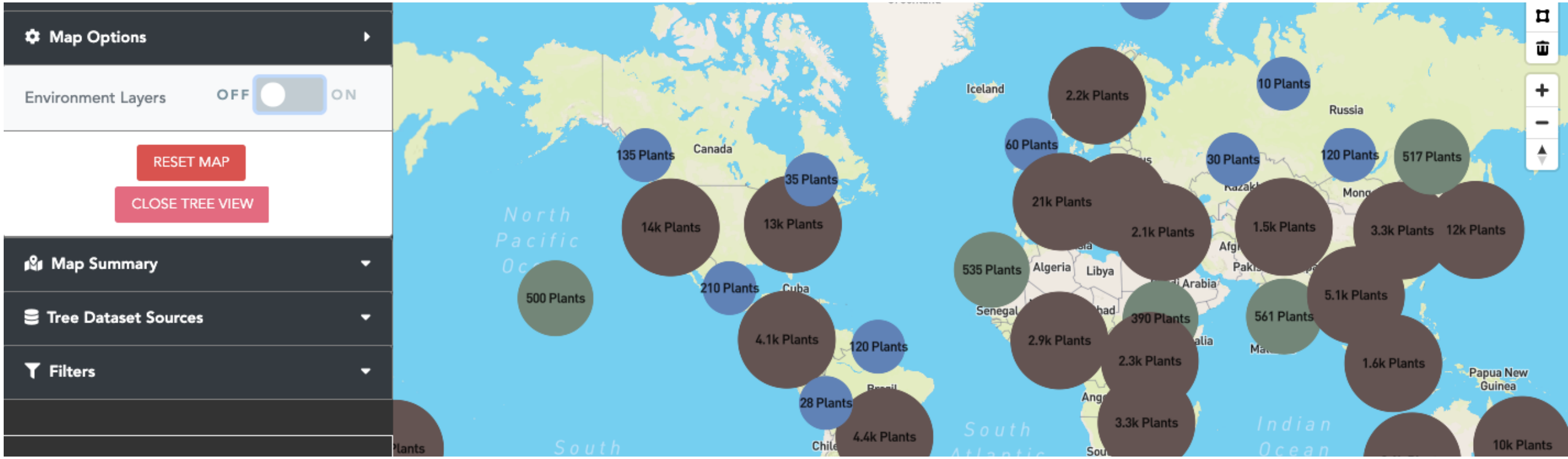
CartograPlant: Cyberinfrastructure to improve plant health and productivity in the context of a changing climate



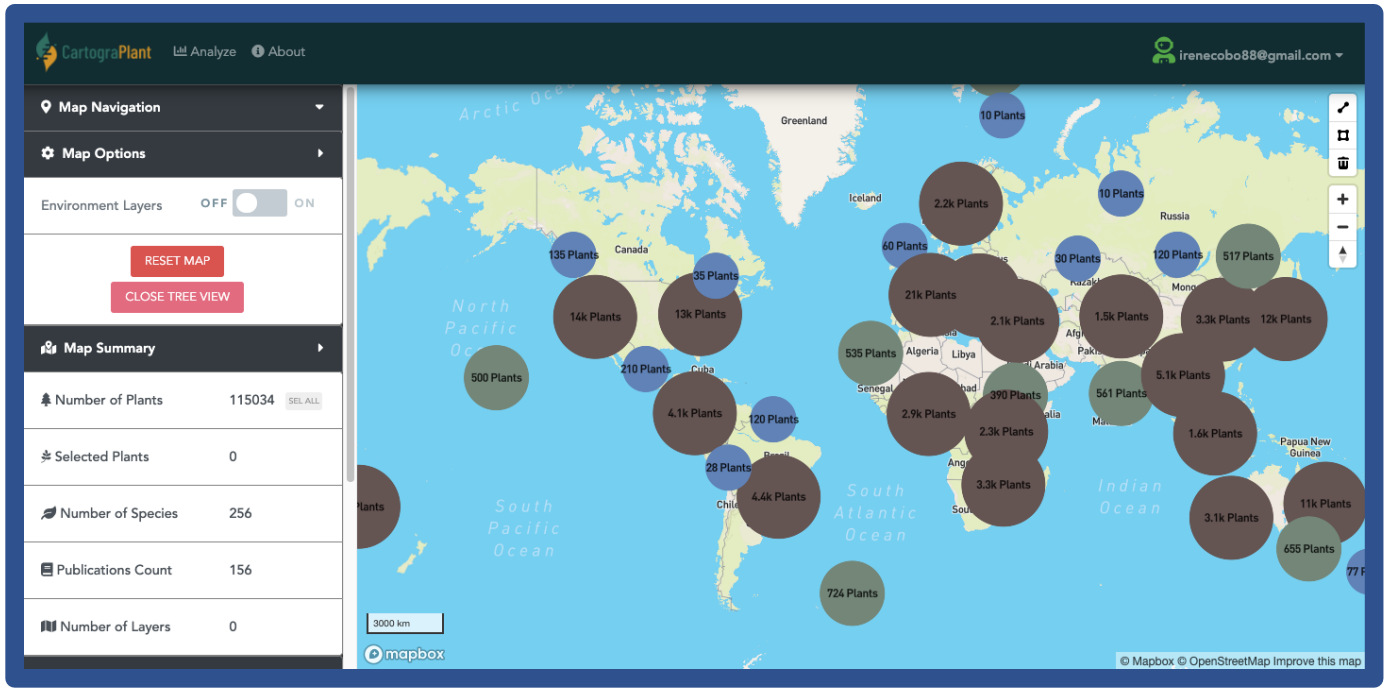
Irene Cobo Simón, PhD.

