Cartograplant: Cyberinfrastructure to Improve Plant Health and Productivity in the Context of a Changing Climate

Jill L. Wegrzyn[1], Irene Cobo-Simon[1], Risharde Ramnath[1], Emily Grau[1], Sean Buehler[1], Meghan Myles[1], Victoria Burton[1], Madison Gadomski[1], Emily Strickland[1], Isabella Harding[1], Shay Muhonen[1], Vlad Savitsky[1], Gabriel Barrett[1], Abdullah Almsaeed[2], Nic Herndon[3], and Margaret Staton[2]

[1]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA;
[2]Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, TN, USA;
[3]Department of Computer Science, East Carolina University, NC, USA; *jill.wegrzyn@uconn.edu

CartograPlant (https://cartograplant.org/) is the first web-based application that integrates genotype and phenotype data for model and non-model plant systems with global environmental layers. This Tripal-based (tripal.info/) field to analysis framework connects data collection, data submission, ontology-based metadata annotation, and analytics directly to high performance computing resources. Genotype and phenotype metrics are collected through direct submission of studies at the time of publication via the FAIR compliant Tripal Plant PopGen Submit data submission module, and through the biocuration efforts of the affiliated databases (TreeGenes, BIEN, Dryad). CartograPlant also integrates with the mobile forest health application, TreeSnap (https://treesnap.org/). Tripal employs the Chado relational database schema (gmod.org/wiki/Chado) which is itself designed to use ontology terms to describe both primary data and metadata. This ontology-centric structure allows CartograPlant to seamlessly filter and analyze genotypic, phenotypic, and environmental data across multiple independent studies from a variety of sources. Phenotypes are composed of several separate elements including structure, observed trait, and units, all of which are described with ontology terms from a number of ontologies including PATO, PO, CHEBI, ENVO, and Woody Plant Ontology. This allows CartograPlant to standardize and group phenotypes despite differing sources.

Today, CartograPlant houses genotype, phenotype metrics from over 300 studies describing 636 plant species from 278 genera. The associated environmental metrics describe abiotic and biotic descriptors including range maps, seed zones, canopy height, land use, forest fragmentation, soil type, precipitation, temperature, aridity, fire, and biotic damage. The underlying analysis framework, developed in Galaxy, provides a mechanism to span studies (meta-analysis) to estimate diversity, calculate population structure, and perform association mapping/landscape genomics.