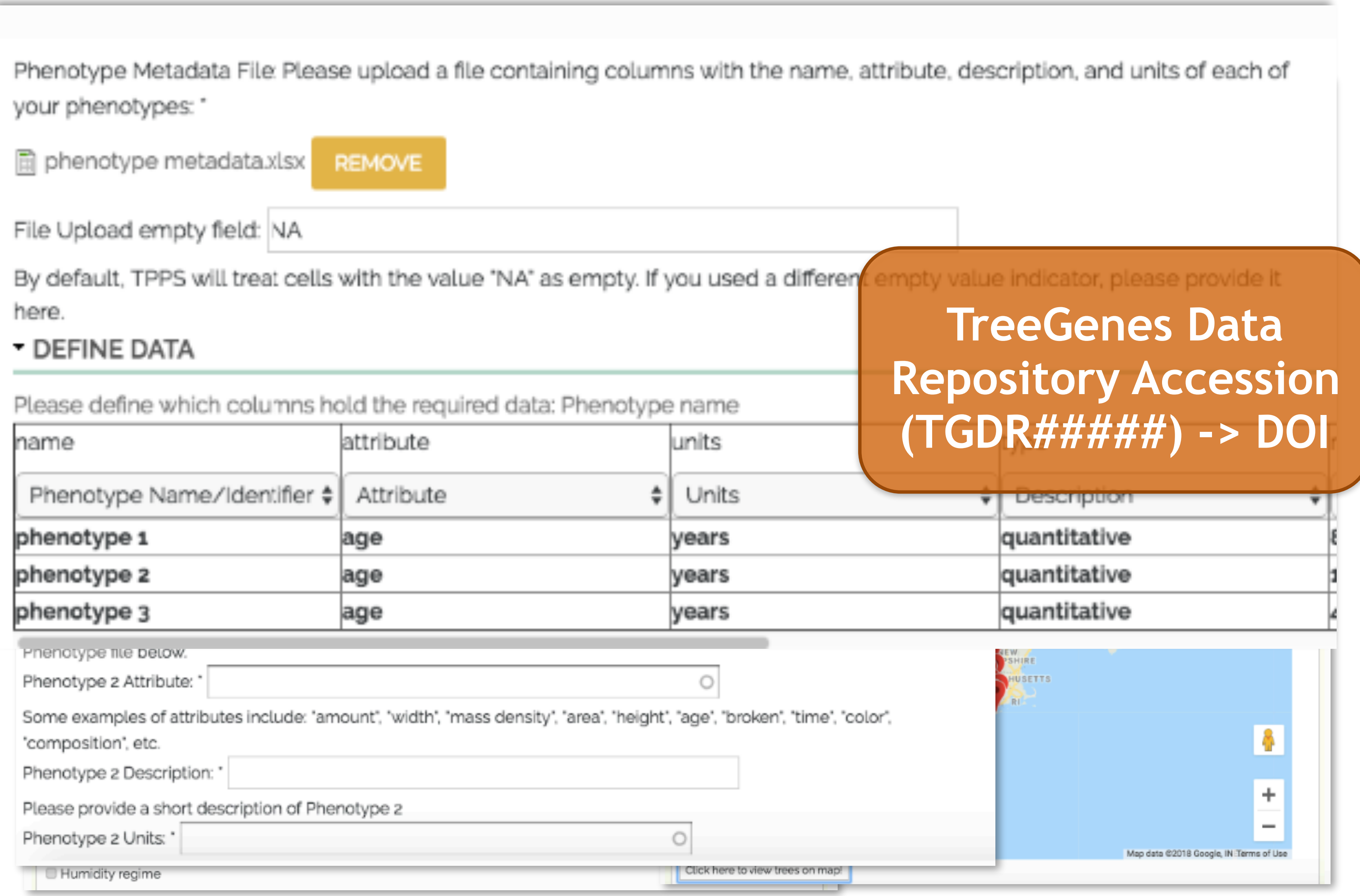
MIAPPE v1.1

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 Ontology (ENVO)
- Chemical Ontology (CHEBI)

Tripal Plant PopGen Submit (TPPS)



