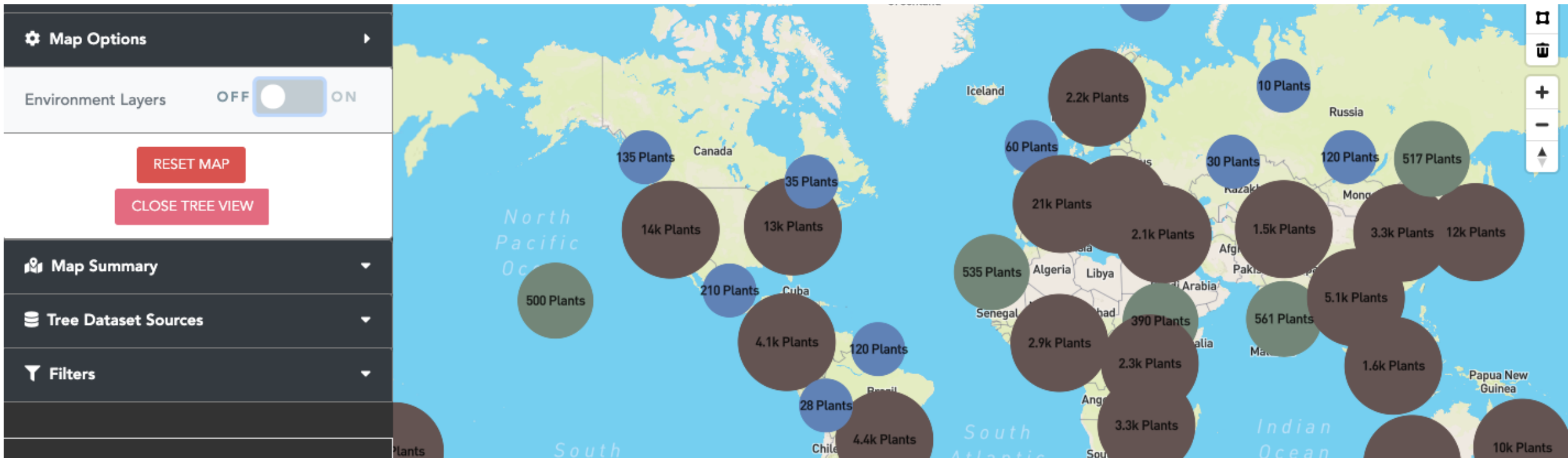


CartograPlant: Cyberinfrastructure to improve plant health and productivity in the context of a changing climate



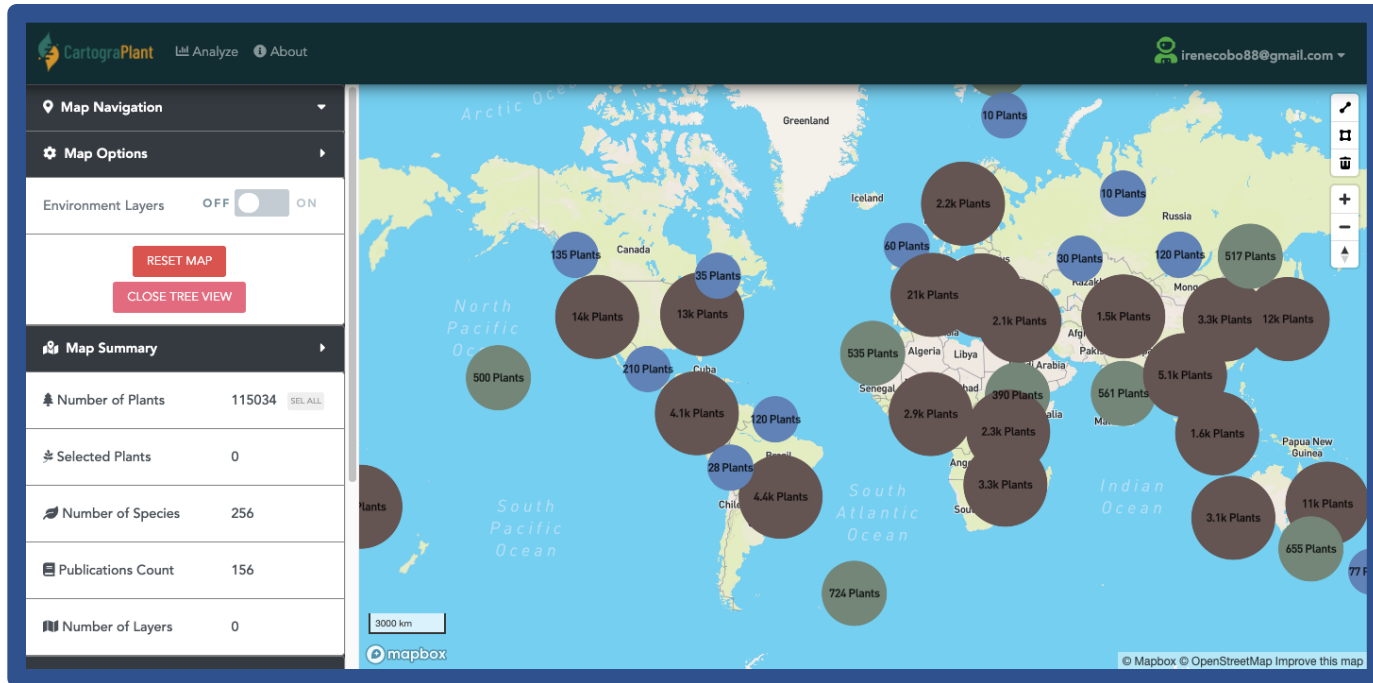
Irene Cobo Simón, PhD.

Postdoctoral Fellow

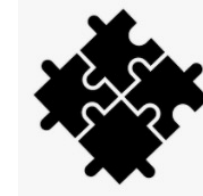
Institute of Forest Science (ICIFOR-INIA-CSIC, Spain)

BACKGROUND

What is CartograPlant?



<https://cartograplant.org/>



INTEGRATES



VISUALIZES



ANALYZES



Genotypes



Phenotypes



Environments



From georeferenced plants
(including forest trees)

BACKGROUND

DATA
INTEGRATION

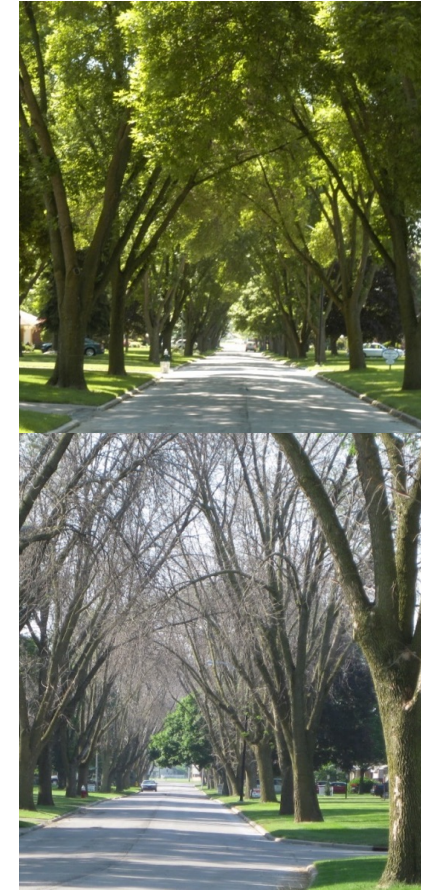
DATA
VISUALIZATION

DATA ANALYSIS

BACKGROUND

Why is CartograPlant relevant and timely?

- **Climate change** is threatening **plant health and productivity**



Green ashes affected by the pest emerald ash borer

BACKGROUND

DATA
INTEGRATION

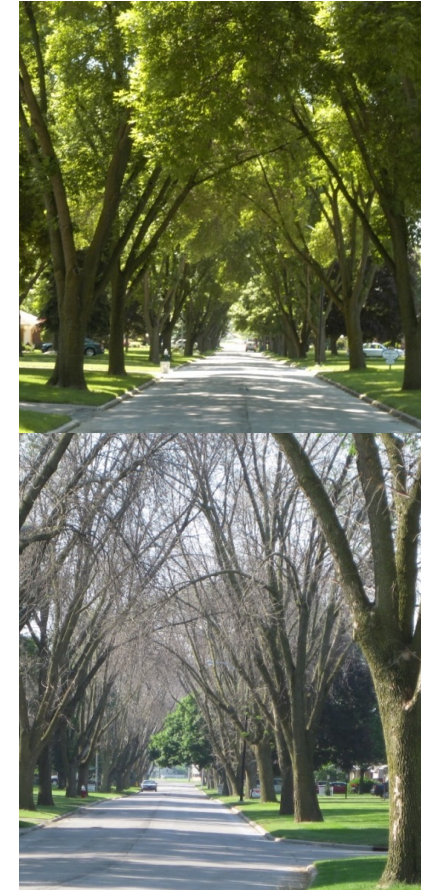
DATA
VISUALIZATION

DATA ANALYSIS

BACKGROUND

Why is CartograPlant relevant and timely?

- **Climate change** is threatening **plant health and productivity**
 - Can **plant breeding** keep pace with the **rate and direction of environmental change?**



Green ashes affected by the pest emerald ash borer

BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

BACKGROUND

Why is CartograPlant relevant and timely?

- **Climate change** is threatening **plant health and productivity**
 - Can **plant breeding** keep pace with the **rate and direction of environmental change**?
 - Increasing **invasive pests and pathogens**



Green ashes affected by the pest emerald ash borer

BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

BACKGROUND

Why is CartograPlant relevant and timely?

- **Climate change** is threatening **plant health and productivity**
 - Can **plant breeding** keep pace with the **rate and direction of environmental change?**
 - Increasing **invasive pests and pathogens**
- **Illegal logging and deforestation** (forest trees)



Green ashes affected by the pest emerald ash borer

BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

BACKGROUND

Why is CartograPlant relevant and timely?

- **Match between genotypes, phenotypes and new environments**



Green ashes affected by the pest emerald ash borer

BACKGROUND

DATA
INTEGRATION

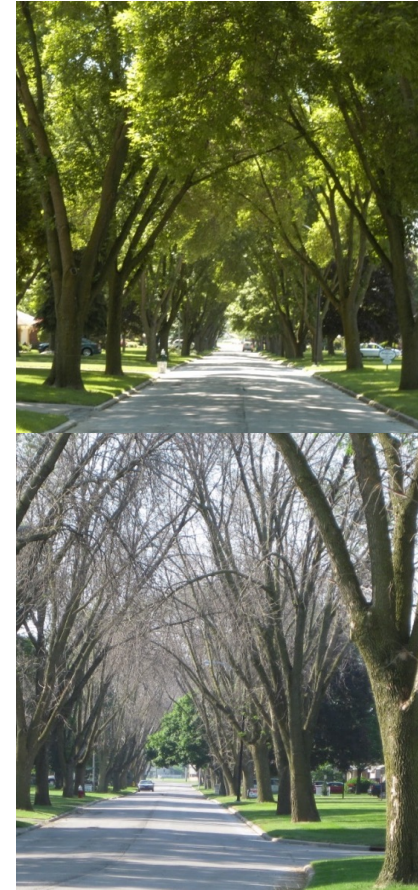
DATA
VISUALIZATION

DATA ANALYSIS

BACKGROUND

Why is CartograPlant relevant and timely?

- **Match between genotypes, phenotypes and new environments**
 - Candidate gene identification (resilience)



Green ashes affected by the pest emerald ash borer

BACKGROUND

DATA
INTEGRATION

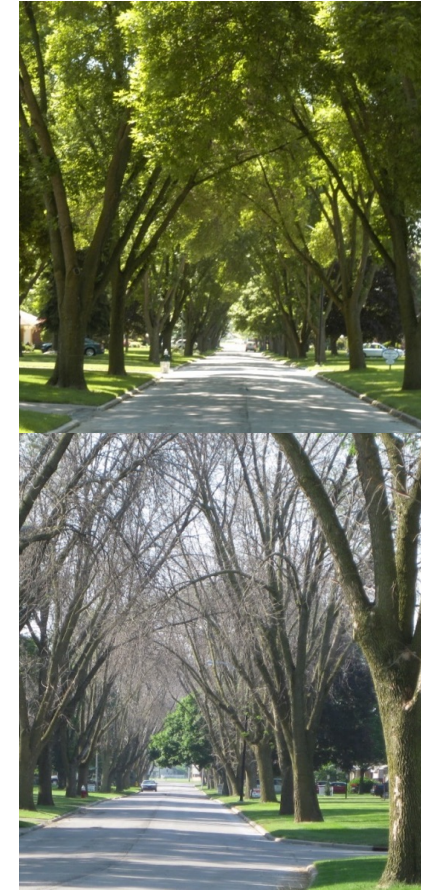
DATA
VISUALIZATION

DATA ANALYSIS

BACKGROUND

Why is CartograPlant relevant and timely?

