INTEGRATING, VISUALIZING AND ANALYZING PLANT ENVIRONMENTS, PHENOTYPES AND GENOTYPES USING CARTOGRAPLANT, WILDTYPE AND TRIPAL GALAXY

Irene Cobo-Simon¹, Nic Herndon², Emily Grau¹, Sean Buehler¹, Peter Richter¹, Risharde Ramnath¹, Charles Demurjian¹, Barnaly Pande¹, Emily Strickland¹, Alicia Abrams², Alex Lowe², Brian Holguin-Herrera², Abdullah Almsaeed³, Margaret Staton³, Jill Wegrzyn¹

1. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, US; 2. Department of Computer Science, East Carolina University, NC, US; 3. Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, TN, US

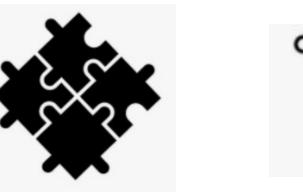
Current statistics

- ✓ 166 studies, 274 species
- ✓ Genotypes and phenotypes from 147,366 plants
- √ 913 environmental layers

WHAT IS CARTOGRAPLANT?



CartograPlant (cartograplant.org) is a web-based application that:



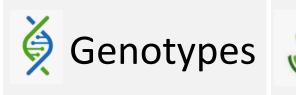


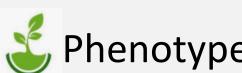


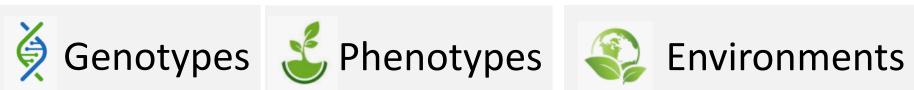
INTEGRATES

VISUALIZES

ANALYZES









From georeferenced plants (including forest trees)

Why is CartograPlant relevant and timely?

1. Climate change is threatening plant health and productivity

2. It is unknown whether plant breeding can keep pace with the rate and direction of environmental change

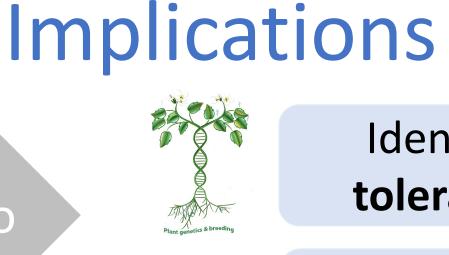
3. The frequency and impact of invasive pests and pathogens is increasing due to globalization and climate change

4. Threats specific to forests: illegal logging and deforestation

Future plant health and productivity will depend on the match of



Necessary to



Identify genes controlling traits that provide tolerance to these biotic and abiotic stresses

Tackle illegal logging and deforestation by identifying the timber origin

Tools that collect, integrate and facilitate this assessment, such as CartograPlant, are critical

Library of worldwide tree samples to help determine timber origin and tackle illegal

logging and deforestation

be used by scientists





New

environments

data types.

DATA INTEGRATION

TreeSnap METADATA + RAW DATA + Citizen science mobile phone **ONTOLOGIES + STANDARDS** application which collects GxPxE phenotypic data from trees to

DATA ANALYSIS

CartograPlant

Galaxy

WildType

where **scientists** can define the

constraints of data collection, targeted

towards landscape-based studies.

Forest tree **genomics** and database

CARTOGRAPLANT FEATURES

BPEN

Database for global patterns of plant diversity, function and distribution.

ENVIRONMENTAL LAYERS

CLIMATE DATA ECOREGIONS SOIL TYPE More customized mobile phone platform

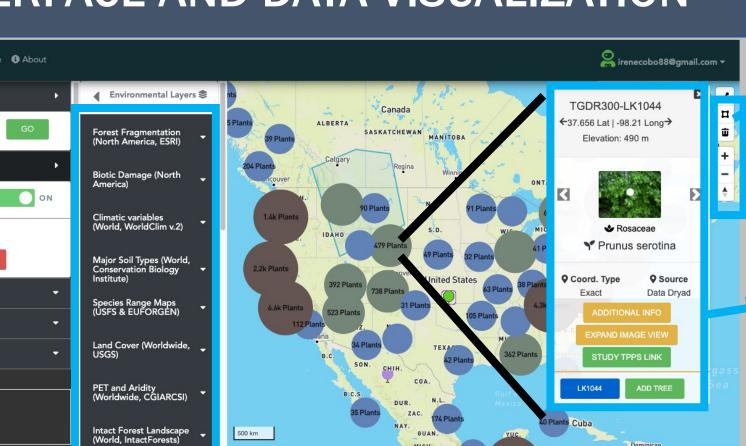
FOREST FRAGMENTATION - FOREST CANOPY - NEON FIELD STATIONS - SEED ZONES **BIODIVERSITY HOTSPOTS** - AGRICULTURE AREAS

- POPULATION DENSITY LOW IMPACT AREAS. - NATIONAL FORESTS

- SPECIES RANGES - PET/ARIDITY - LAND COVER - BIOTIC DAMAGE - NDVI (PLANT HEALTH)

- HUMAN IMPACT - PROTECTED AREAS - INTACT FORESTS - WORLD FOREST ID

INTERFACE AND DATA VISUALIZATION



Zoom/polygon drawing selection

PLANT INDIVIDUAL FEATURES

By clicking on a plant, detailed information is displayed

Users can **navigate** on the map by using the **zoom buttons** or by

typing the coordinates of interest

ENVIRONMENTAL LAYERS

By clicking on the map, information about the layer is



DATA FILTERING



Once data is selected, it can be **filtered by** individual, marker and allele frequency, before data analysis

DATA SELECTION

Plants can be selected for data analysis by multiple filtering criteria (taxonomy, study, data source...) or by drawing the area of interest directly on the map

ANALYTIC WORKFLOWS

NGS: QC and manipu

- CartograPlant offers several analytic workflows via Galaxy (usegalaxy.org), a user-friendly platform that allows highthroughput data analysis without the need of coding skills, thanks to the module **Tripal Galaxy**.
- These flexible analytic workflows can facilitate a diversity of biological questions.
- They also take advantage of the strengths of CartograPlant: the integration of a diversity of data types (genotypic, phenotypic and envrironmental) from different studies, thanks to the metadata collection using ontologies and standards.

Candidate Local genes adaptation 5. ASSOCIATION 6. LANDSCAPE **PHENOTYPE GENOTYPE ENVIRONMENT MAPPING GENOMICS** 4. POPULATION 1.SNP QUALITY FILTERING STRUCTURE/KINSHIP 2. SNP IMPUTATION 3. SNP REMAPPING **SUMMARY STATISTICS METADATA + ONTOLOGIES** + STANDARDS 7. META-

(miappe

WORKFLOW DESCRIPTION

- 1. SNP quality filtering allows to retain high quality SNPs to ensure the
- trustworthiness of further analysis
- 2. Imputation allows to infer the lacking SNPs, enhancing association signals 3. SNPs can be remapped against the newest version of the reference genome. It ensures the transferability and interoperability of SNPs from
- different studies and sequencing techniques. 4. Population structure and kinship calculation is useful to ask population genomic questions but also to avoid spurious associations during association
- mapping (5) and landscape genomics (6) analyses. 7. Meta-analysis is enabled thanks to the metadata collected in CartograPlant. Meta-analysis can increase the power to detect association signals by increasing sample size and by examining more variants















ANALYSIS





throughout the genome than each dataset alone.