Postdoctoral Fellow

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

UCONN
UNIVERSITY OF CONNECTICUT

What is CartograPlant?



https://cartograplant.org/

INTEGRATES

VISUALIZES

ANALYZES

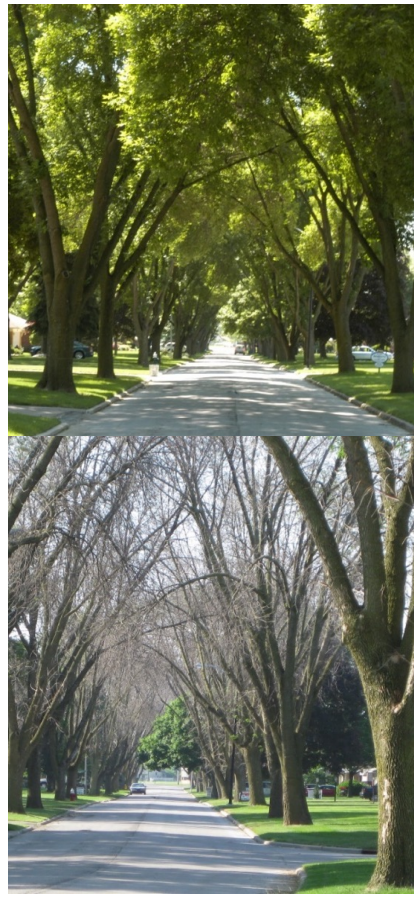
- Genotypes
- Phenotypes
- Environments

From georeferenced plants (including forest trees)

BACKGROUND → DATA INTEGRATION → DATA VISUALIZATION → DATA ANALYSIS

Why is CartograPlant relevant and timely?

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



Green ashes affected by the pest emerald ash borer

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

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

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

Why is CartograPlant relevant and timely?

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

Green ashes affected by the pest emerald ash borer

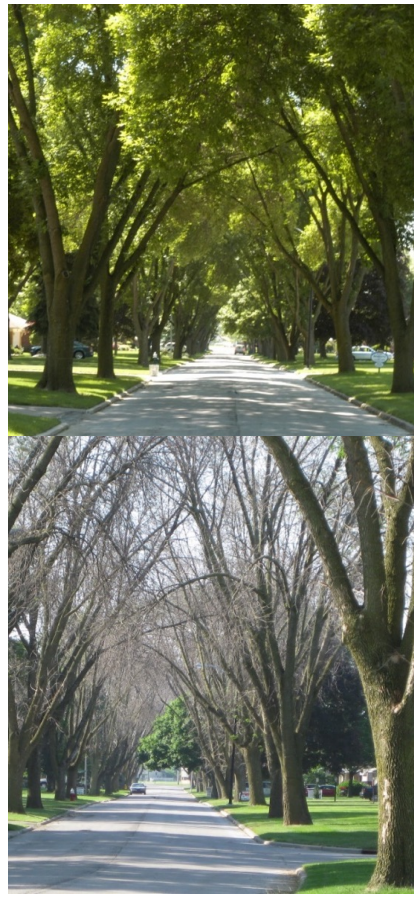
Why is CartograPlant relevant and timely?