- **Match between genotypes, phenotypes and new environments**
 - Candidate gene identification (resilience)
 - **Timber origin identification, using chemical, genetic and anatomic tree data**



Green ashes affected by the pest emerald ash borer

BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

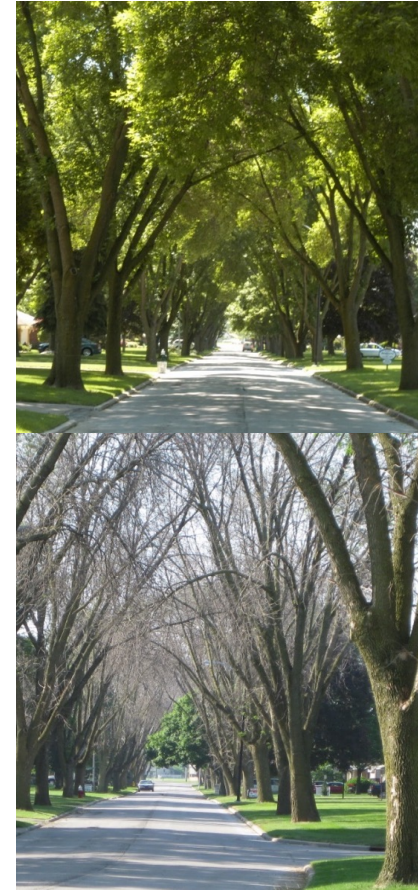
DATA ANALYSIS

BACKGROUND

Why is CartograPlant relevant and timely?

- **Match between genotypes, phenotypes and new environments**
 - Candidate gene identification (resilience)
 - **Timber origin identification, using chemical, genetic and anatomic tree data**

Tools that collect, integrate and facilitate these data, such as CartograPlant, are critical



Green ashes affected by the pest emerald ash borer

BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA INTEGRATION

DATA TYPES INTEGRATED IN CARTOGRAPLANT

1. RAW DATA



GENOTYPIC



PHENOTYPIC



ENVIRONMENTAL



BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA INTEGRATION

DATA TYPES INTEGRATED IN CARTOGRAPLANT

1. RAW DATA



GENOTYPIC



PHENOTYPIC



ENVIRONMENTAL

How can these disparate data types, from different studies, be integrated?



BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA INTEGRATION

DATA TYPES INTEGRATED IN CARTOGRAPLANT

1. RAW DATA



GENOTYPIC

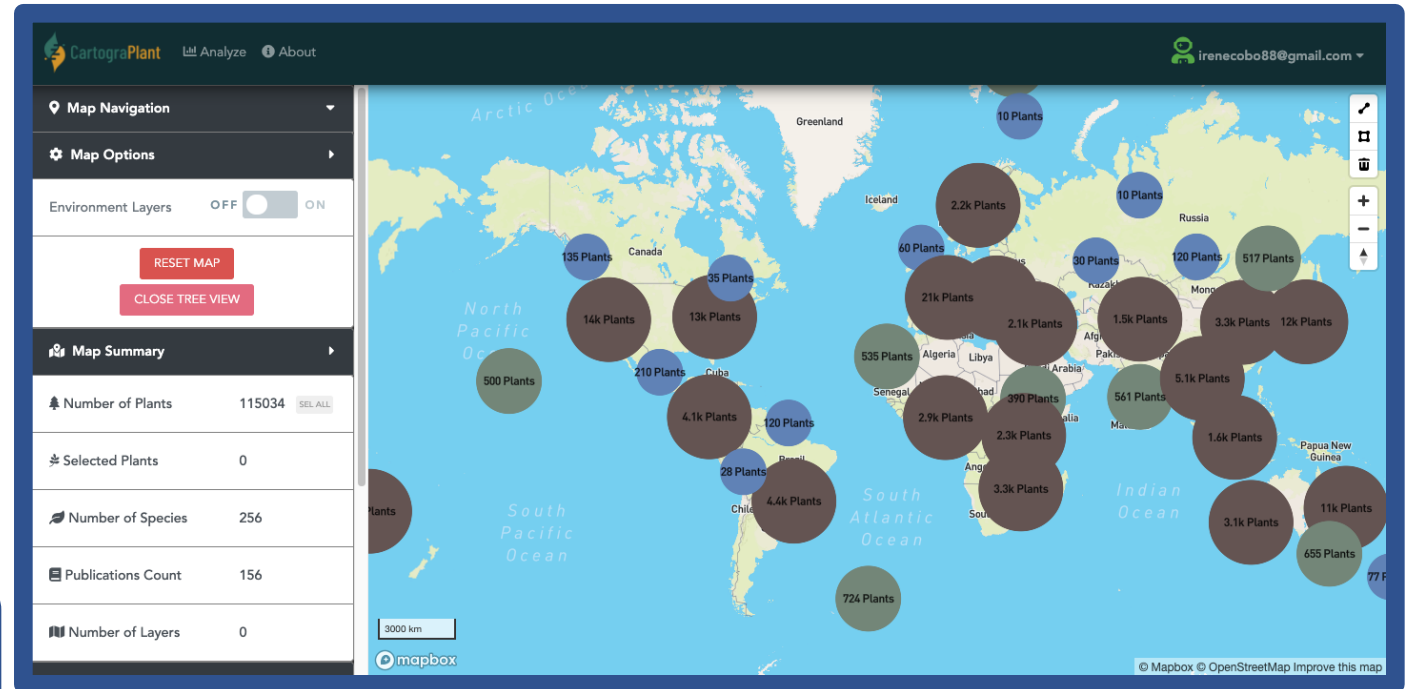


PHENOTYPIC



ENVIRONMENTAL

2. METADATA + ONTOLOGIES + STANDARDS



BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA INTEGRATION

DATA TYPES INTEGRATED IN CARTOGRAPLANT

1. RAW DATA



GENOTYPIC



PHENOTYPIC



ENVIRONMENTAL

2. METADATA + ONTOLOGIES
+ STANDARDS

Direct submission of studies



**Tripal Plant PopGen Submit
(TPPS) pipeline**

BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA INTEGRATION

DATA TYPES INTEGRATED IN CARTOGRAPLANT

1. RAW DATA



GENOTYPIC



PHENOTYPIC



ENVIRONMENTAL

2. METADATA + ONTOLOGIES
+ STANDARDS

Direct submission of studies



**Tripal Plant PopGen Submit
(TPPS) pipeline**

Biocuration efforts of affiliated
databases



BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

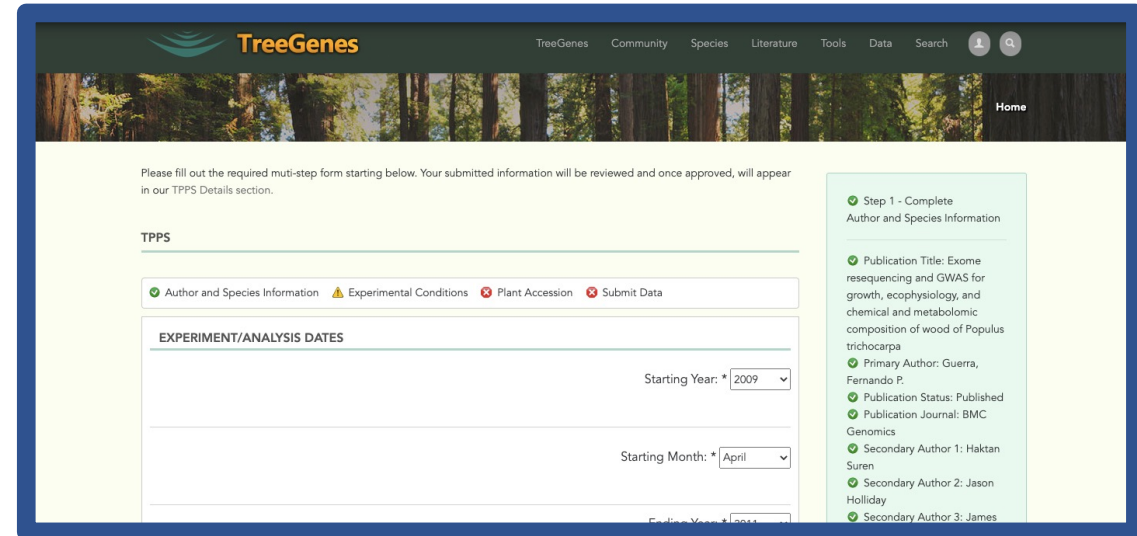
DATA ANALYSIS

DATA INTEGRATION

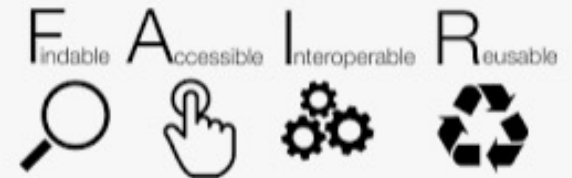
Direct submission: Tripal Plant PopGen Submit (TPPS) pipeline

<https://treegenesdb.org/tpps>

- **Genotype, phenotype, environmental data and metadata** and provides a **DOI**



The screenshot shows the TreeGenes TPPS submission form. The form is titled "TPPS" and includes a progress bar with four steps: "Author and Species Information" (completed), "Experimental Conditions" (warning), "Plant Accession" (error), and "Submit Data" (error). The "EXPERIMENT/ANALYSIS DATES" section includes fields for "Starting Year" (2009) and "Starting Month" (April). A sidebar on the right shows a list of completed steps and their details, including "Step 1 - Complete Author and Species Information" with a list of authors and publication information.



BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA INTEGRATION

Direct submission: Tripal Plant PopGen Submit (TPPS) pipeline

<https://treegenesdb.org/tpps>

- **Genotype, phenotype, environmental data and metadata** and provides a DOI
- **Population genomics, association mapping, and landscape genomic studies**

Please fill out the required multi-step form starting below. Your submitted information will be reviewed and once approved, will appear in our TPPS Details section.

TPPS

Author and Species Information Experimental Conditions Plant Accession Submit Data

EXPERIMENT/ANALYSIS DATES

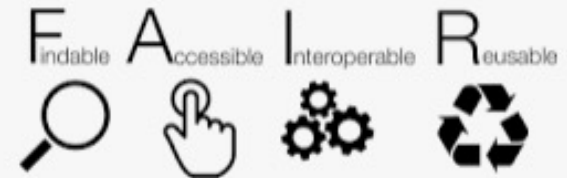
Starting Year: * 2009

Starting Month: * April

Ending Year: * 2011

Step 1 - Complete Author and Species Information

- Publication Title: Exome resequencing and GWAS for growth, ecophysiology, and chemical and metabolomic composition of wood of Populus trichocarpa
- Primary Author: Guerra, Fernando P.
- Publication Status: Published
- Publication Journal: BMC Genomics
- Secondary Author 1: Haktan Suren
- Secondary Author 2: Jason Holliday
- Secondary Author 3: James



BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA INTEGRATION

Direct submission: Tripal Plant PopGen Submit (TPPS) pipeline

<https://treegenesdb.org/tpps>

- **Genotype, phenotype, environmental data and metadata** and provides a **DOI**
- **Population genomics, association mapping, and landscape genomic studies**
- **Metadata is collected using ontologies and standards (MIAPPE)**

Please fill out the required multi-step form starting below. Your submitted information will be reviewed and once approved, will appear in our TPPS Details section.

TPPS

Author and Species Information Experimental Conditions Plant Accession Submit Data

EXPERIMENT/ANALYSIS DATES

Starting Year: * 2009

Starting Month: * April

Ending Year: * 2011

Step 1 - Complete Author and Species Information

Publication Title: Exome resequencing and GWAS for growth, ecophysiology, and chemical and metabolomic composition of wood of Populus trichocarpa

Primary Author: Guerra, Fernando P.

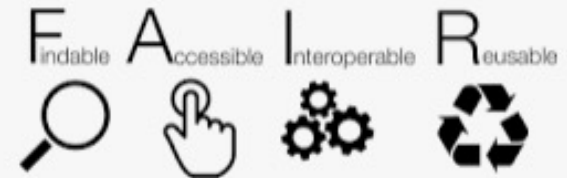
Publication Status: Published

Publication Journal: BMC Genomics

Secondary Author 1: Haktan Suren

Secondary Author 2: Jason Holliday

Secondary Author 3: James



BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

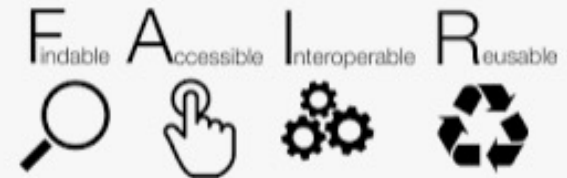
DATA ANALYSIS

DATA INTEGRATION

Direct submission: Tripal Plant PopGen Submit (TPPS) pipeline

<https://treegenesdb.org/tpps>

- **Genotype, phenotype, environmental data and metadata** and provides a **DOI**
- **Population genomics, association mapping, and landscape genomic studies**
- **Metadata** is collected using **ontologies and standards (MIAPPE)**
- Ensures the **FAIRness** of the data



BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

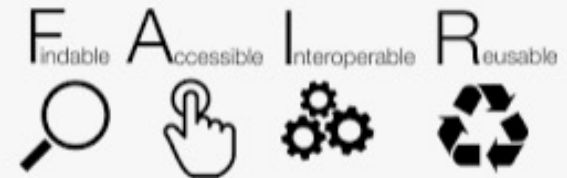
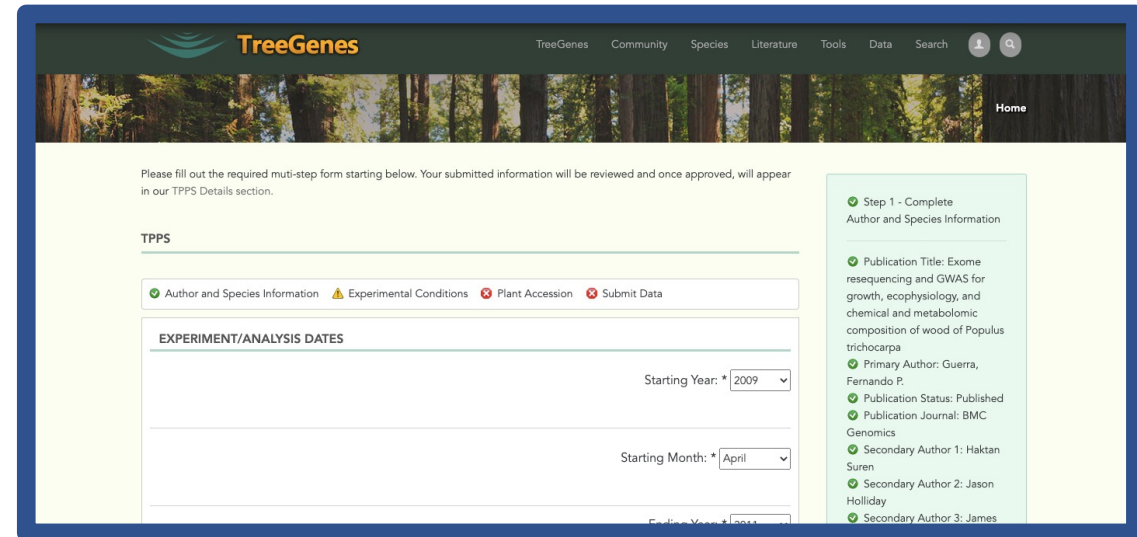
DATA ANALYSIS