TreeGenes Data Repository Accession (TGDR#####) -> DOI

FAIR

Findable

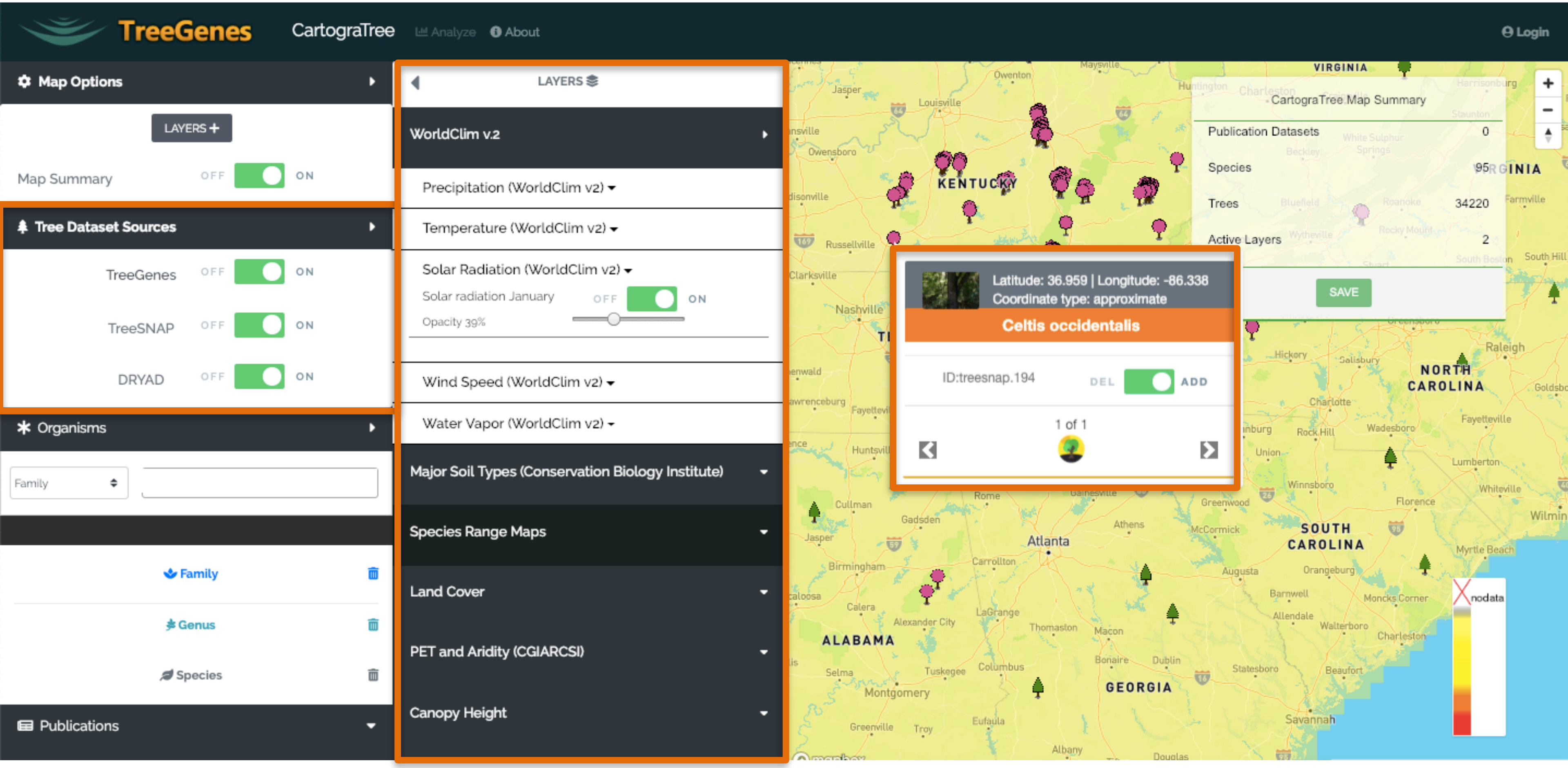
Accessible

TreeGenes to CartograTree

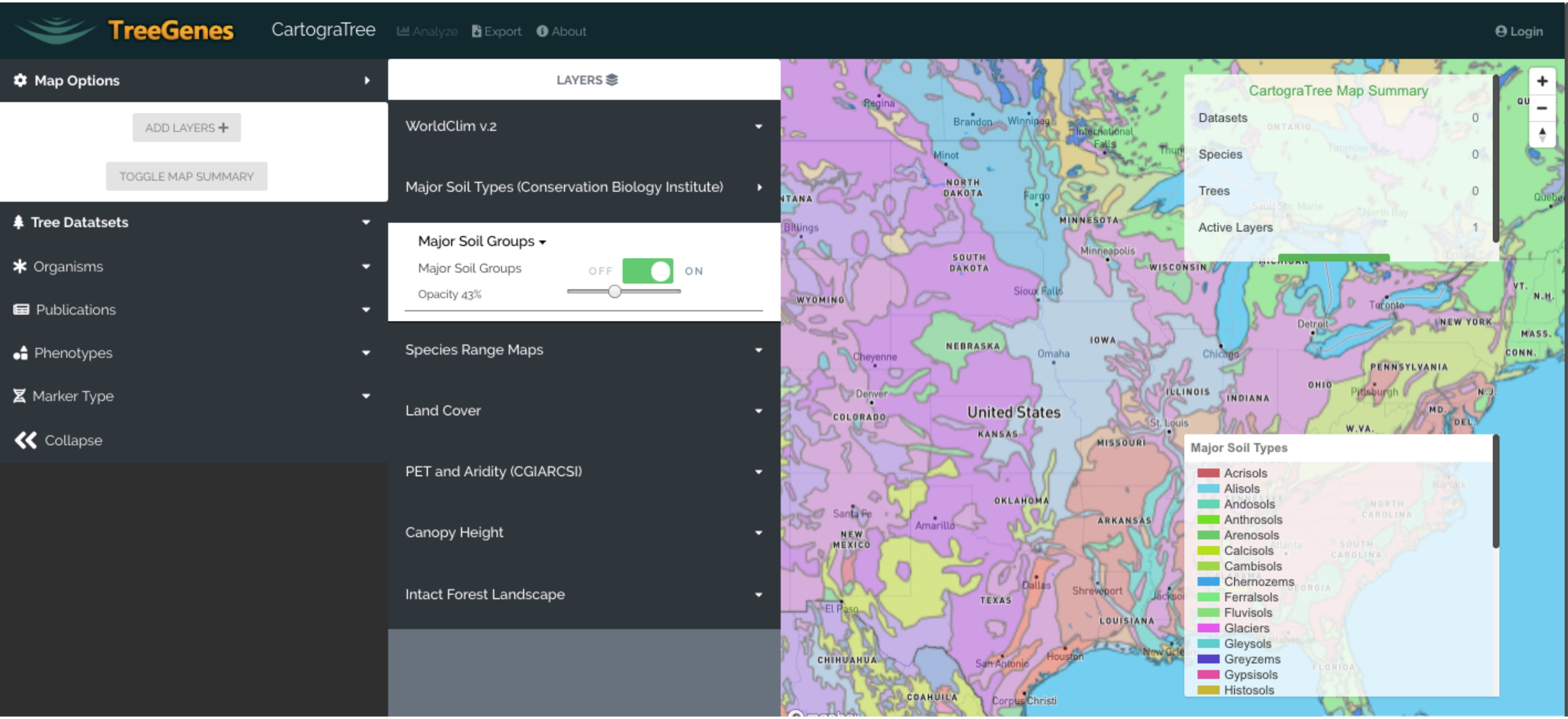
TPPS/TGDR DETAILS FOR TGDR001

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	ASSOCIATION RESULTS FILE 1 FILE GENOTYPES SNP GPS COORDINATES HAPLOTYPE DATA FILE PHENOTYPES PHENOTYPES DEFINITIONS
CartograTree	View in CartograTree
Tree Count	1376
Phenotype Count	4032
Unique Phenotypes	3

