

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

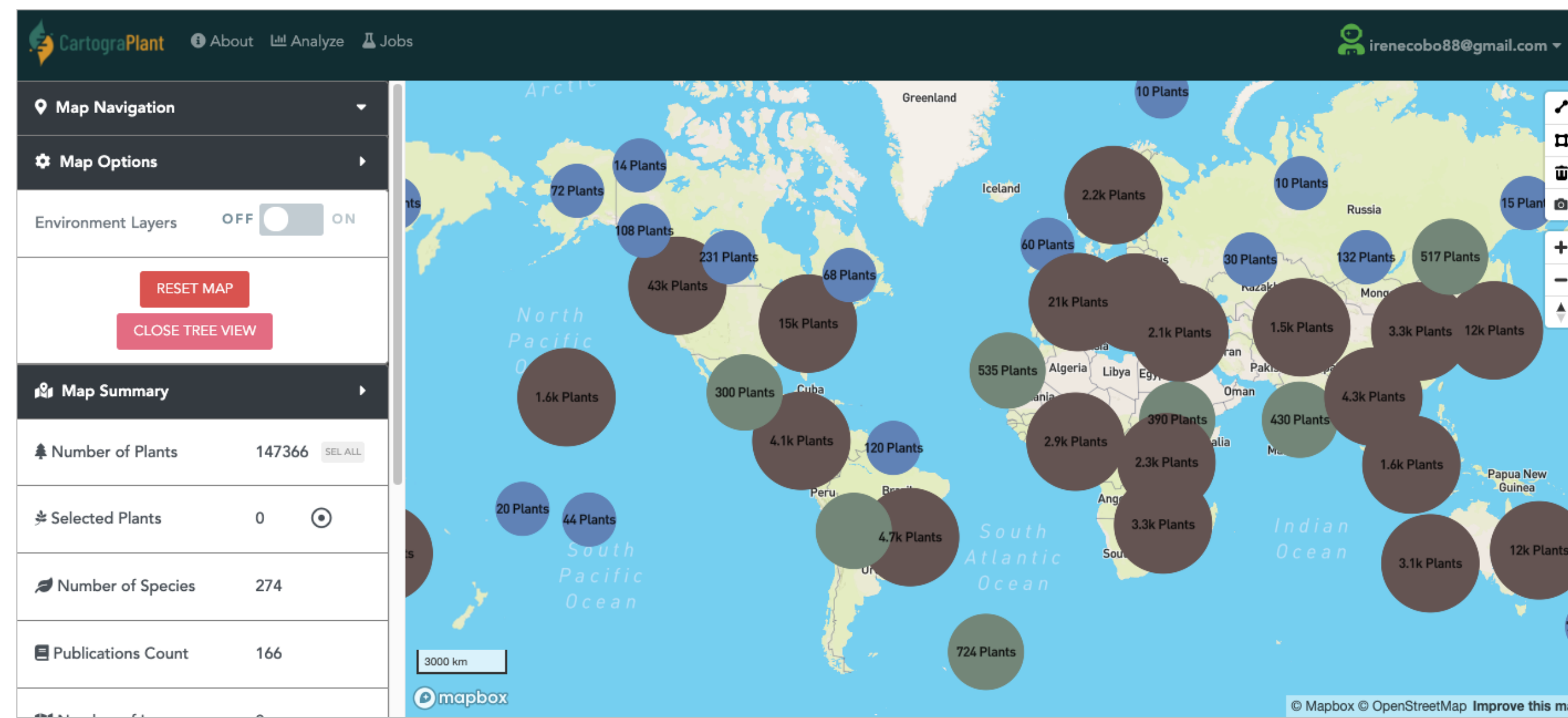
Irene Cobo-Simon¹, Nic Herndon², Emily Grau¹, Sean Buehler¹, Peter Richter¹, Risharde Ramnath¹, Charles Demurjian¹, Barnaly Pande¹, Emily Strickland¹, Alicia Abrams², Alex Lowe², Brian Holguin-Herrera², Abdullah Almsaeed³, Margaret Staton³, Jill Wegrzyn¹