DATA INTEGRATION

Direct submission: Tripal Plant PopGen Submit (TPPS) pipeline

<https://treegenesdb.org/tpps>

- **Genotype, phenotype, environmental data and metadata** and provides a **DOI**
- **Population genomics, association mapping, and landscape genomic studies**
- **Metadata** is collected using **ontologies and standards (MIAPPE)**
- Ensures the **FAIRness** of the data
- Supports standard **genotyping file formats (VCF)**



BACKGROUND


DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA INTEGRATION

Direct submission: Tripal Plant PopGen Submit (TPPS) pipeline



Full paper | [Free Access](#)

Geographical and environmental gradients shape phenotypic trait variation and genetic structure in *Populus trichocarpa*

Athena D. McKown, Robert D. Guy, Jaroslav Klápště, Armando Galdes, Michael Friedmann
... [See all authors](#)

First published: 25 November 2013 | <https://doi.org/10.1111/nph.12601> | Citations: 128

SECTIONS PDF TOOLS

Summary

- Populus trichocarpa* is widespread across western North America spanning extensive variation in photoperiod, growing season and climate. We investigated trait variation in *P. trichocarpa* using over 2000 trees from a common garden at Vancouver, Canada representing replicate plantings of 461 genotypes originating from 136 provenance localities.

MOLECULAR ECOLOGY

Fine scale genetic structure in the wild ancestor of maize (*Zea mays ssp. parviglumis*)

JOOST VAN HEERWAARDEN, JEFFREY ROSS-IBARRA, JOHN DOEBLEY, JEFFREY C. GLAUBITZ
... [See all authors](#)

First published: 23 February 2010 | <https://doi.org/10.1111/j.1365-294X.2010.04559.x> | Citations: 25

✉ Jeffrey Ross-Ibarra, E-mail: rossibarra@ucdavis.edu

Read the full text > PDF TOOLS SHARE

Abstract

Analysis of fine scale genetic structure in continuous populations of outcrossing plant species has traditionally been limited by the availability of sufficient markers. We used a set of 468 SNPs to characterize fine-scale genetic structure within and between two dense stands of the wild ancestor of maize, teosinte (*Zea mays ssp. parviglumis*). Our analyses confirmed that teosinte is highly outcrossing and showed little population structure over short distances. We found that the two populations were clearly

UCONN BIOCURATION TEAM

Charles Demurjian, MSc

Emily Strickland, BSc

Victoria Burton, BSc

Meghan Myles

Maddie Gadowski

The vast majority of the TPPS submitted studies (and their associated genotype, phenotype and environmental data) are available in CartograPlant **thanks to our Biocuration team**

BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

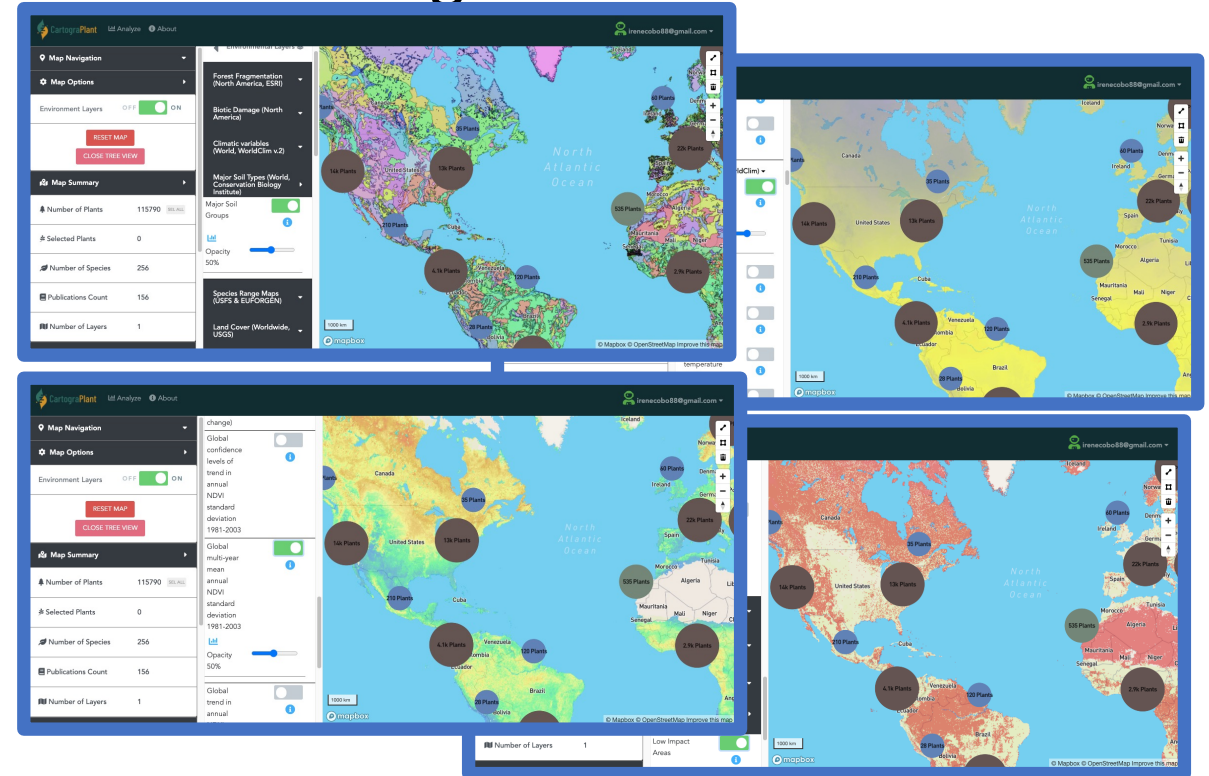
DATA ANALYSIS

DATA INTEGRATION

Environmental layers

950 environmental layers are now available in CartograPlant

- CLIMATE DATA
- ECOREGIONS
- SOIL TYPE
- FOREST FRAGMENTATION
- NEON FIELD STATIONS
- SEED ZONES
- BIODIVERSITY HOTSPOTS
- WORLD FOREST ID DATA
- POPULATION DENSITY
- LOW IMPACT AREAS.
- NATIONAL FORESTS
- SPECIES RANGES
- PET/ARIDITY
- LAND COVER
- CANOPY HEIGHT
- BIOTIC DAMAGE
- NDVI (PLANT HEALTH)
- HUMAN IMPACT
- PROTECTED AREAS
- INTACT FORESTS
- WORLD FOREST ID DATA
- POPULATION DENSITY



BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA INTEGRATION

CartograPlant current statistics

- Plants **8,439,968**
- Species **635**
- Genera **277**
- Countries **43**
- Studies **313**
- Genotypes **771,763,817**
- Phenotypes **1,741,822**
- Environmental layers **950**



BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA VISUALIZATION

LEFT PANEL

Action panel to interact with the map and plants, located to the left of the screen



BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA VISUALIZATION

LEFT PANEL

Action panel to interact with the map and plants, located to the left of the screen



RIGHT PANEL

An interactive map, showing the selected plants and environmental layers on the left panel

BACKGROUND

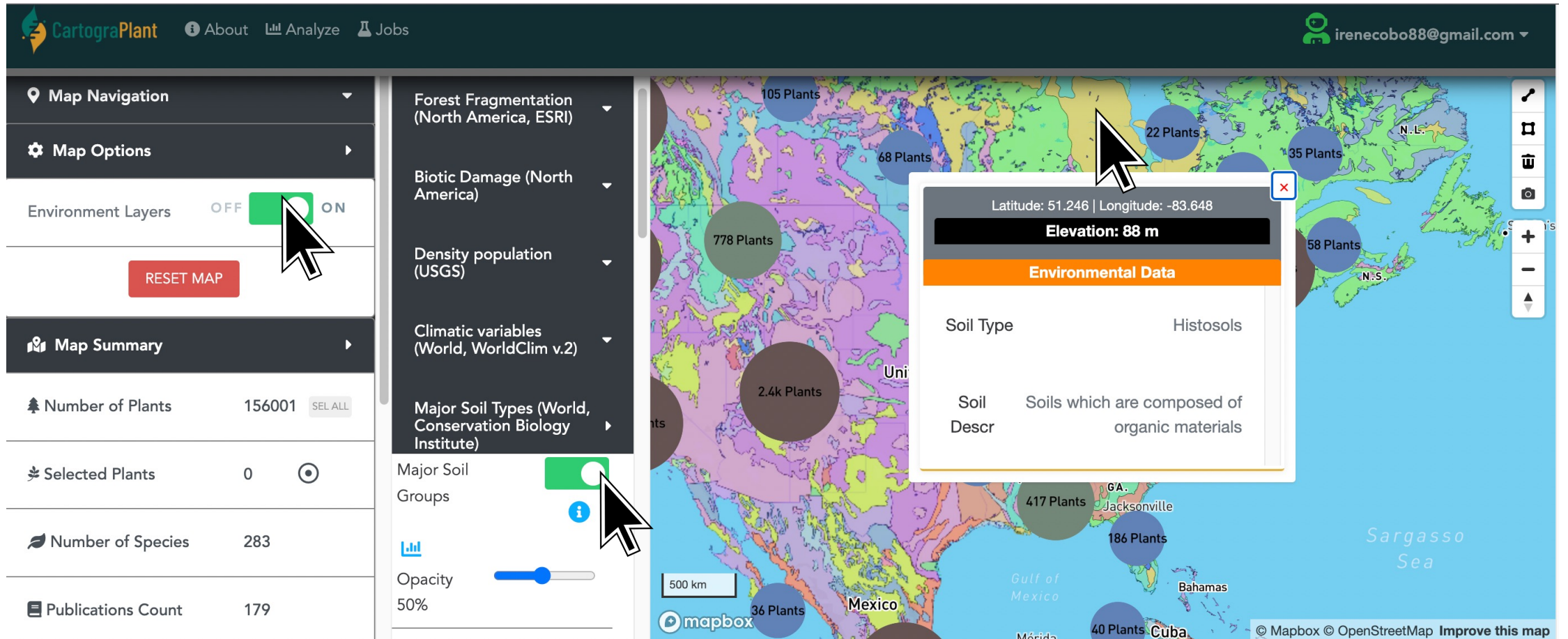
DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA VISUALIZATION

<https://cartograplant.org/>



BACKGROUND

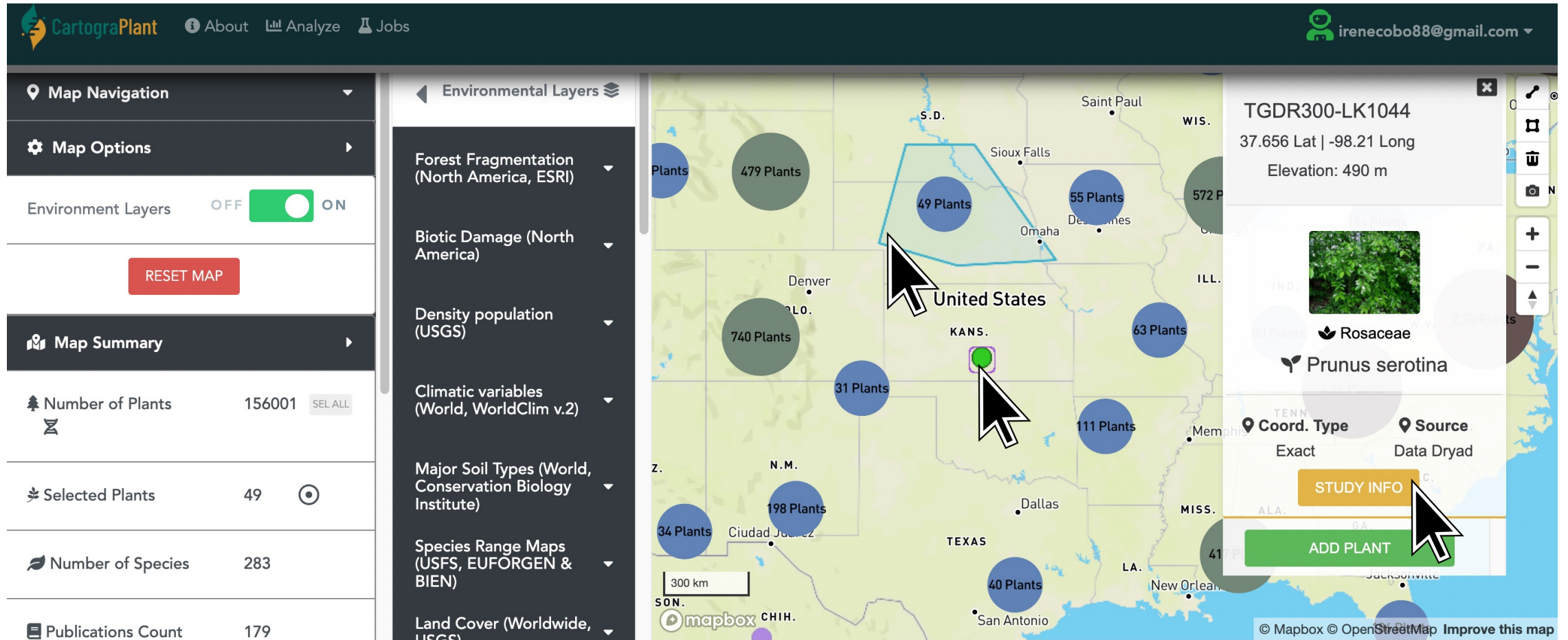
DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA VISUALIZATION