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

Green ashes affected by the pest emerald ash borer

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**

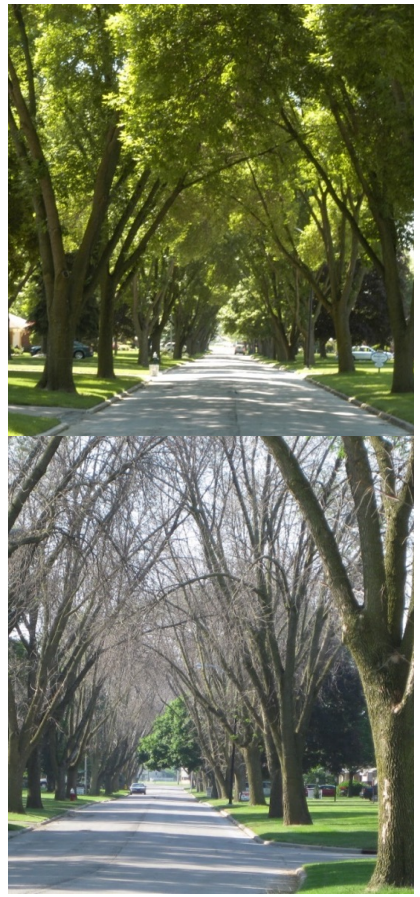


Green ashes affected by the pest emerald ash borer

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

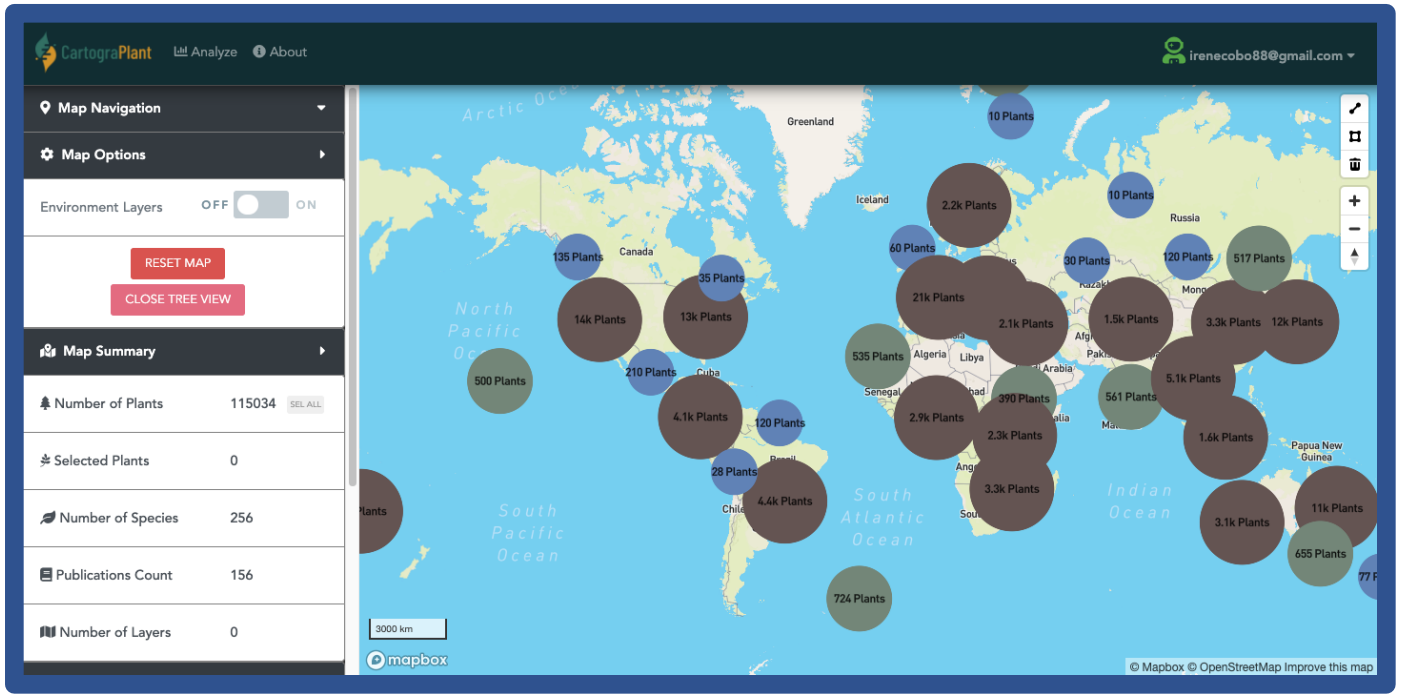


Green ashes affected by the pest emerald ash borer

DATA TYPES INTEGRATED IN CARTOGRAPLANT

1. RAW DATA

- **GENOTYPIC**
- **PHENOTYPIC**
- **ENVIRONMENTAL**



BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA TYPES INTEGRATED IN CARTOGRAPLANT