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

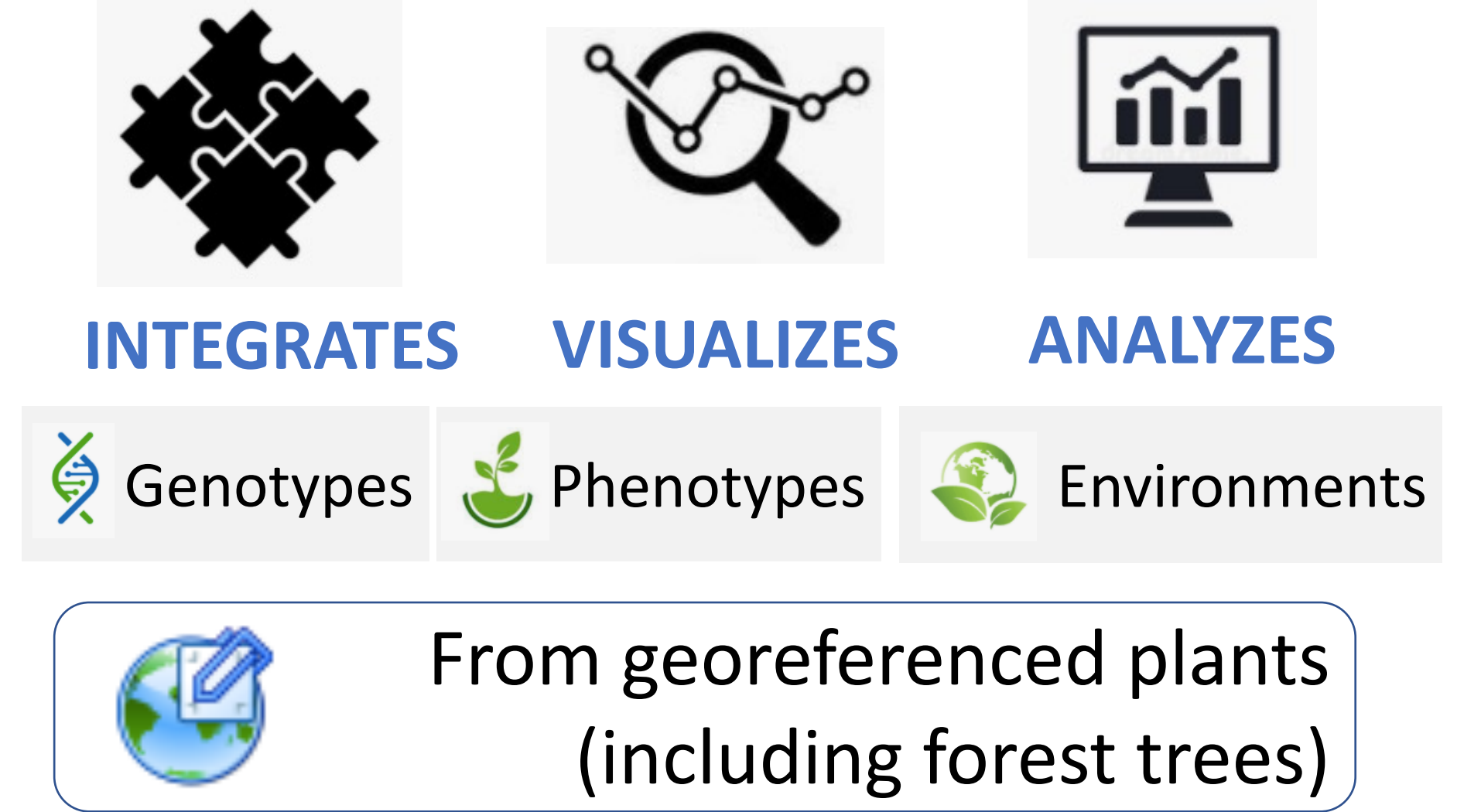
WHAT IS CARTOGRAPLANT?