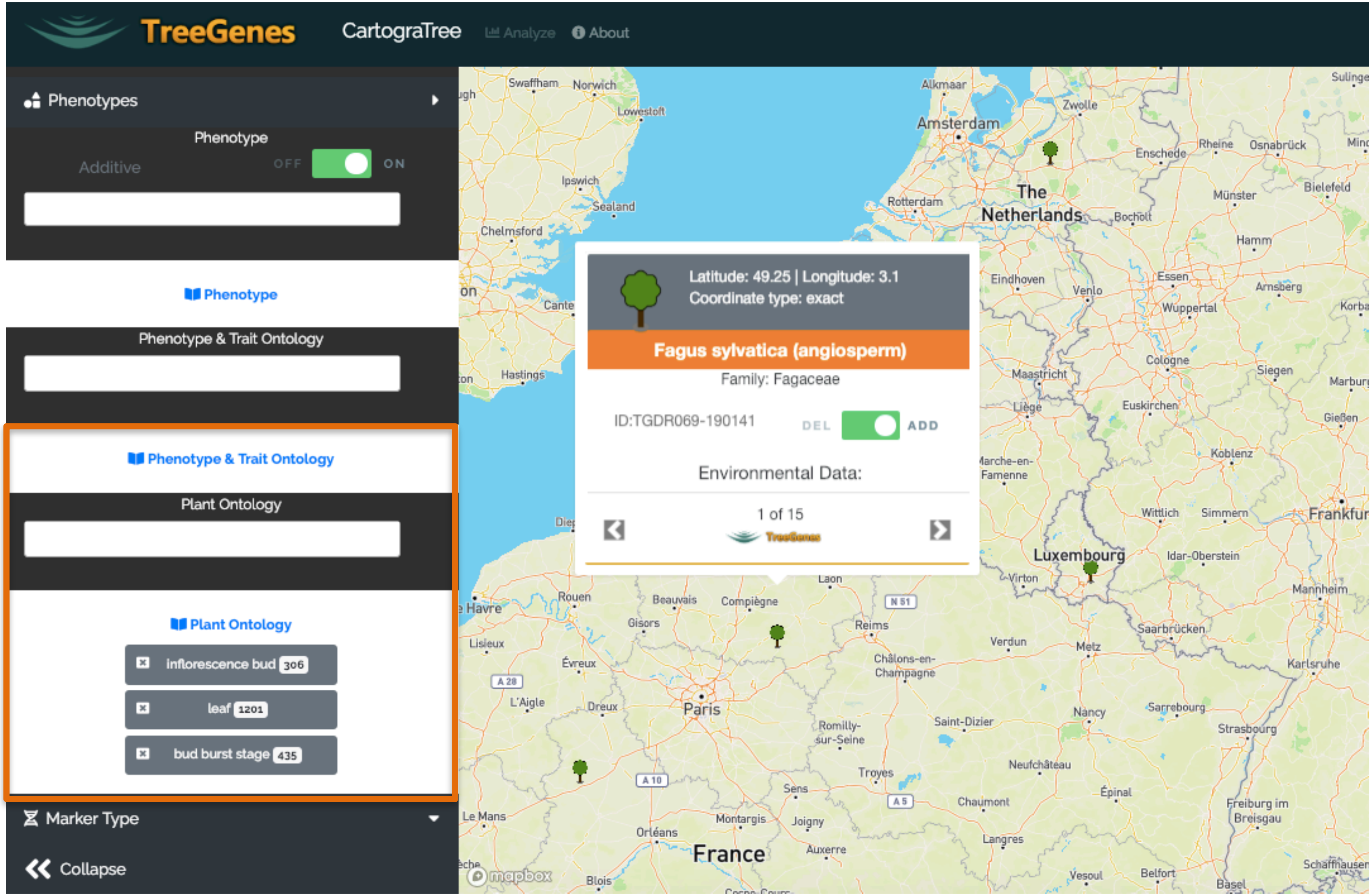
CartograTree: Integrating environmental layers with georeferenced trees



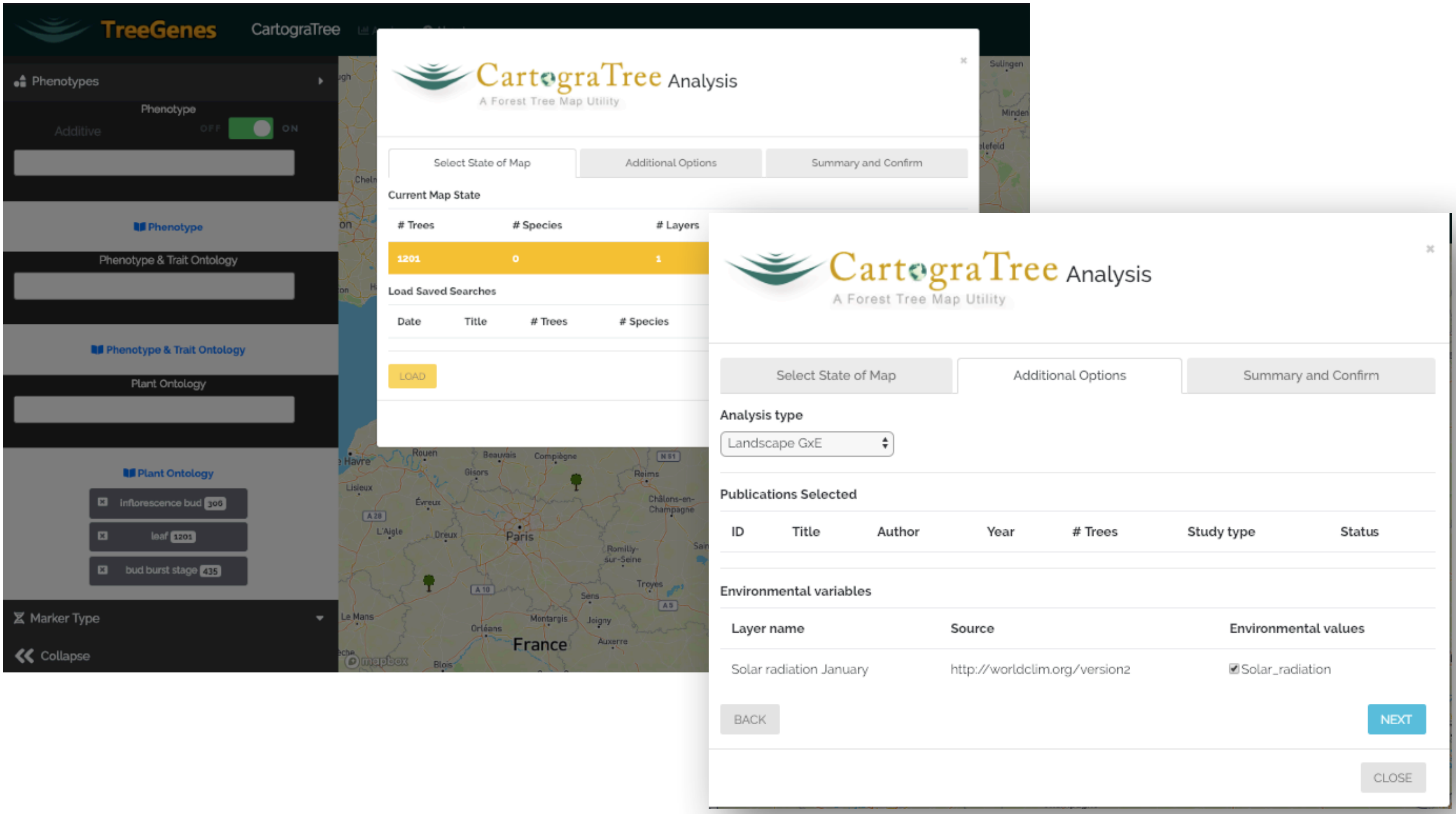
CartograTree: Integrating environmental layers with georeferenced trees