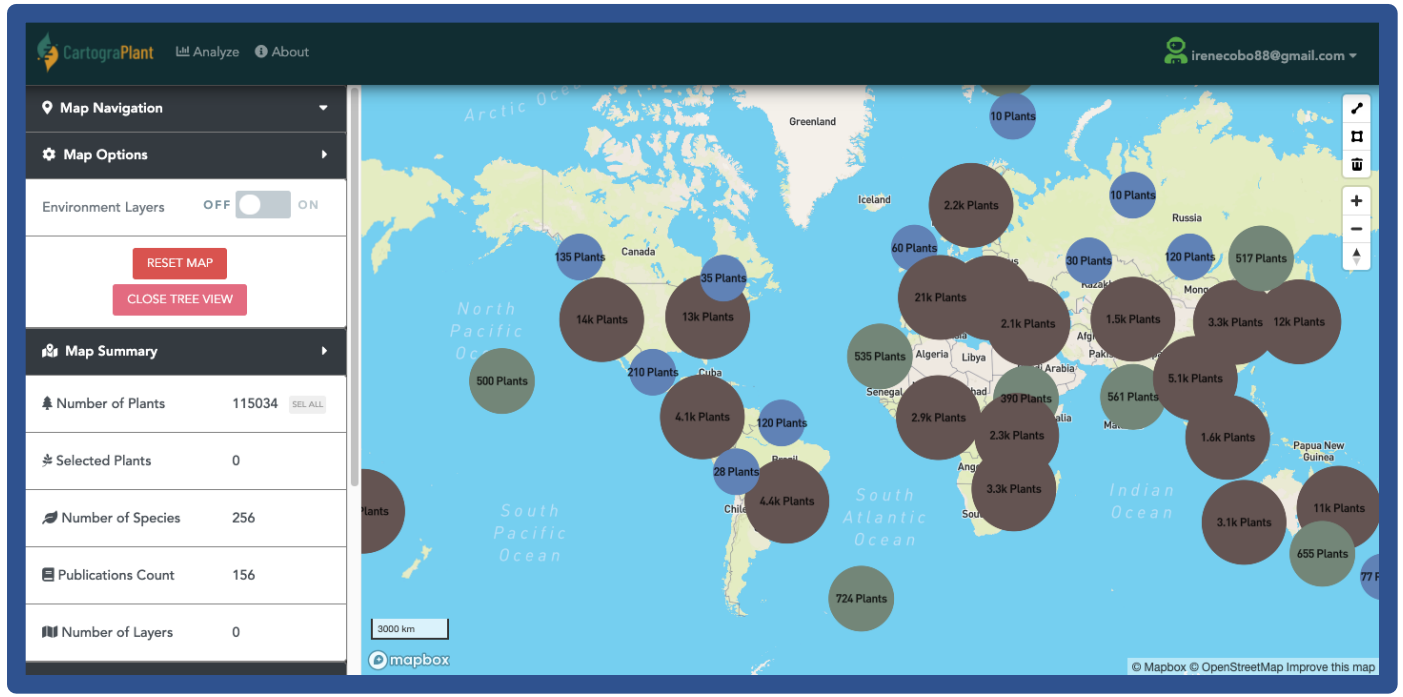
1. RAW DATA

GENOTYPIC

PHENOTYPIC

ENVIRONMENTAL

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



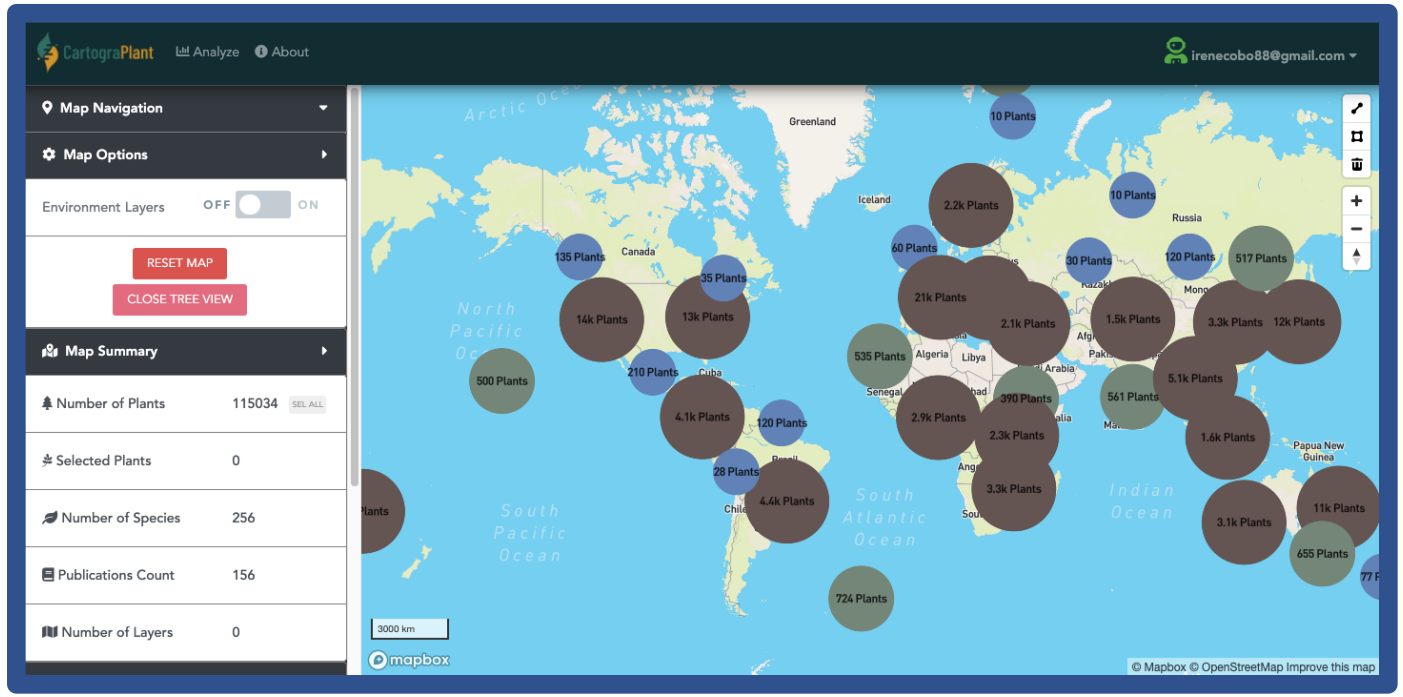
BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA TYPES INTEGRATED IN CARTOGRAPLANT

1. RAW DATA

- **GENOTYPIC**
- **PHENOTYPIC**