Current statistics

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



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



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

Implications

Future plant health and productivity will depend on the match of

- Genotypes
- Phenotypes
- New environments

Necessary to

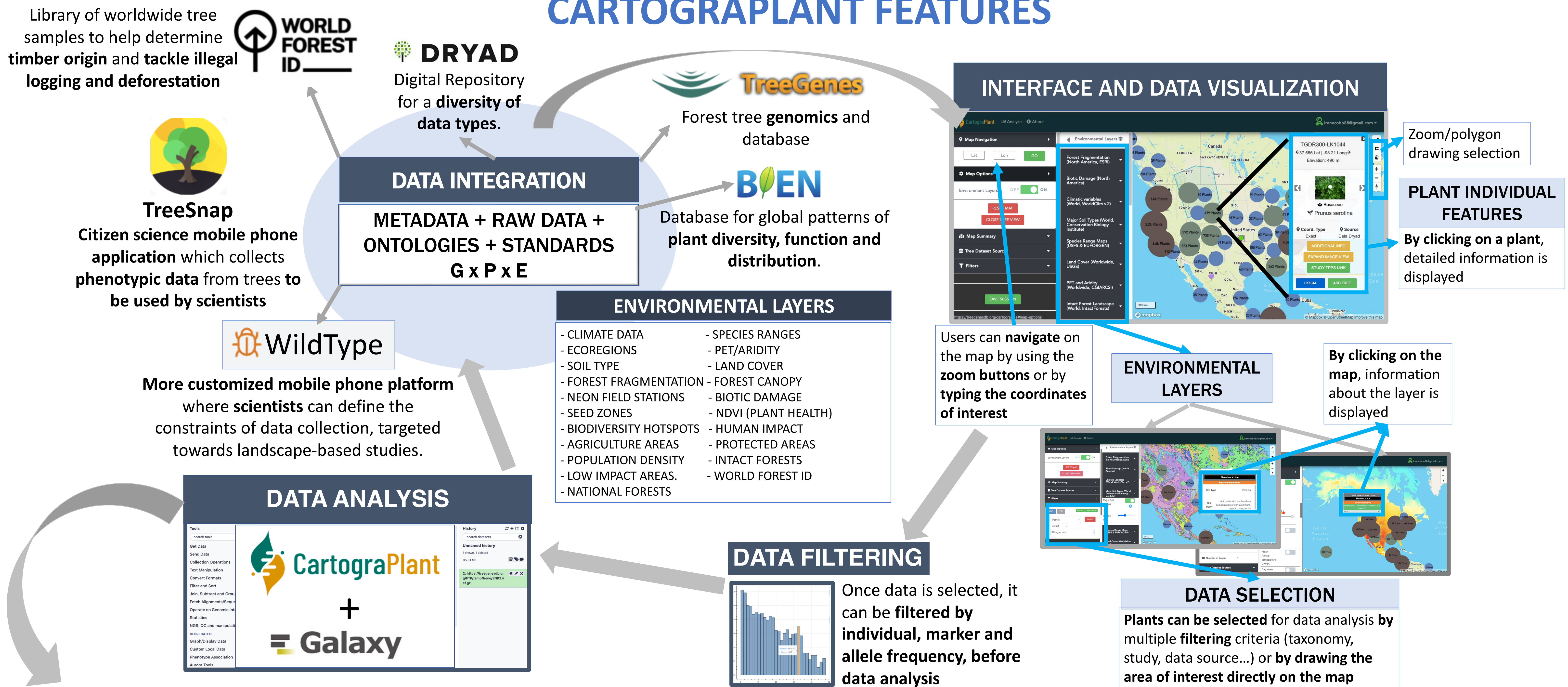


Identify genes controlling traits that provide tolerance to these biotic and abiotic stresses