<https://cartograplant.org/>



BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

CartograPlant

Map Navigation

Map Options

Environment Layer

Map Summary

Number of Plants

Selected Plants

Number of Species

Publications Co

TGDR300-LK1044
✕

BIOME Temperate Grasslands, Savannas & Shrublands

Study Associated

A rangewide herbarium-derived dataset indicates high levels of gene flow in black cherry (*Prunus serotina*)

Konrade, Lauren 2019 [View Additional Details](#)

STUDY FILE DOWNLOADS

Phenotype Accession

Genotype SSRs/cpSSRs

Markers

microsatellite

Markers Count

28000

Genotypic Data

Marker Name	Genotype	Marker Type
509_48	NA	microsatellite
506_34	230	microsatellite
510_50	175	microsatellite

88@gmail.com

44

ong

an

02

is

ae

erotina

55 Plants

Source

Data Dryad

47 Plants

reetMap Improve this map

BACKGROUND

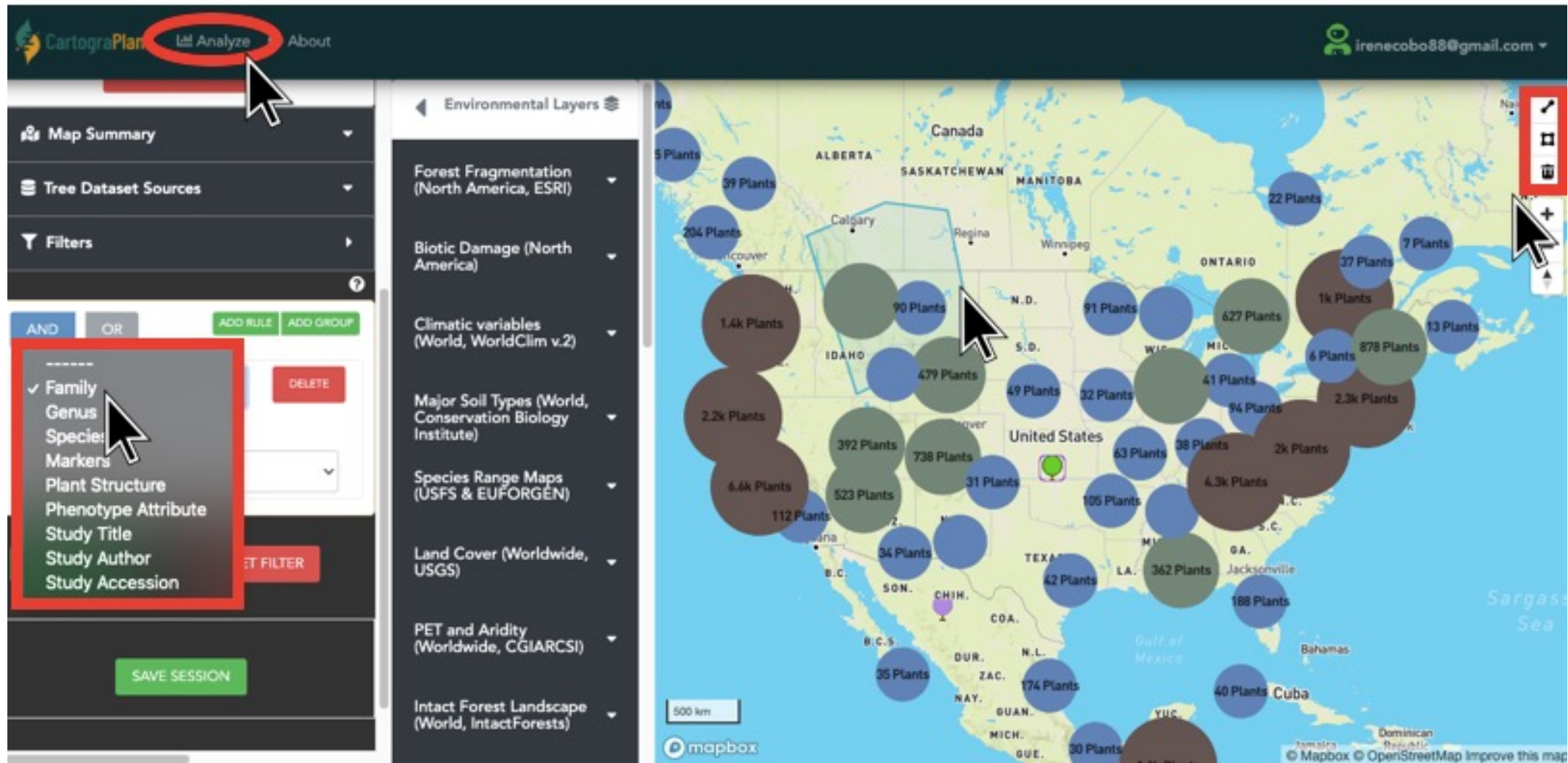
DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA VISUALIZATION

<https://cartograplant.org/>



BACKGROUND

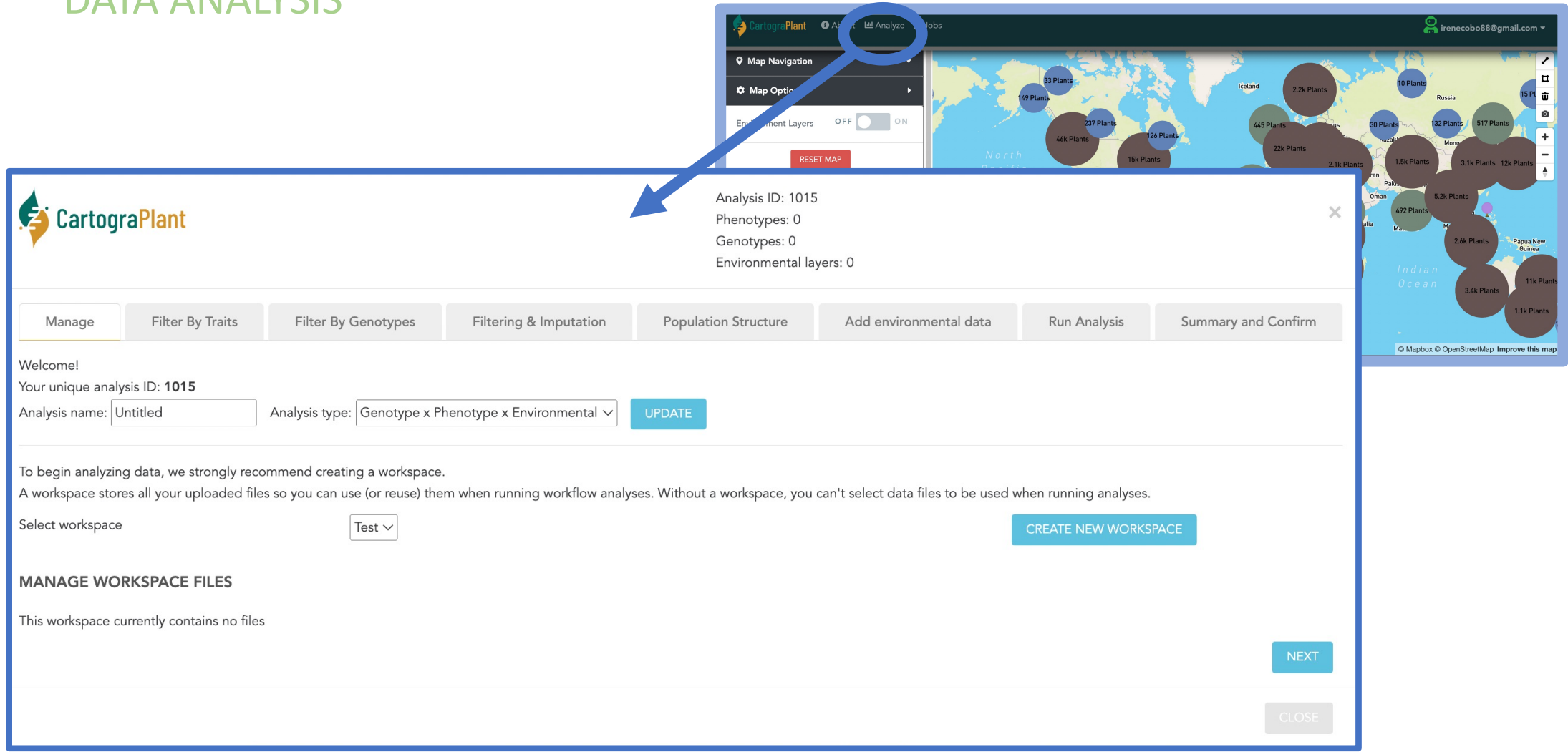
DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA ANALYSIS

<https://cartograplant.org/>



The image shows a screenshot of the CartograPlant web application. At the top right, there is a navigation menu with 'Analyze' highlighted. Below it, a map displays various plant locations with circular markers of different sizes and colors, each labeled with the number of plants (e.g., 33 Plants, 149 Plants, 2.2k Plants, 10 Plants, 15 Plants, 30 Plants, 132 Plants, 517 Plants, 445 Plants, 22k Plants, 2.1k Plants, 1.5k Plants, 3.1k Plants, 12k Plants, 492 Plants, 5.2k Plants, 2.6k Plants, 11k Plants, 3.4k Plants, 1.1k Plants). A blue arrow points from the 'Analyze' button to a summary panel.

Analysis ID: 1015
Phenotypes: 0
Genotypes: 0
Environmental layers: 0

Manage | Filter By Traits | Filter By Genotypes | Filtering & Imputation | Population Structure | Add environmental data | Run Analysis | Summary and Confirm

Welcome!
Your unique analysis ID: **1015**

Analysis name: Analysis type: **UPDATE**

To begin analyzing data, we strongly recommend creating a workspace.
A workspace stores all your uploaded files so you can use (or reuse) them when running workflow analyses. Without a workspace, you can't select data files to be used when running analyses.

Select workspace: **CREATE NEW WORKSPACE**

MANAGE WORKSPACE FILES
This workspace currently contains no files

NEXT
CLOSE

BACKGROUND

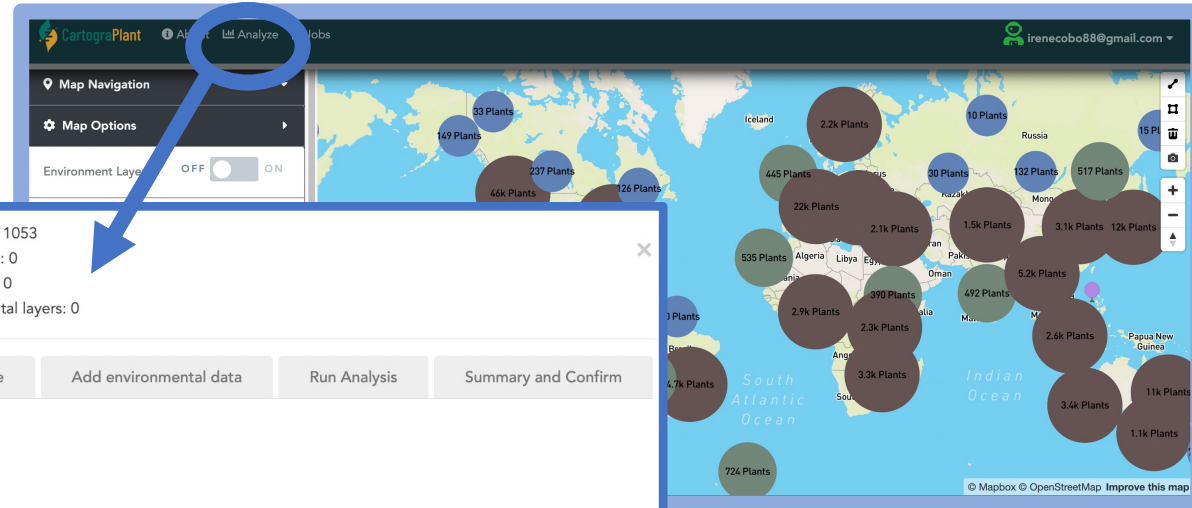
DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA ANALYSIS

<https://cartograplant.org/>



CartograPlant

Analysis ID: 1053
Phenotypes: 0
Genotypes: 0
Environmental layers: 0

Manage | Filter By Traits | Filter By Genotypes | Filtering & Imputation | Population Structure | Add environmental data | Run Analysis | Summary and Confirm