- **ENVIRONMENTAL**

2. METADATA + ONTOLOGIES + STANDARDS



BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

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

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

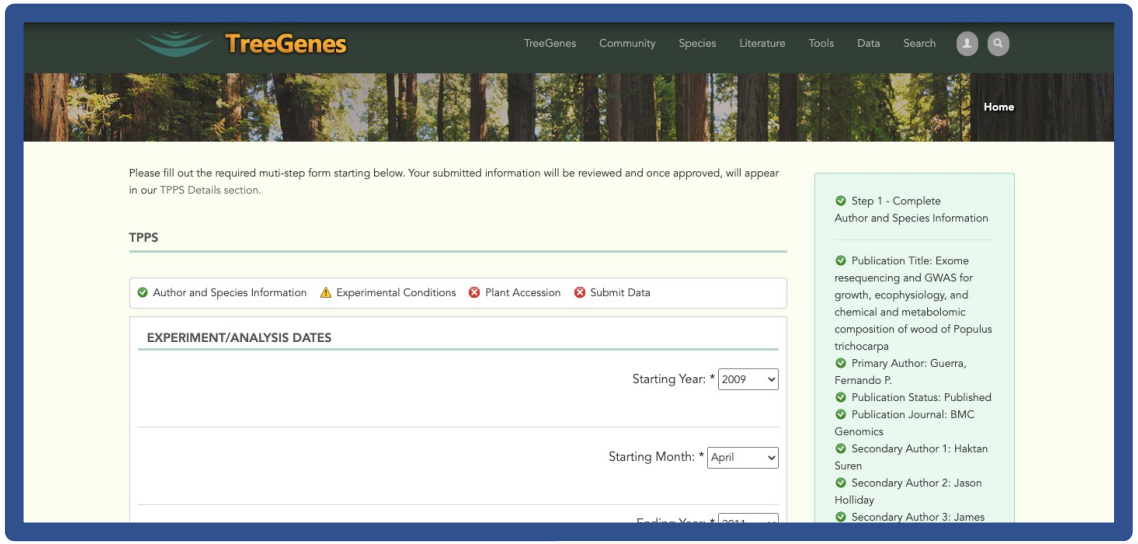
Biocuration efforts of affiliated databases