CartograTree: Integrating environmental layers with georeferenced trees

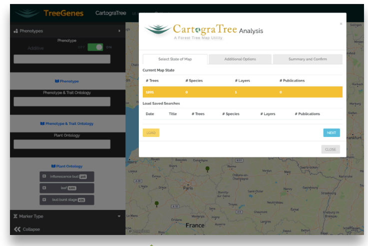


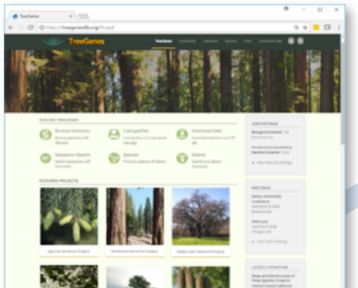
CartograTree: Save searches locally and select for meta-analysis



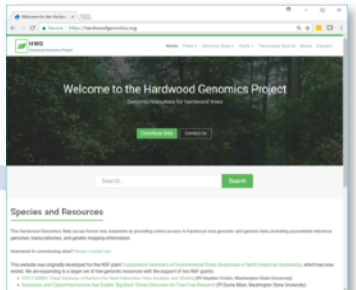
Bringing Analytical Capacity to the Data

CartograTree

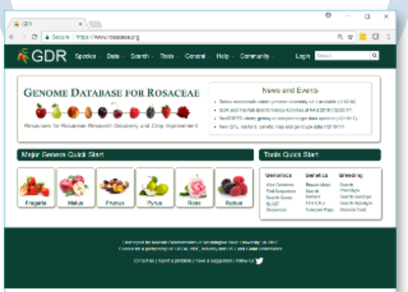




Tree Genes



Hardwood Genomics

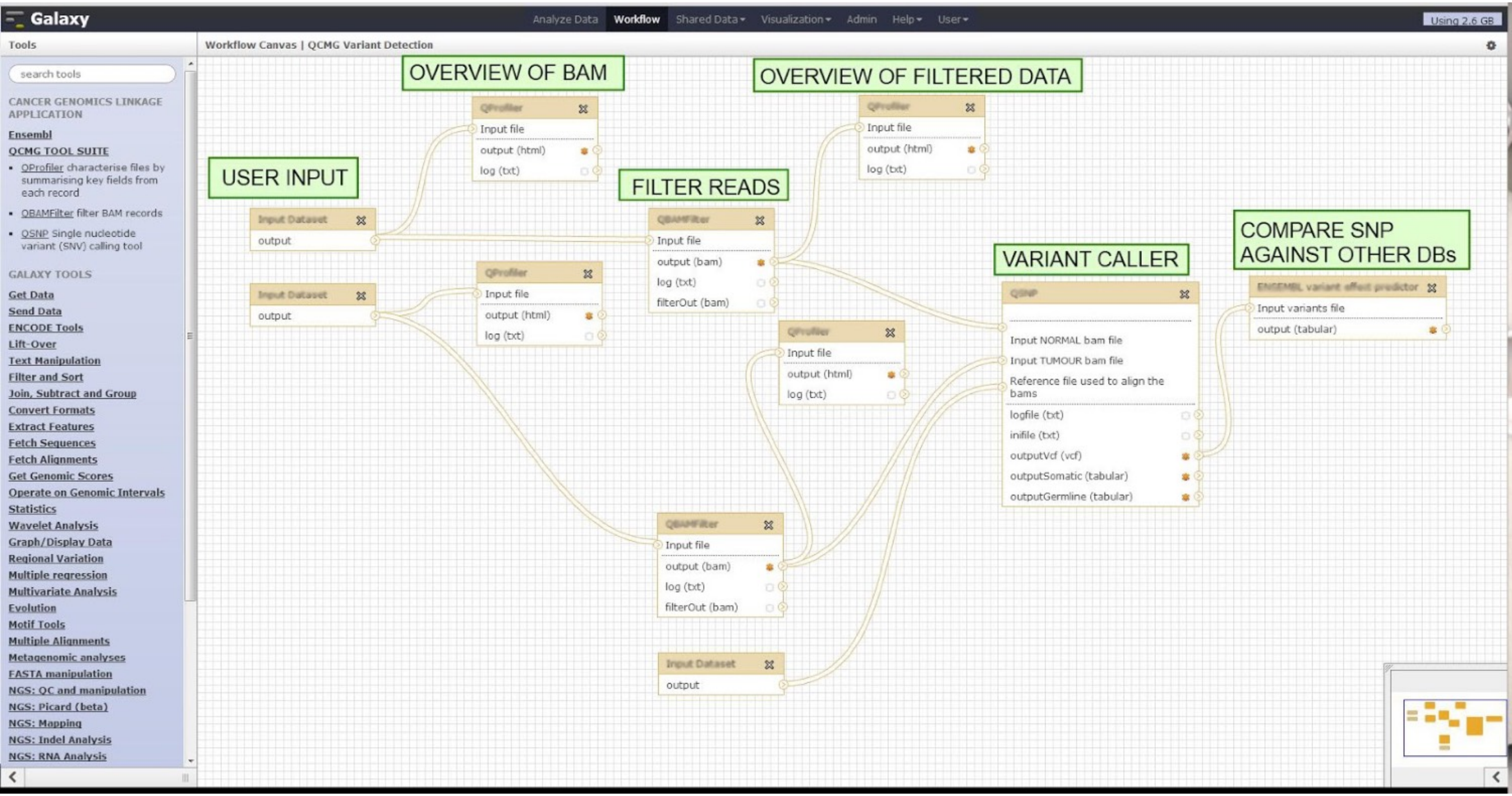


Genome Database for Rosaceae