Filter By Traits

2 studies detected based on the trees you selected on the map

TGDR674 | TGDR682

<input type="checkbox"/> bud break	1836 phenotypes	Overlaps with all studies
<input type="checkbox"/> bud set	2754 phenotypes	Overlaps with all studies
<input type="checkbox"/> carbon to nitrogen ratio	1836 phenotypes	Overlaps with all studies
<input type="checkbox"/> frost free days	918 phenotypes	Overlaps with all studies
<input type="checkbox"/> leaf chlorophyll content	2754 phenotypes	Overlaps with all studies
<input type="checkbox"/> leaf nitrogen 15 content	1836 phenotypes	Overlaps with all studies
<input type="checkbox"/> leaf shape	918 phenotypes	Overlaps with all studies
<input type="checkbox"/> mean annual precipitation	918 phenotypes	Overlaps with all studies
<input type="checkbox"/> mean annual temperature	918 phenotypes	Overlaps with all studies
<input type="checkbox"/> plant height	3672 phenotypes	Overlaps with all studies
<input type="checkbox"/> ratio of average warmest month temperature to average summer precipitation	918 phenotypes	Overlaps with all studies
<input type="checkbox"/> volume	2754 phenotypes	Overlaps with all studies
<input type="checkbox"/> whole plant mass	918 phenotypes	Overlaps with all studies
<input type="checkbox"/> wood carbon 13 content	918 phenotypes	Overlaps with all studies

BACKGROUND

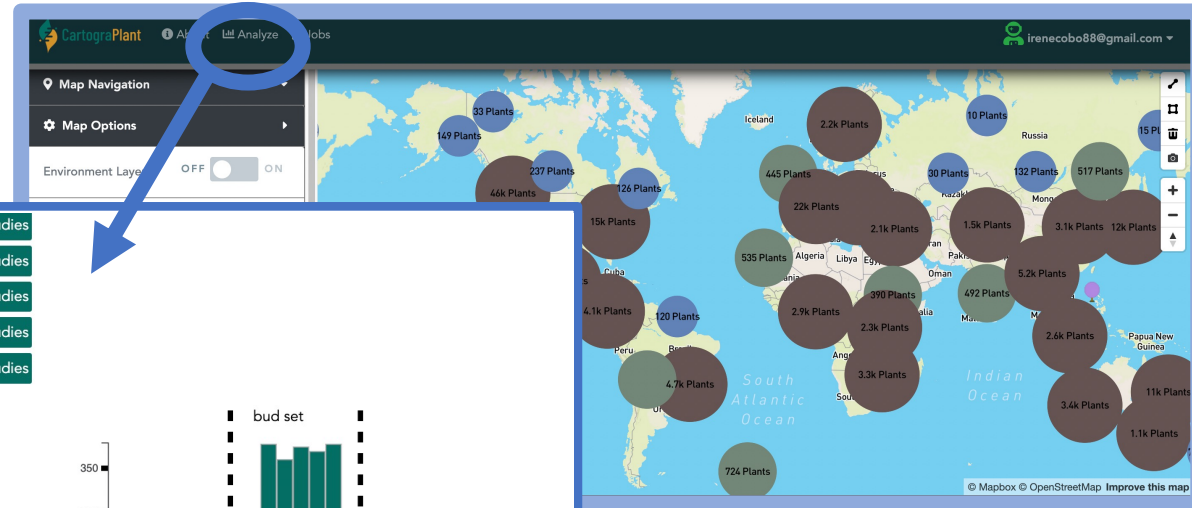
DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA ANALYSIS

<https://cartograplant.org/>



- plant height
- ratio of average warmest month temperature to average summer precipitation
- volume
- whole plant mass
- wood carbon 13 content

3672 phenotypes	Overlaps with all studies
918 phenotypes	Overlaps with all studies
2754 phenotypes	Overlaps with all studies
918 phenotypes	Overlaps with all studies
918 phenotypes	Overlaps with all studies

bud break

Adjust thresholds

bud set

Adjust thresholds

carbon to nitrogen ratio

Adjust thresholds

bud set

Save adjustment

frost free days

leaf chlorophyll content

leaf nitrogen 15 content

BACKGROUND

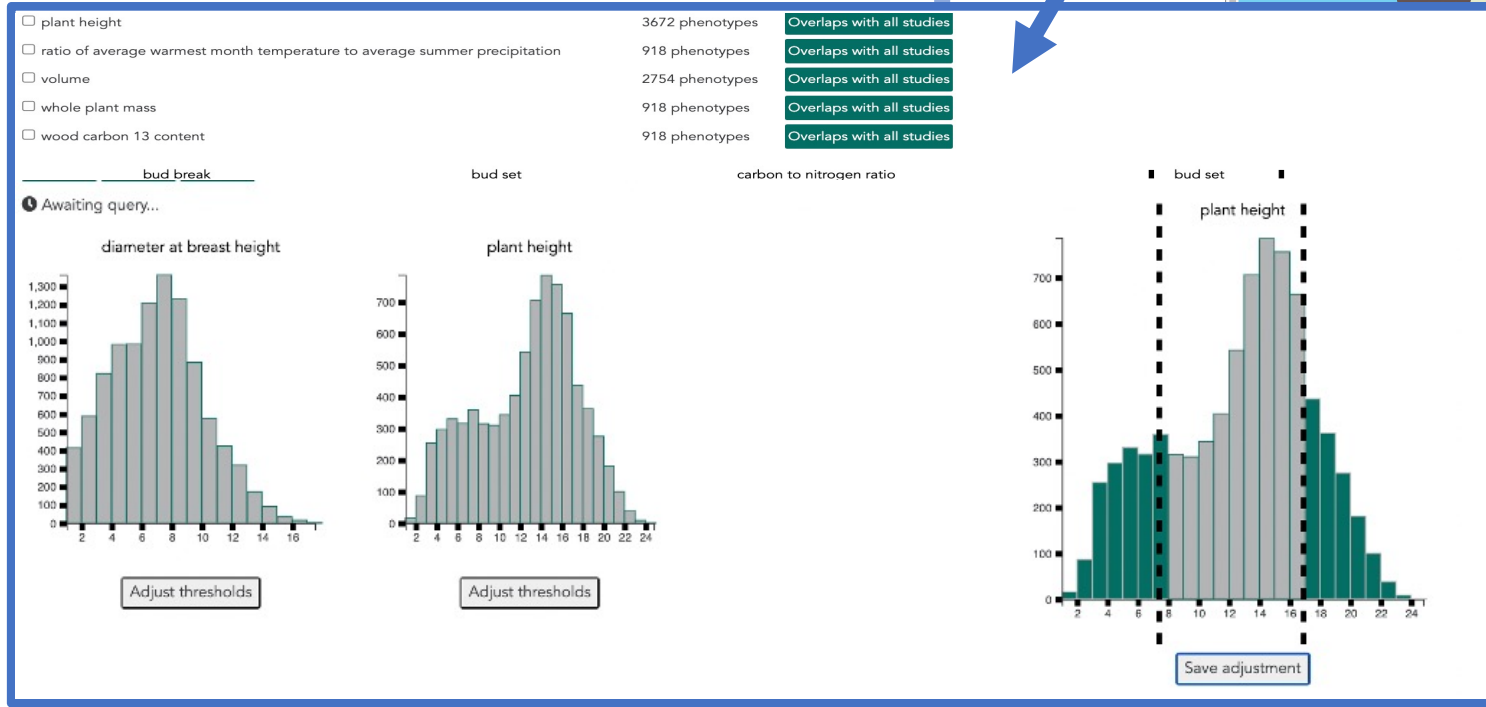
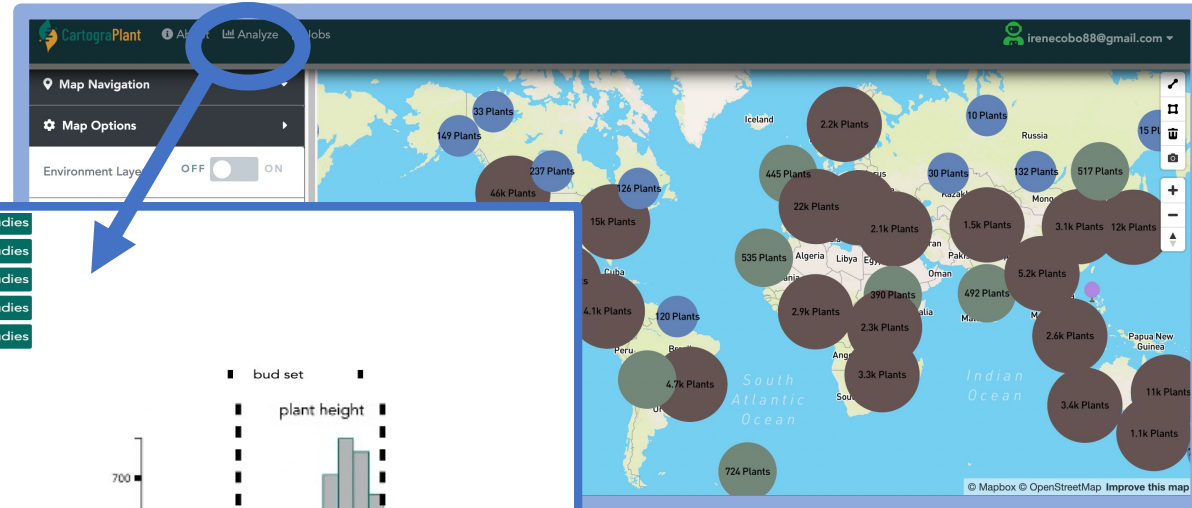
DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA ANALYSIS

<https://cartograplant.org/>



BACKGROUND

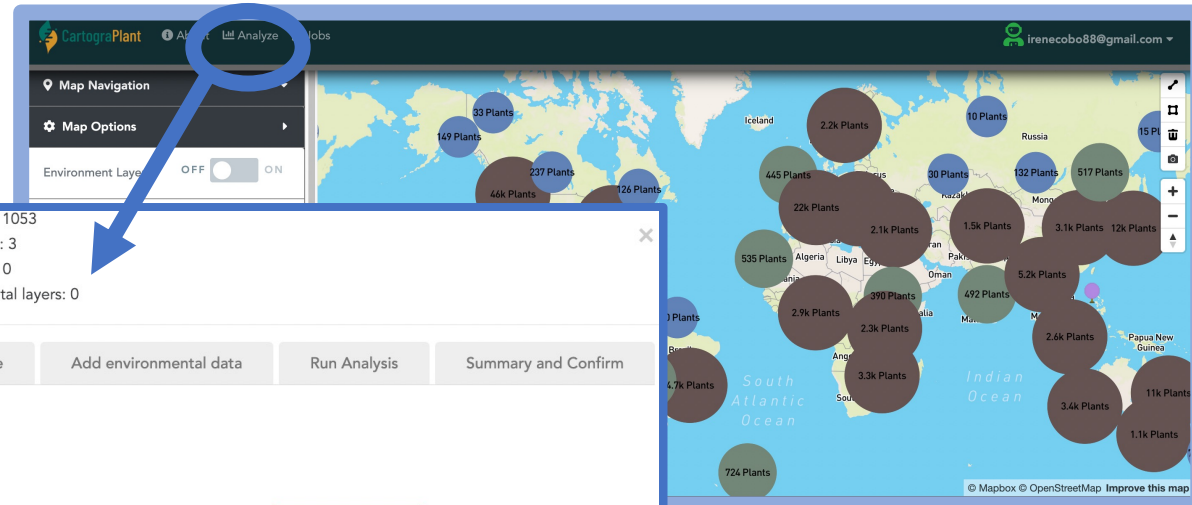
DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA ANALYSIS

<https://cartograplant.org/>



Analysis ID: 1053
Phenotypes: 3
Genotypes: 0
Environmental layers: 0

Manage | Filter By Traits | Filter By Genotypes | Filtering & Imputation | Population Structure | Add environmental data | Run Analysis | Summary and Confirm

Filter By Traits

2 studies detected based on the trees you selected on the map

TGDR674 | TGDR682

<input checked="" type="checkbox"/> bud break	1836 phenotypes	Overlaps with all studies
<input checked="" type="checkbox"/> bud set	2411 phenotypes	Overlaps with all studies
<input type="checkbox"/> carbon to nitrogen ratio	1836 phenotypes	Overlaps with all studies
<input type="checkbox"/> frost free days	918 phenotypes	Overlaps with all studies
<input type="checkbox"/> leaf chlorophyll content	2754 phenotypes	Overlaps with all studies
<input type="checkbox"/> leaf nitrogen 15 content	1836 phenotypes	Overlaps with all studies
<input type="checkbox"/> leaf shape	918 phenotypes	Overlaps with all studies
<input type="checkbox"/> mean annual precipitation	918 phenotypes	Overlaps with all studies
<input type="checkbox"/> mean annual temperature	918 phenotypes	Overlaps with all studies
<input checked="" type="checkbox"/> plant height	3672 phenotypes	Overlaps with all studies
<input type="checkbox"/> ratio of average warmest month temperature to average summer precipitation	918 phenotypes	Overlaps with all studies
<input type="checkbox"/> volume	2754 phenotypes	Overlaps with all studies
<input type="checkbox"/> whole plant mass	918 phenotypes	Overlaps with all studies
<input type="checkbox"/> wood carbon 13 content	918 phenotypes	Overlaps with all studies

PCA (3 phenotypes)

standardized PC2 (33.2% explained var.)

standardized PC1 (63.1% explained var.)