TreeGenes, DRYAD, TreeSnap, BIEN, WildType, WORLD FOREST ID

Direct submission: Tripal Plant PopGen Submit (TPPS) pipeline

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

https://treegenesdb.org/tpps

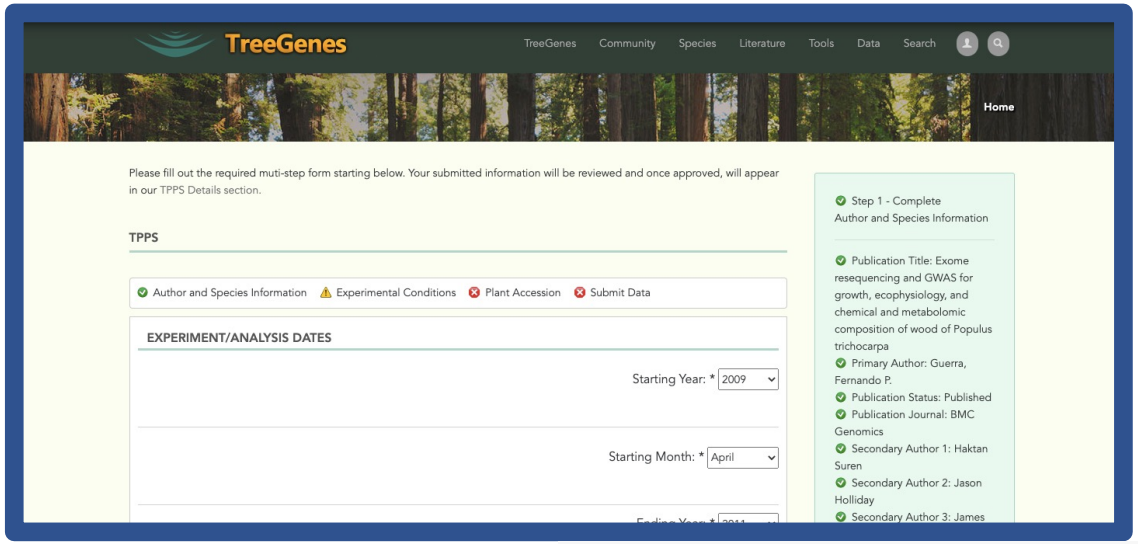


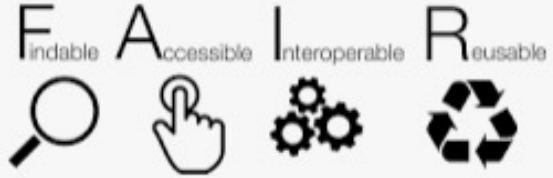
miappe

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

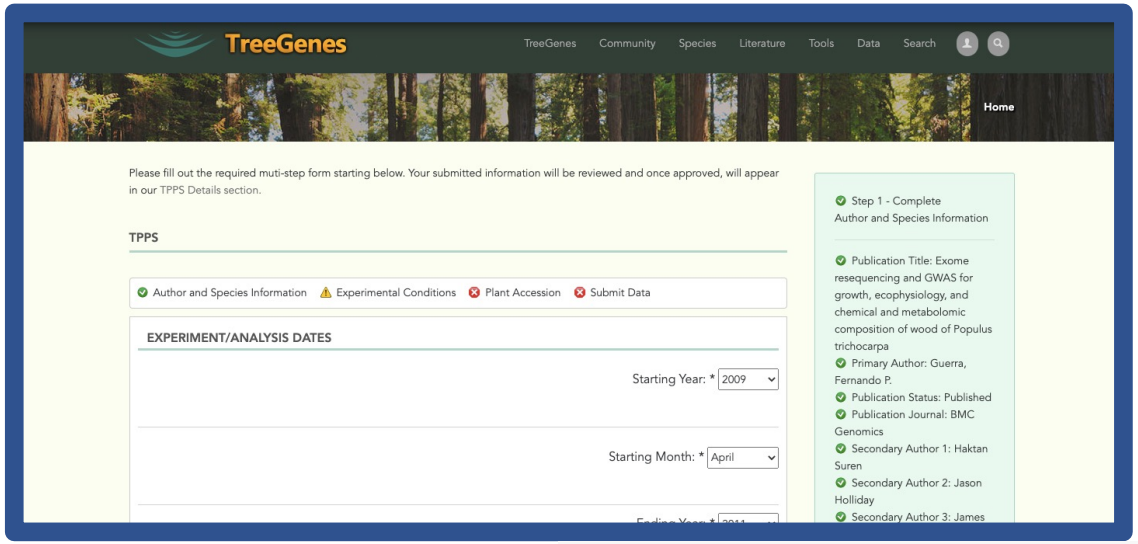