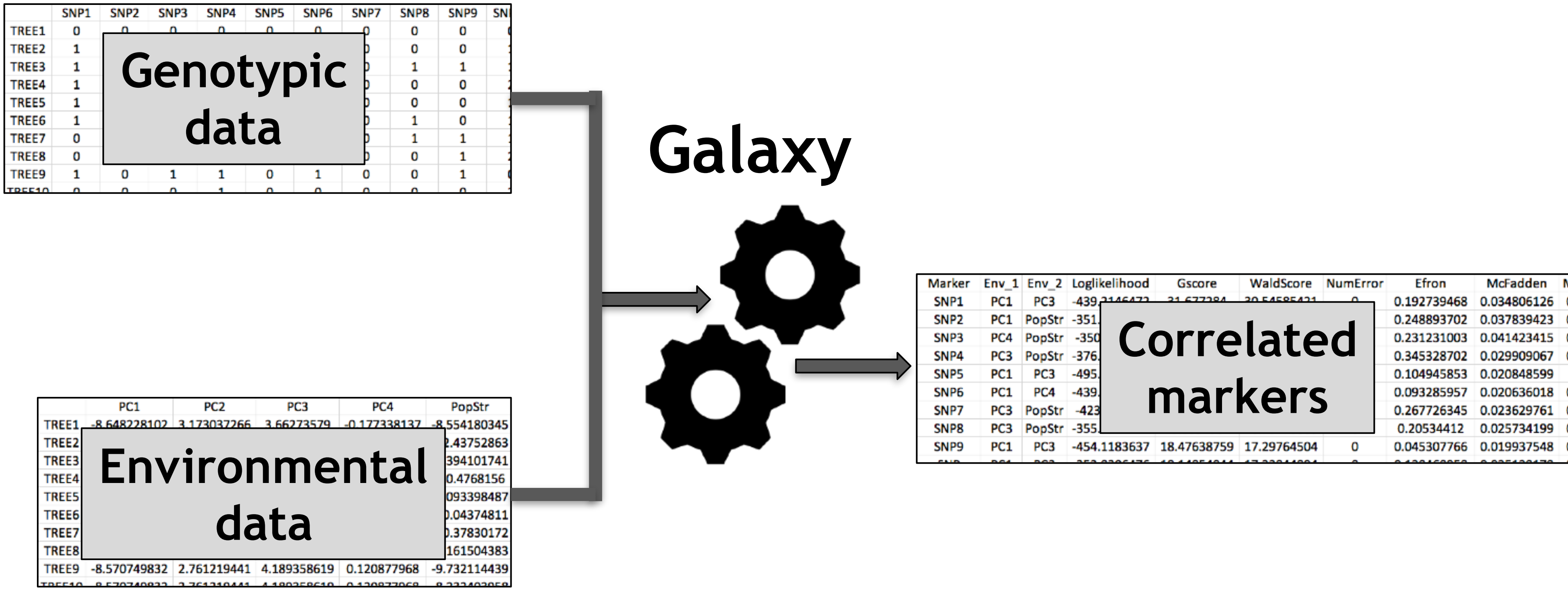
Galaxy PROJECT

Galaxy: Open Source Web-based Platform for Bioinformatic Analysis

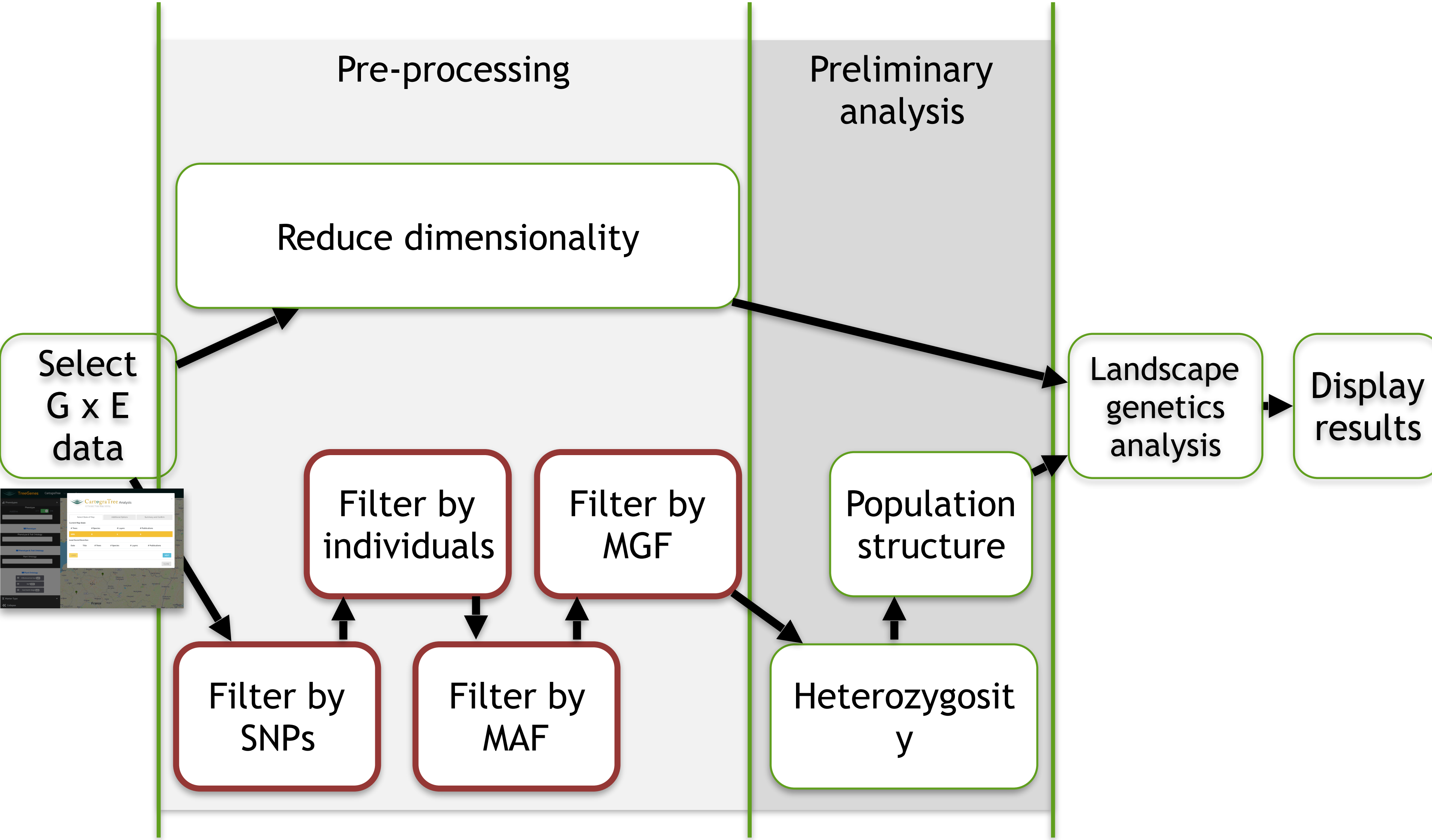
Galaxy: Open Source Web-based Platform for Bioinformatic Analysis