Tackle illegal logging and deforestation by identifying the timber origin

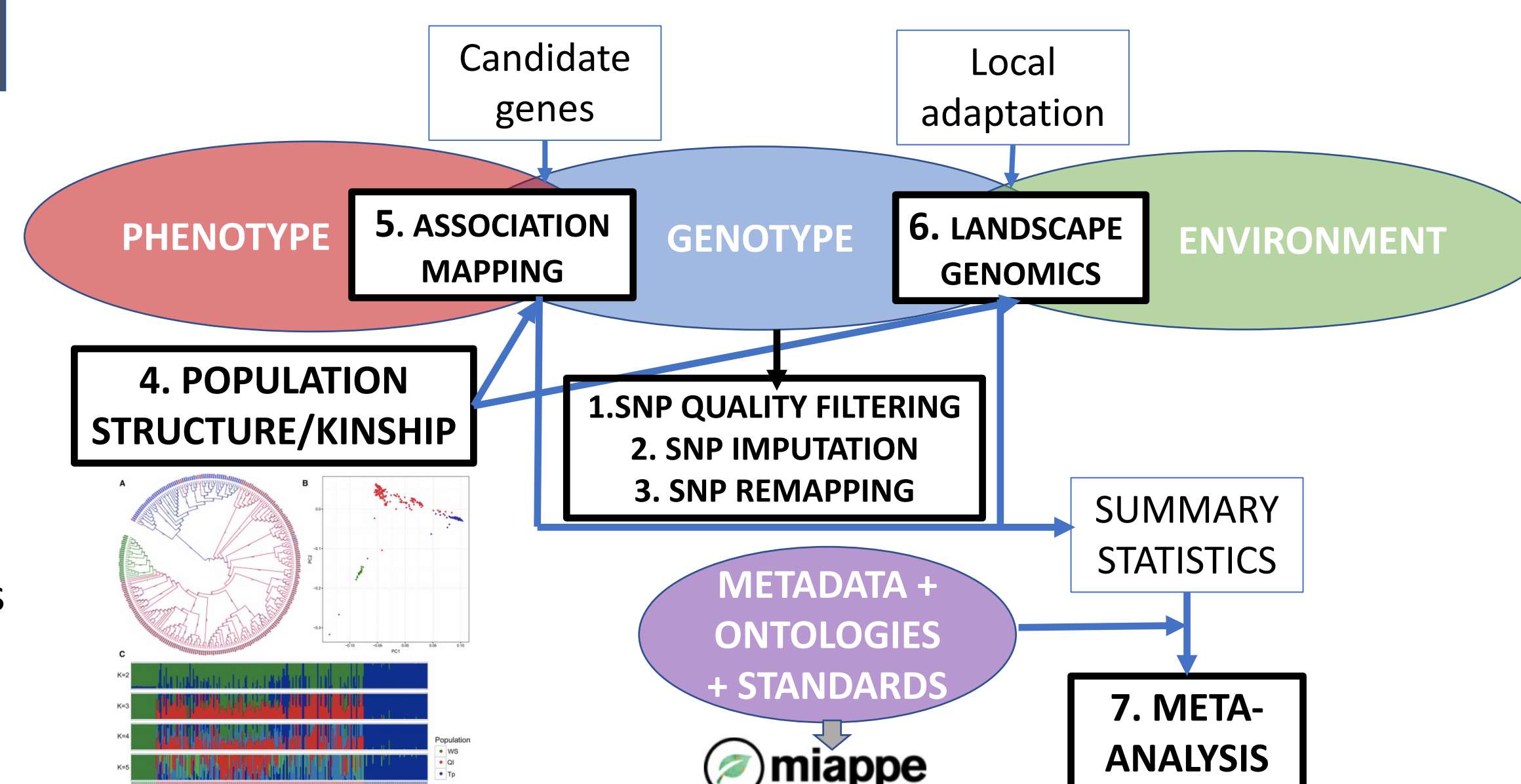
Tools that collect, integrate and facilitate this assessment, such as CartograPlant, are critical

CARTOGRAPLANT FEATURES



ANALYTIC WORKFLOWS

- CartograPlant offers several analytic workflows via Galaxy (usegalaxy.org), a user-friendly platform that allows high-throughput 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 environmental) from different studies, thanks to the metadata collection using ontologies and standards.



WORKFLOW DESCRIPTION

- SNP quality filtering allows to retain high quality SNPs to ensure the trustworthiness of further analysis
- Imputation allows to infer the lacking SNPs, enhancing association signals
- 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.
- 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.
- 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 throughout the genome than each dataset alone.