Direct submission: Tripal Plant PopGen Submit (TPPS) pipeline

https://treegenesdb.org/tpps

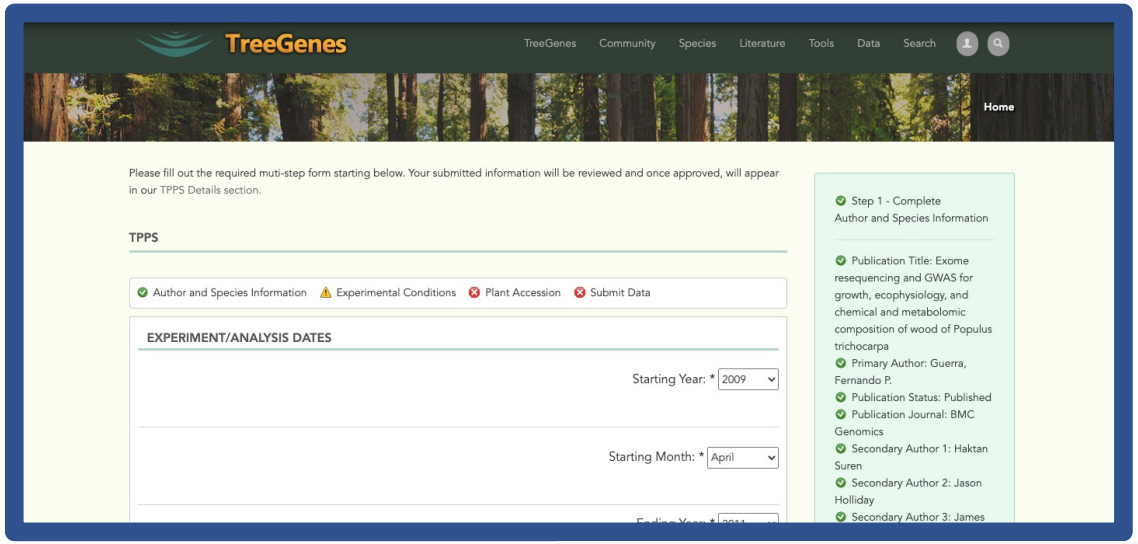
- **Genotype, phenotype, environmental data** and **metadata** and provides a **DOI**
- **Population genomics, association mapping, and landscape genomic** studies
- **Metadata** is collected using **ontologies and standards** (**MIAPPE**)