■ bud set ■

artogratree#analysis-filter-snp

bud set

carbon to nitrogen ratio

BACKGROUND

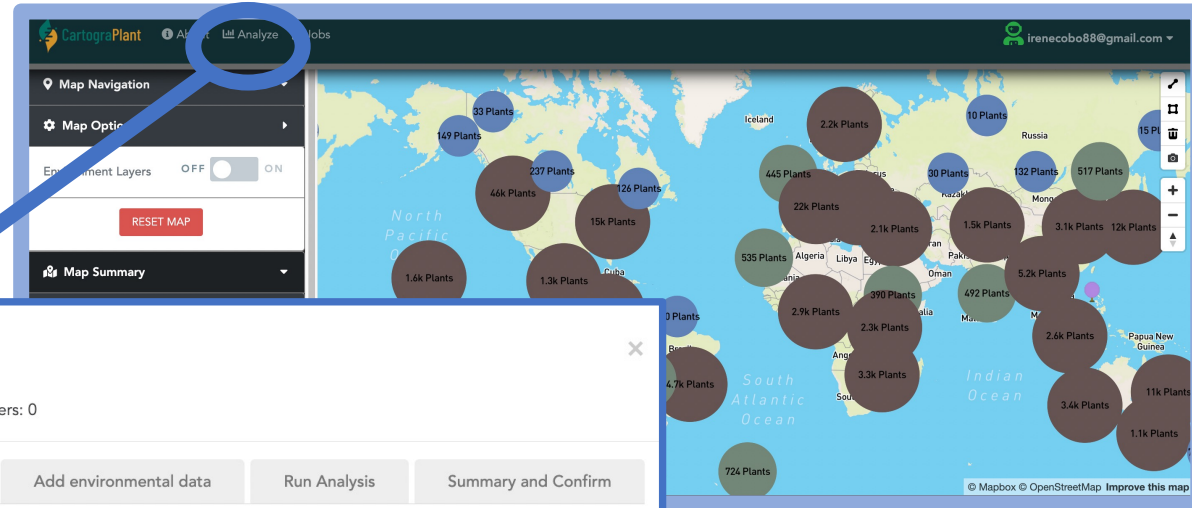
DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA ANALYSIS

<https://cartograplant.org/>



CartograPlant

Analysis ID: 1053
Phenotypes: 3
Genotypes: 0
Environmental layers: 0

Manage | Filter By Traits | **Filter By Genotypes** | Filtering & Imputation | Population Structure | Add environmental data | Run Analysis | Summary and Confirm

Filter By Genotypes
2 studies detected based on the trees you selected on the map
TGDR674 **TGDR682**
Overlapping genotypes: 24887

Insights
We analyzed 2 studies and discovered the following:
There are a total of 24887 SNP overlaps across all 2 studies

- SNP overlaps between TGDR674, TGDR682: 24887
- TGDR674 has 24887 SNPs
- TGDR682 has 33070 SNPs

TGDR674 has 0 non-overlaps
TGDR682 has 8183 non-overlaps

TGDR674 0 **24887** **TGDR682 8183**

NEXT
CLOSE

BACKGROUND

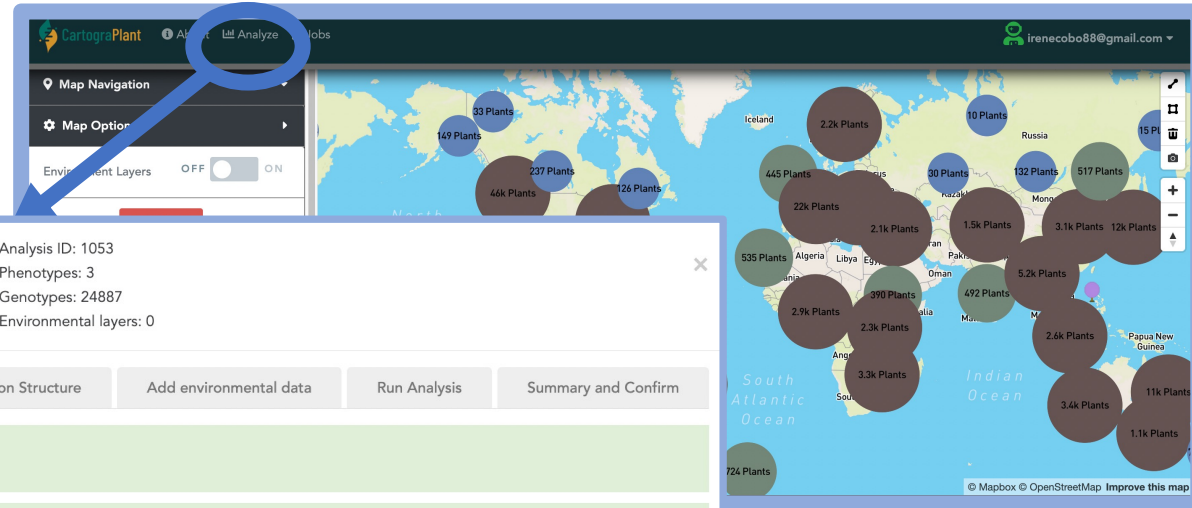
DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA ANALYSIS

<https://cartograplant.org/>



CartograPlant

Analysis ID: 1053
Phenotypes: 3
Genotypes: 24887
Environmental layers: 0

Manage | Filter By Traits | Filter By Genotypes | **Filtering & Imputation** | Population Structure | Add environmental data | Run Analysis | Summary and Confirm

🌟 VCF files successfully uploaded to workspace

✅ Found Genotype VCF for TGDR674
✅ Found Genotype VCF for TGDR682

Quality filtering method: SNP Quality Filtering Step 1.4: Missingness per individual, filtering

SNP QUALITY FILTERING STEP 1.4: MISSINGNESS PER INDIVIDUAL, FILTERING ANALYSIS CONFIGURATION

Missingness per individual, filtering

⚙️ SNP quality filtering. Step 1 (part 4): Missingness per individual, filtering

📁 Select the file to use. If files don't appear, please try refreshing the workflow

Workspace available files: 🔄

⚙️ Output from the step 1 (part 3)

📁 Select the file to use. If files don't appear, please try refreshing the workflow

BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA ANALYSIS

<https://cartograplant.org/>

The screenshot displays the CartograPlant web application interface. At the top, a navigation bar includes 'Map Navigation', 'Map Options', and 'Environment Layers' (OFF/ON). A blue circle highlights the 'Analyze' button. Below the navigation bar, a world map shows plant distribution data with bubbles of varying sizes and colors (blue, green, brown) representing different plant counts in various regions. A pop-up window in the center-left provides analysis details: Analysis ID: 1053, Phenotypes: 3, Genotypes: 24887, and Environmental layers: 0. Below this, a series of tabs includes 'Manage', 'Filter By Traits', 'Filter By Genotypes', 'Filtering & Imputation' (selected), 'Population Structure', 'Add environmental data', 'Run Analysis', and 'Summary and Confirm'. A green message box states 'VCF files successfully uploaded to workspace'. Below that, two green checkmarks indicate 'Found Genotype VCF for TGDR674' and 'Found Genotype VCF for TGDR682'. A dropdown menu for 'SNP Quality Filtering' is open, listing steps 1.1 through 1.4, with 'SNP Quality Filtering Step 3: Minimum quality score' selected. The interface also includes 'Workspace available files' and 'Output from the step 1' sections, each with a 'Select the file to use' button. A refresh button and a message 'If files don't appear, please try refreshing the workflow' are visible at the bottom right.

BACKGROUND

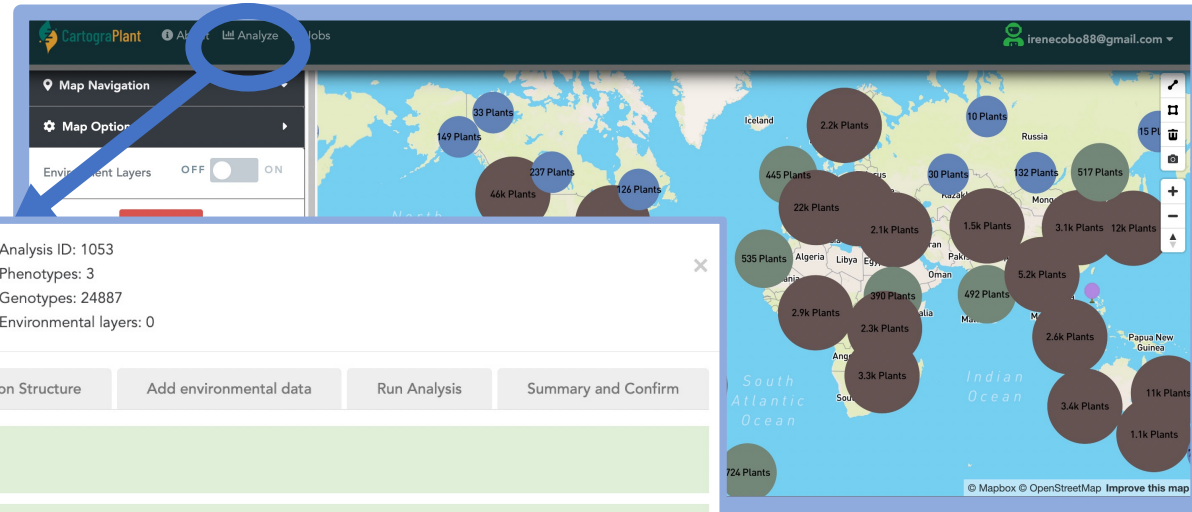
DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA ANALYSIS

<https://cartograplant.org/>



CartograPlant

Analysis ID: 1053
Phenotypes: 3
Genotypes: 24887
Environmental layers: 0

Manage | Filter By Traits | Filter By Genotypes | **Filtering & Imputation** | Population Structure | Add environmental data | Run Analysis | Summary and Confirm

🌟 VCF files successfully uploaded to workspace

✅ Found Genotype VCF for TGDR674
✅ Found Genotype VCF for TGDR682

Quality filtering method: SNP Quality Filtering Step 1.4: Missingness per individual, filtering

SNP QUALITY FILTERING STEP 1.4: MISSINGNESS PER INDIVIDUAL, FILTERING ANALYSIS CONFIGURATION

Missingness per individual, filtering

⚙️ SNP quality filtering. Step 1 (part 4): Missingness per individual, filtering

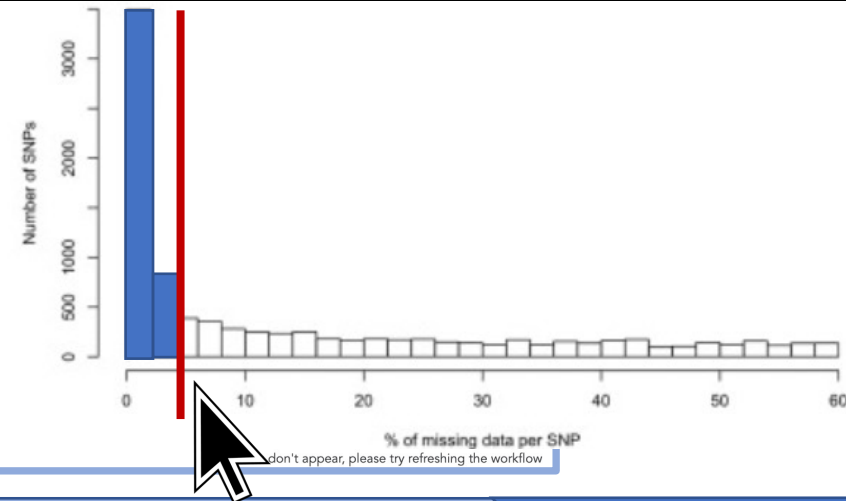
📁 Select the file to use.

Workspace available files:

⚙️ Output from the step 1 (part 3)

📁 Select the file to use.

Histograms to help decide the quality filtering thresholds available soon!



BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA ANALYSIS

<https://cartograplant.org/>

The image shows a screenshot of the CartograPlant web application. The top part features a world map with numerous circular markers of varying sizes and colors (blue, green, brown) representing plant populations. Each marker is labeled with the number of plants, such as '33 Plants', '149 Plants', '237 Plants', '44k Plants', '15k Plants', '126 Plants', '1.6k Plants', '1.3k Plants', '2.2k Plants', '10 Plants', '15 Plants', '445 Plants', '22k Plants', '2.1k Plants', '1.5k Plants', '30 Plants', '132 Plants', '517 Plants', '3.1k Plants', '12k Plants', '535 Plants', '2.9k Plants', '390 Plants', '492 Plants', '5.2k Plants', '2.3k Plants', '2.6k Plants', '3.3k Plants', '3.4k Plants', '11k Plants', and '1.1k Plants'. The map interface includes a sidebar with 'Map Navigation', 'Map Options', 'Environment Layers' (OFF/ON), and 'Map Summary'. A blue circle highlights the 'Analyze' button in the top navigation bar, with a blue arrow pointing to the analysis configuration panel below.

The analysis configuration panel, titled 'CartograPlant', displays the following information:

- Analysis ID: 1053
- Phenotypes: 3
- Genotypes: 24887
- Environmental layers: 0

The panel contains a series of tabs: Manage, Filter By Traits, Filter By Genotypes, Filtering & Imputation, Population Structure (selected), Add environmental data, Run Analysis, and Summary and Confirm. Under the 'Population Structure' tab, the 'Number of populations' is set to 2. Below this, there is a 'Select VCF file' dropdown menu with the text 'SELECT VCF FILE'. A blue button labeled 'GENERATE FAST STRUCTURE' is highlighted with a mouse cursor. At the bottom right of the panel, there are 'NEXT' and 'CLOSE' buttons.

BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA ANALYSIS

<https://cartograplant.org/>

The screenshot displays the CartograPlant web application. At the top, a navigation bar includes 'Analyze' and 'Jobs'. A sidebar on the left contains 'Map Navigation', 'Map Options', 'Environment Layers' (OFF/ON), and 'Map Summary'. The main area shows a world map with circular markers indicating the number of plants in various regions, such as 33 Plants in North America and 1.6k Plants in South America. A detailed analysis window is open, showing the following information:

- Analysis ID: 1053
- Phenotypes: 3
- Genotypes: 24887
- Environmental layers: 0

The analysis window features a menu with options: Manage, Filter By Traits, Filter By Genotypes, Filtering & Imputation, Population Structure (selected), Add environmental data, Run Analysis, and Summary and Confirm. The 'Population Structure' section includes a 'Number of populations' dropdown set to 2, a 'Select VCF file' button, and a 'GENERATE FAST STRUCTURE' button. Below this is a PCA plot with PC1 on the x-axis and PC2 on the y-axis. To the right of the plot are four horizontal bar charts for K=2, K=3, K=4, and K=5, representing different numbers of clusters. A legend indicates three populations: WS (green), QI (red), and Tp (blue).

Population structure calculation (PCA and DAPC) and visualization (fastSTRUCTURE, PCA and DAPC) available soon!!!

BACKGROUND

DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA ANALYSIS

<https://cartograplant.org/>

CartograPlant

Analyze Jobs

irenecho88@gmail.com

Analysis: 1015
Phenotypes: 3
Genotypes: 32,007
Environmental layers: 0

Manage Filter By Traits Filter By Genotypes Filtering & Imputation Population Structure **Add environmental data** Run Analysis Summary and Confirm

Choose environmental layers

- CATEGORY US
- GROUP Biotic Damage (North America)
- GROUP Climatic variables (World, ClimateWNA)
 - PROPERTY Annual Heat Moisture Index CWNA (AHM)
 - PROPERTY Cooling degree-days CWNA (DD18)
 - PROPERTY Summer Heat Moisture Index (CWNA) (SHM)
 - PROPERTY Temperature difference between MWMT and MCMT CWNA (TD)
 - PROPERTY Mean Annual Temperature CWNA (MEAN_ANNUAL_TEMP)
 - PROPERTY Day when FFP begins CWNA (BFFP)
 - PROPERTY Growing Degree Days Above 5°C CWNA (DD5)
 - PROPERTY Extreme maximum temperature over 30 years CWNA (EXT)
 - PROPERTY Mean warmest monthly temperature CWNA (MWMT)
 - PROPERTY Chilling Days Below 0 °C CWNA (DD0)
 - PROPERTY Frost-free period CWNA (FFP)
 - PROPERTY The day when FFP ends CWNA (EFFP)
 - PROPERTY Hargreaves climatic moisture deficit CWNA (CMD)
 - PROPERTY Mean Annual Precipitation CWNA (MEAN_ANNUAL_PREC)
 - PROPERTY Heating degree-days CWNA (DD_18)
 - PROPERTY Snow precipitation CWNA (PAS)

PRECACHE VALUES

GATHER AND UPLOAD TO WORKSPACE

Found 145 Mean Annual Precipitation CWNA (MEAN_ANNUAL_PREC) values.

0.29

Annual Temperature CWNA (MEAN_ANNUAL_TEMP)

Annual Precipitation CWNA (MEAN_ANNUAL_PREC)

BACKGROUND

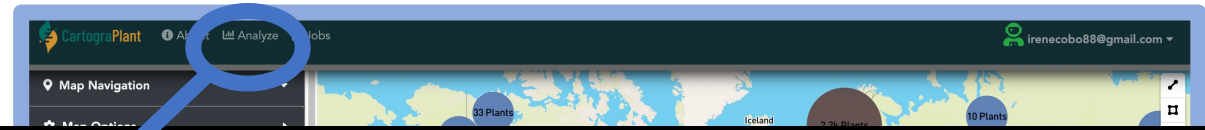
DATA
INTEGRATION

DATA
VISUALIZATION

DATA ANALYSIS

DATA ANALYSIS

<https://cartograplant.org/>



CartograPlant

Analysis ID: 1015
Phenotypes: 3
Genotypes: 32597
Environmental layers: 0

Manage | Filter By Traits | Filter By Genotypes | Filtering & Imputation | Population Structure | Add environmental data | Run Analysis | Summary and Confirm

Step 1 - Select workflow: LinkImputeR Step 1 - Accuracy Mode ▾

Step 2 - Setup analysis

REFRESH

NEXT

CLOSE



DATA ANALYSIS

<https://cartograplant.org/>

Analysis ID: 1015
Phenotypes: 3
Genotypes: 32597
Environmental layers: 0

Manage Filter By Traits Filter By Genotypes Filtering & Imputation Population Structure Add environmental data Run Analysis Summary and Confirm

Step 1 - Select workflow
Step 2 - Setup analysis

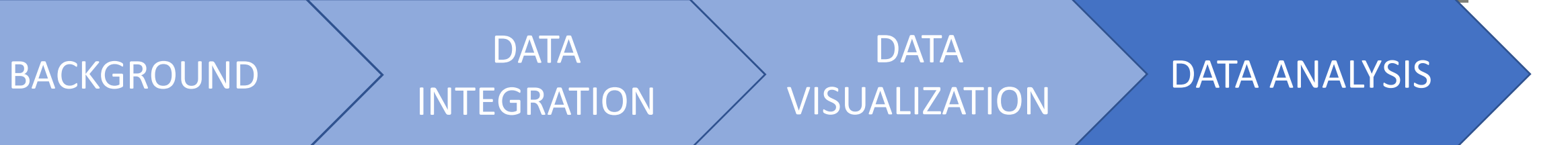
GWAS with EMMAX
Landscape genomics with Sambac
Landscape genomics with Bayenv
Multiple testing correction (FDR)
BLUP calculation (phenotypes)
Meta-analysis with METASOFT
Meta-analysis with PLINK

REFRESH

NEXT

CLOSE

Other analytic workflows available soon!



DATA ANALYSIS

<https://cartograplant.org/>

Analysis ID: 1053
Phenotypes: 3
Genotypes: 24887
Environmental layers: 0

Manage Filter By Traits Filter By Genotypes Filtering & Imputation Population Structure Add environmental data Run Analysis Summary and Confirm

Step 1 - Select workflow GWAS with EMMAX REFRESH

Step 2 - Setup analysis

GWAS WITH EMMAX ANALYSIS CONFIGURATION

This GWAS pipeline performs association mapping using the EMMAX software, a mixed model accounting for the sample structure. In addition to the computational efficiency obtained by EMMA algorithm, EMMAX takes advantage of the fact that each loci explains only a small fraction of complex traits, which allows us to avoid repetitive variance component estimation procedure, resulting in a significant amount of increase in computational time of association mapping using mixed model. EMMAX is a linear mixed model (LMM) to perform GWAS, correcting by a wide range of sample structures (which encompasses population stratification and hidden relatedness). To this end, it uses an internally calculated kinship matrix as a random effect, and population structure externally calculated as a fixed effect. More information about EMMAX program can be found here: <https://genome.sph.umich.edu/wiki/EMMAX>

Genotype file containing SNPs in vcf or vcf.gz format

Select the file to use.

Seleccionar archivo Ninguno ...hivo selec. Upload to workspace

Workspace available files: Analysis 1053: Filtered Phenotypes - bud set - ADJUSTED (10 minutes ago) v

LD filtered vcf

Select the file to use.

Seleccionar archivo Ninguno ...hivo selec. Upload to workspace

Workspace available files: Analysis 1053: Filtered Phenotypes - bud set - ADJUSTED (10 minutes ago) v