Workflows for Landscape Genomics: Integrating across diverse datasets

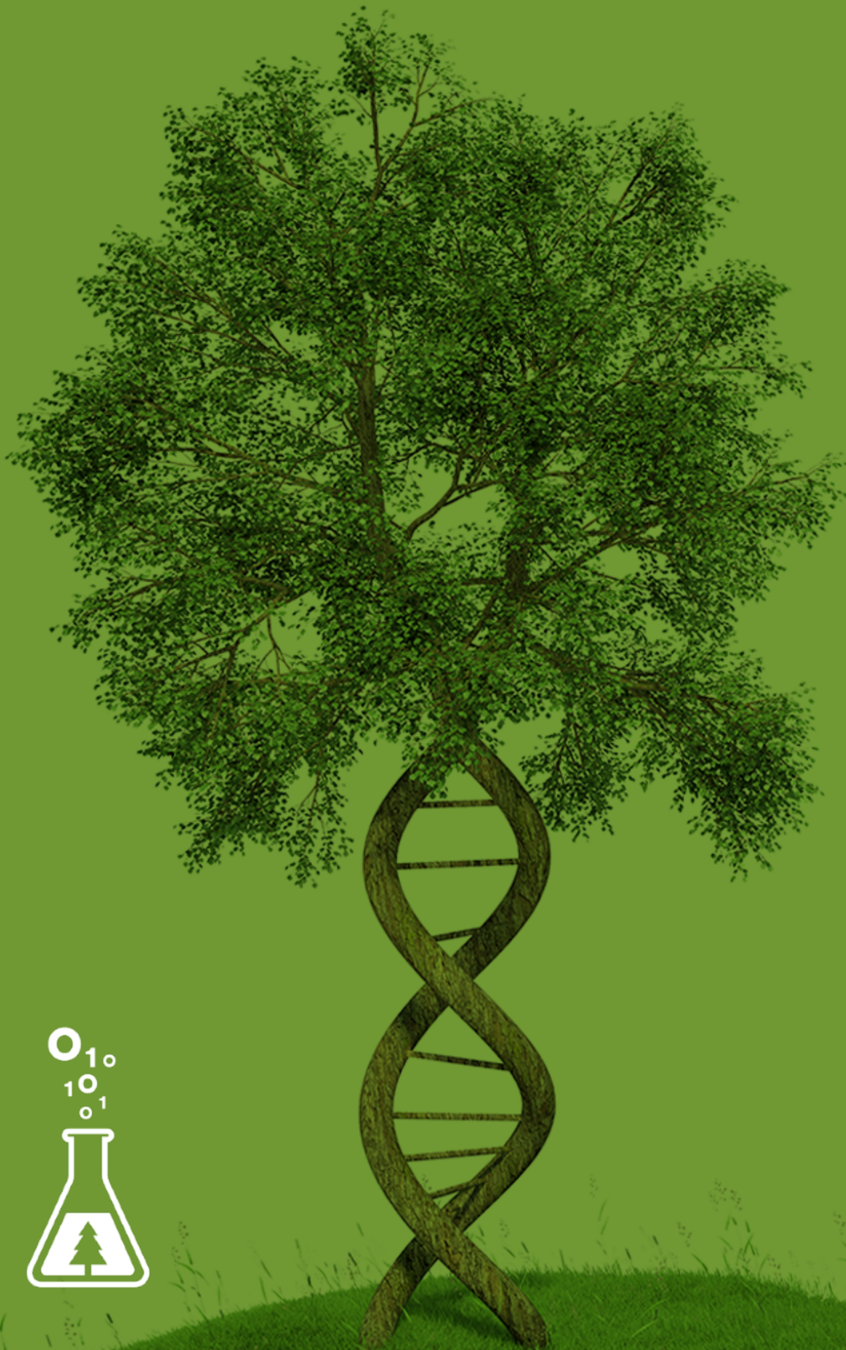


Workflows: Executed in Galaxy with metadata



Future Development:
CartograTree -> CartograPlant

- Developing models to load balance analysis (Galaxy/TACC)
- Leverage Cyverse DataStore to share and store data
- Develop more workflows (association genetics)
- Robust platform for additional plant species



Plant Computational Genomics Lab, University of Connecticut

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

Washington State University

- Stephen Ficklin
- Doreen Main

University of Tennessee

- Margaret Staton
- Ming Chen