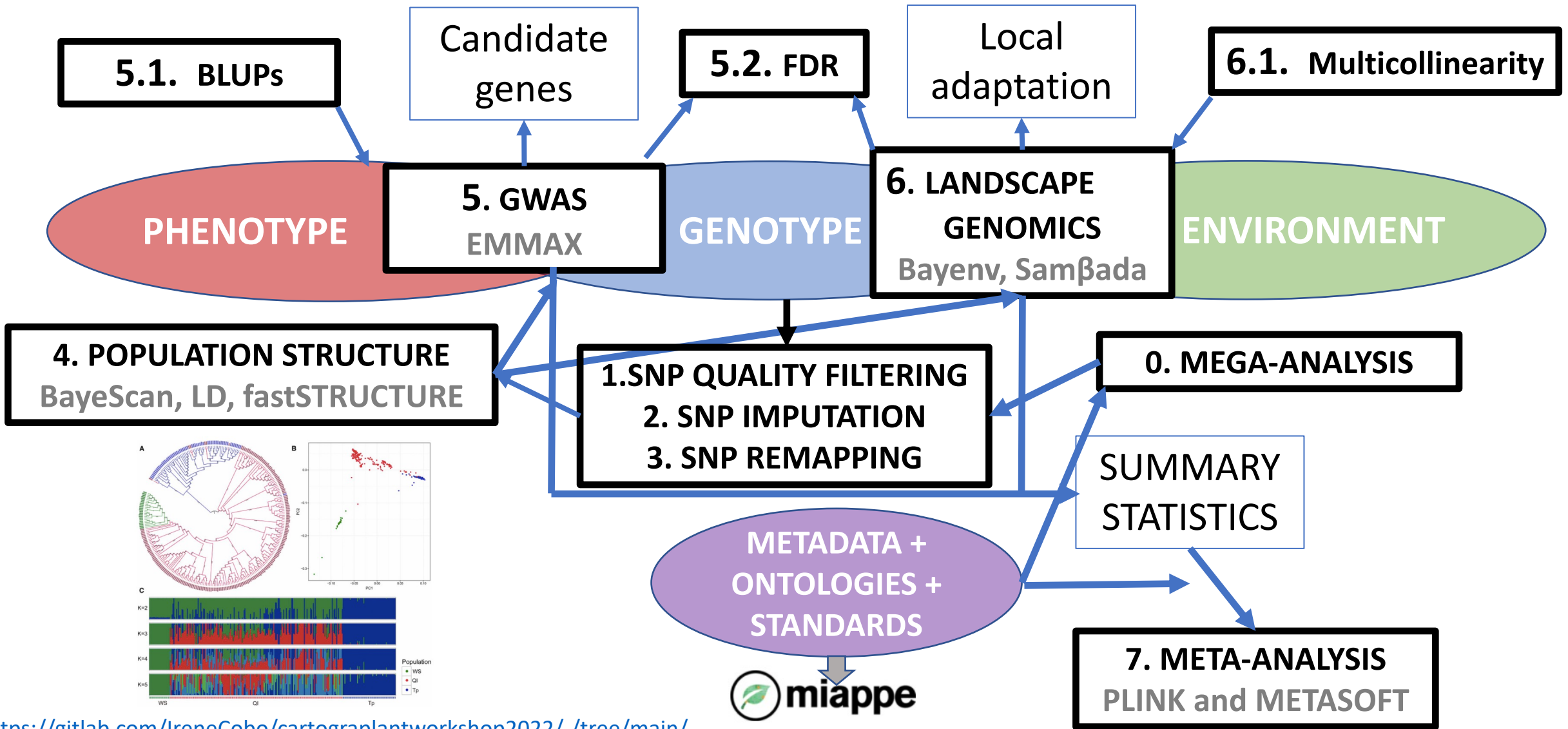
BACKGROUND

DATA
INTEGRATION

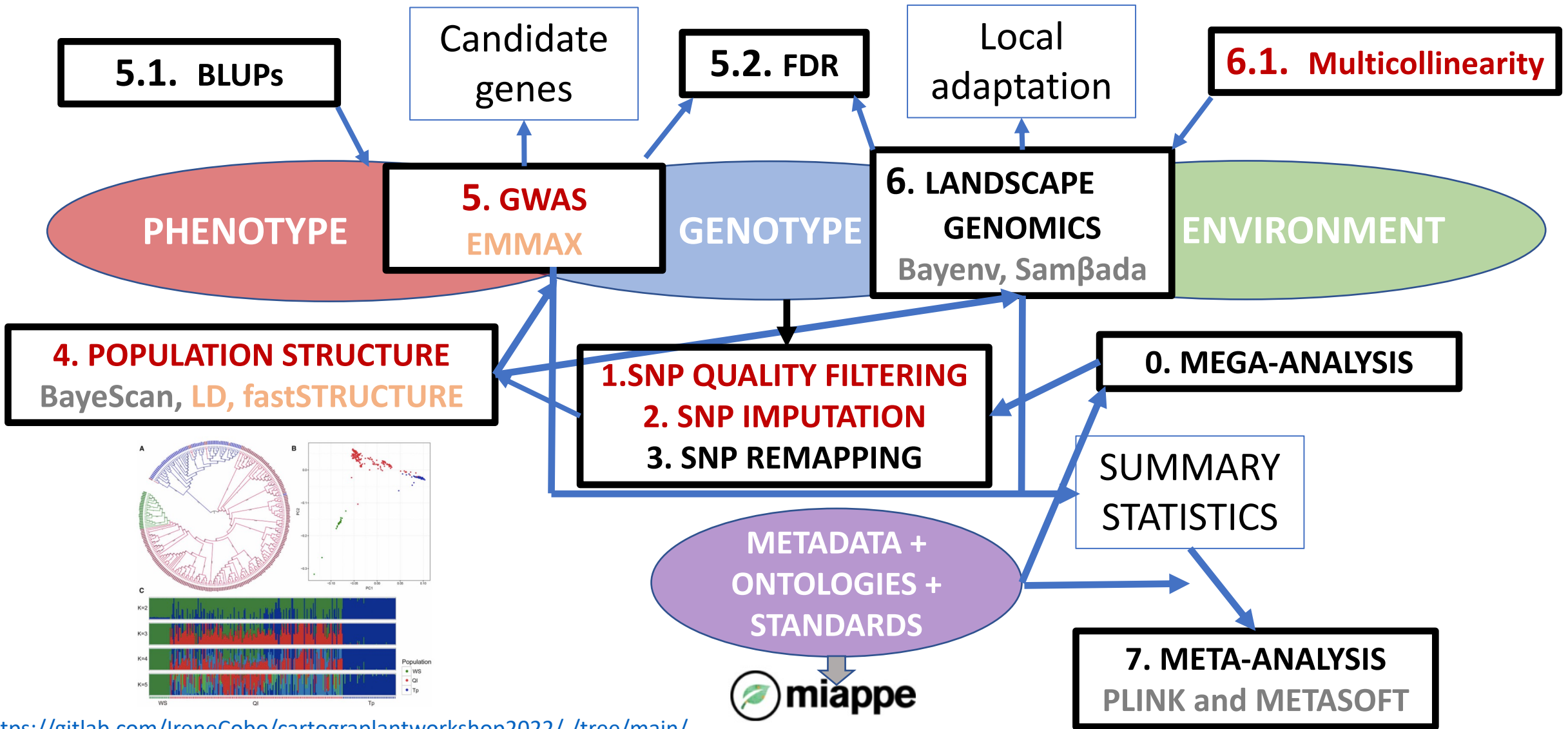
DATA
VISUALIZATION

DATA ANALYSIS

CARTOGRAPLANT WORKFLOWS OVERVIEW

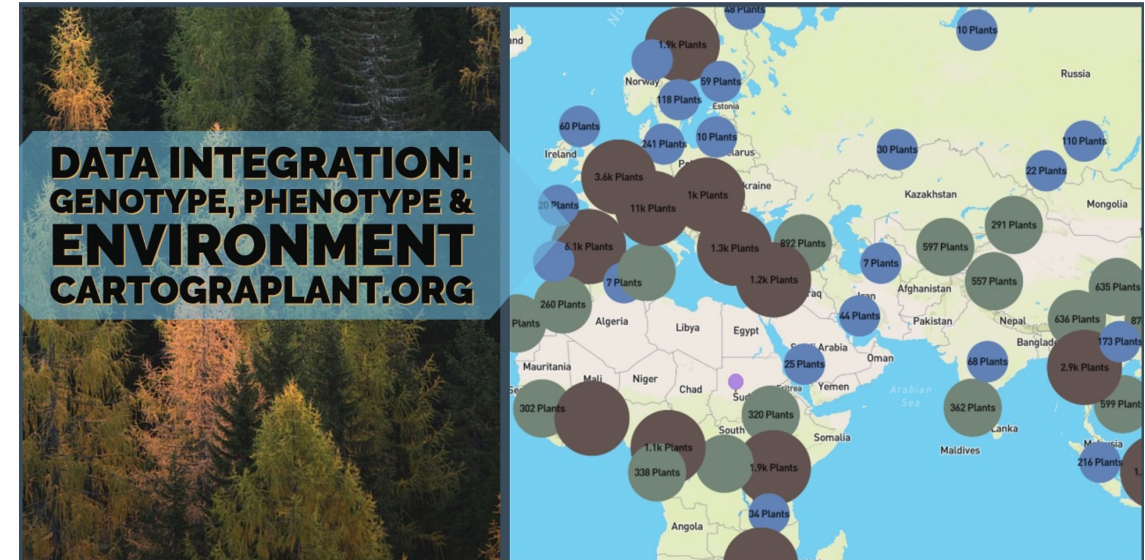


CARTOGRAPLANT WORKFLOWS OVERVIEW



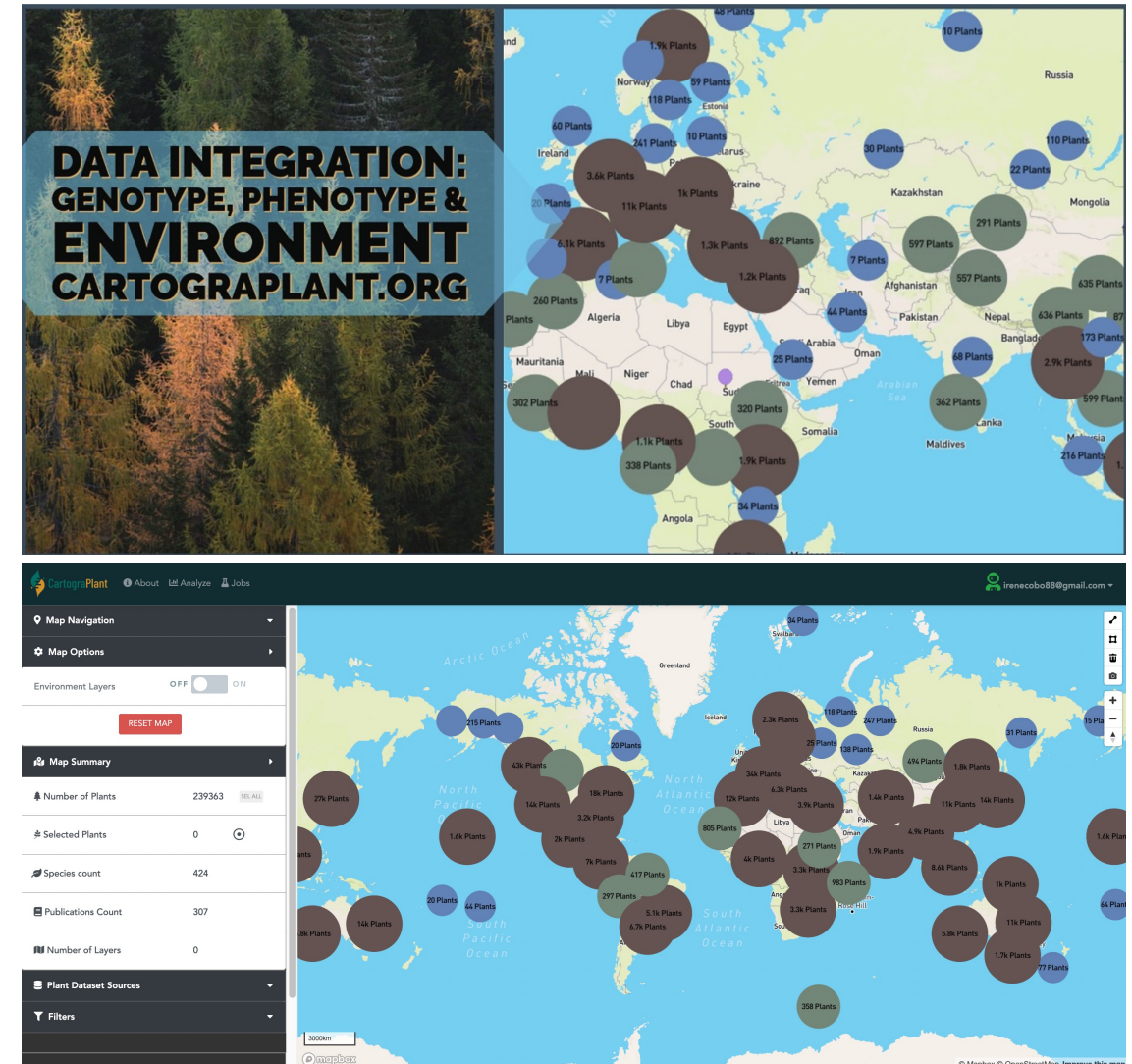
CONCLUSIONS

- **CartograPlant** is intended to serve as a **community resource for Plant Molecular Ecology**.
- **These flexible analytic workflows** allow to analyze a diversity of data types (e.g. SNPs, SSR) and experimental designs (e.g. natural populations, common garden) and facilitate a **diversity of biological questions in CartograPlant**.



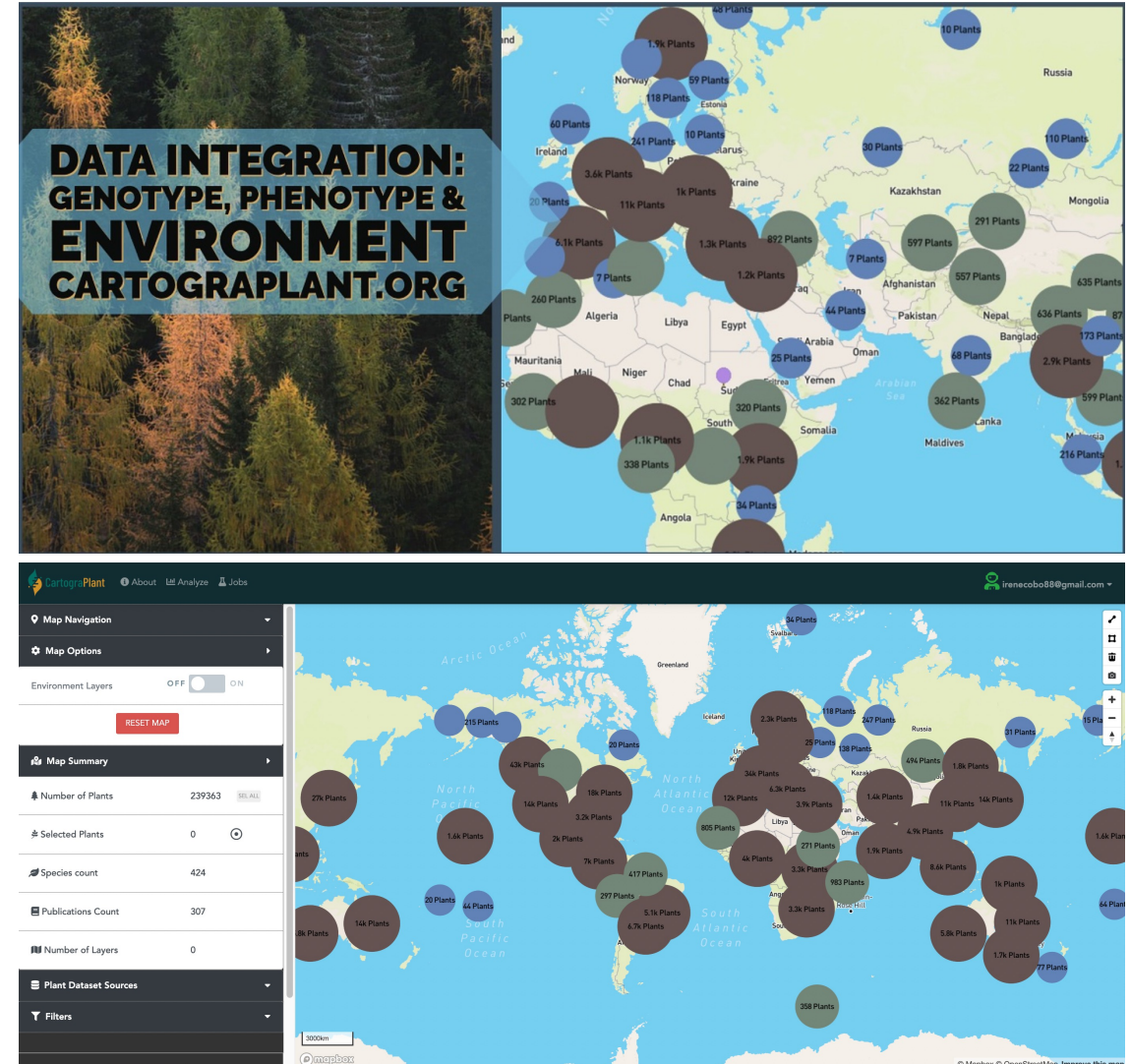
CONCLUSIONS

- **CartograPlant** is intended to serve as a **community resource for Plant Molecular Ecology**.
- **These flexible analytic workflows** allow to analyze a diversity of data types (e.g. SNPs, SSR) and experimental designs (e.g. natural populations, common garden) and facilitate a **diversity of biological questions in CartograPlant**.
- **Mega and meta-analysis take advantage of one of the main strengths of CartograPlant**: the curation and integration of a diversity of data types (genotypic, phenotypic and environmental) from different studies, thanks to the metadata collection using ontologies and standards.
- **CartograPlant is continuously growing to stay up-to-date** with the latest research and statistical methods to analyze high-throughput biological data.



CONCLUSIONS

- **CartograPlant** is intended to serve as a **community resource for Plant Molecular Ecology**.
- **These flexible analytic workflows** allow to analyze a diversity of data types (e.g. SNPs, SSR) and experimental designs (e.g. natural populations, common garden) and facilitate a **diversity of biological questions in CartograPlant**.
- **Mega and meta-analysis take advantage of one of the main strengths of CartograPlant**: the curation and integration of a diversity of data types (genotypic, phenotypic and environmental) from different studies, thanks to the metadata collection using ontologies and standards.
- **CartograPlant is continuously growing to stay up-to-date** with the latest research and statistical methods to analyze high-throughput biological data.
- Having a **centralized and up-to-date platform** to integrate, visualize and analyze high-throughput biological data is **key in the current big data era in plant biology**.



TEAM MEMBERS! QUESTIONS?

Members of the project



- Stephen P. Ficklin



- Nic Herndon



- Emily Grau
- Sean Buehler
- Shay Muhonen
- Risharde Ramnath
- Umed Singh
- Charles Demurjian
- Meghan Myles
- Emily Strickland
- Victoria Burton
- Maddie Gadomski
- Jill Wegrzyn



- Margaret Staton
- Abdullah Almsaeed

This project has been funded by
USDA-NIFA #2018-09223



@Irene_Cobo
@TreeGenes

<https://cartograplant.org/>