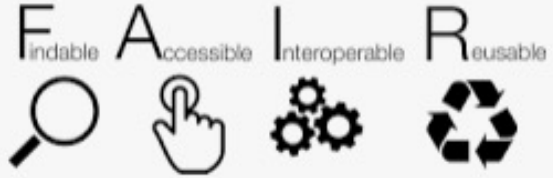


Direct submission: Tripal Plant PopGen Submit (TPPS) pipeline

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

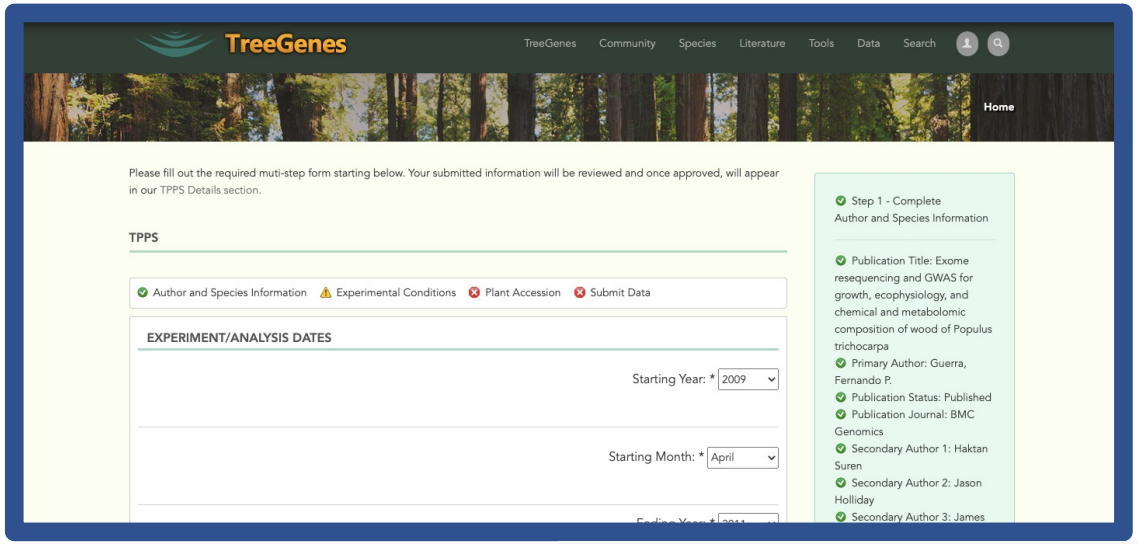


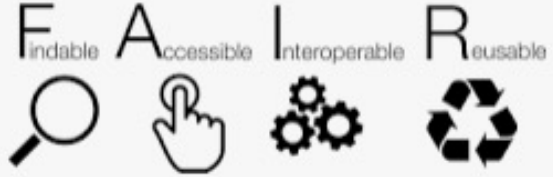


Direct submission: Tripal Plant PopGen Submit (TPPS) pipeline

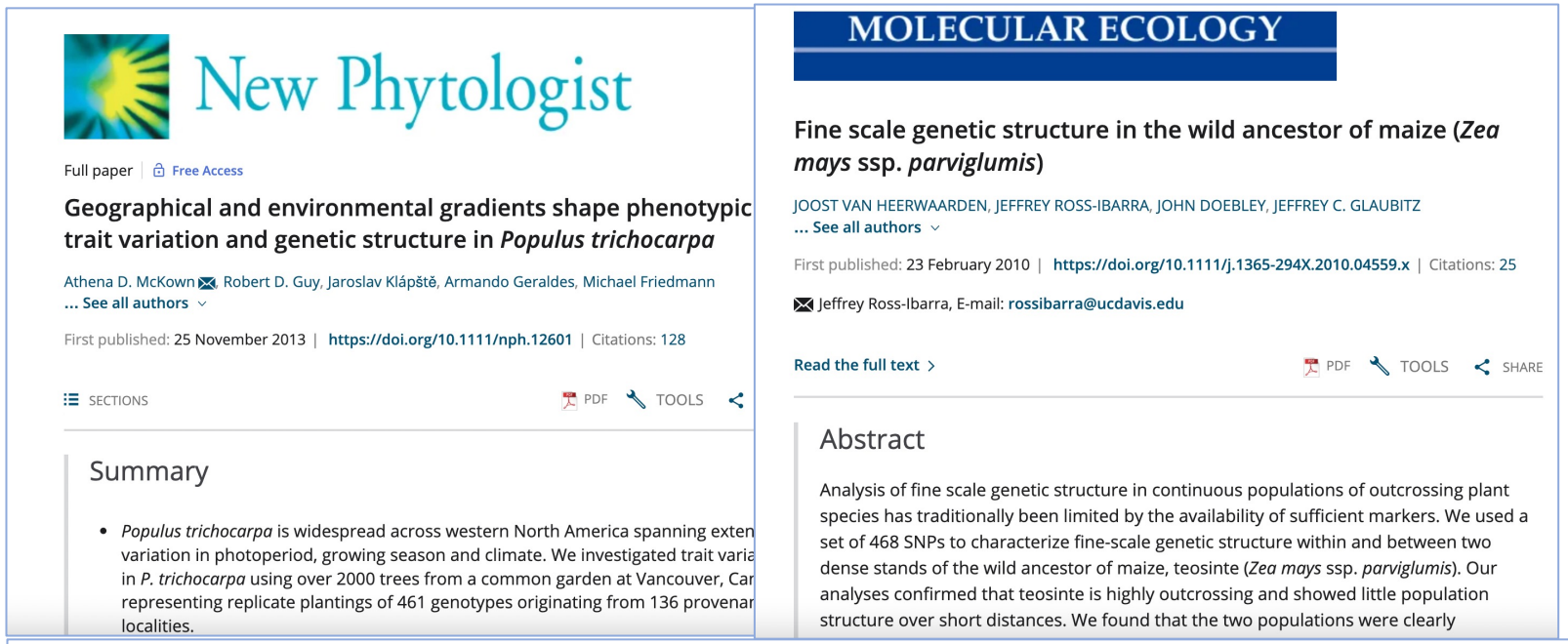
https://treegenesdb.org/tpps

- **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)





Direct submission: Tripal Plant PopGen Submit (TPPS) pipeline



New Phytologist

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 Geraldes, 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 exten variation in photoperiod, growing season and climate. We investigated trait varia in *P. trichocarpa* using over 2000 trees from a common garden at Vancouver, Ca representing replicate plantings of 461 genotypes originating from 136 provenar 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

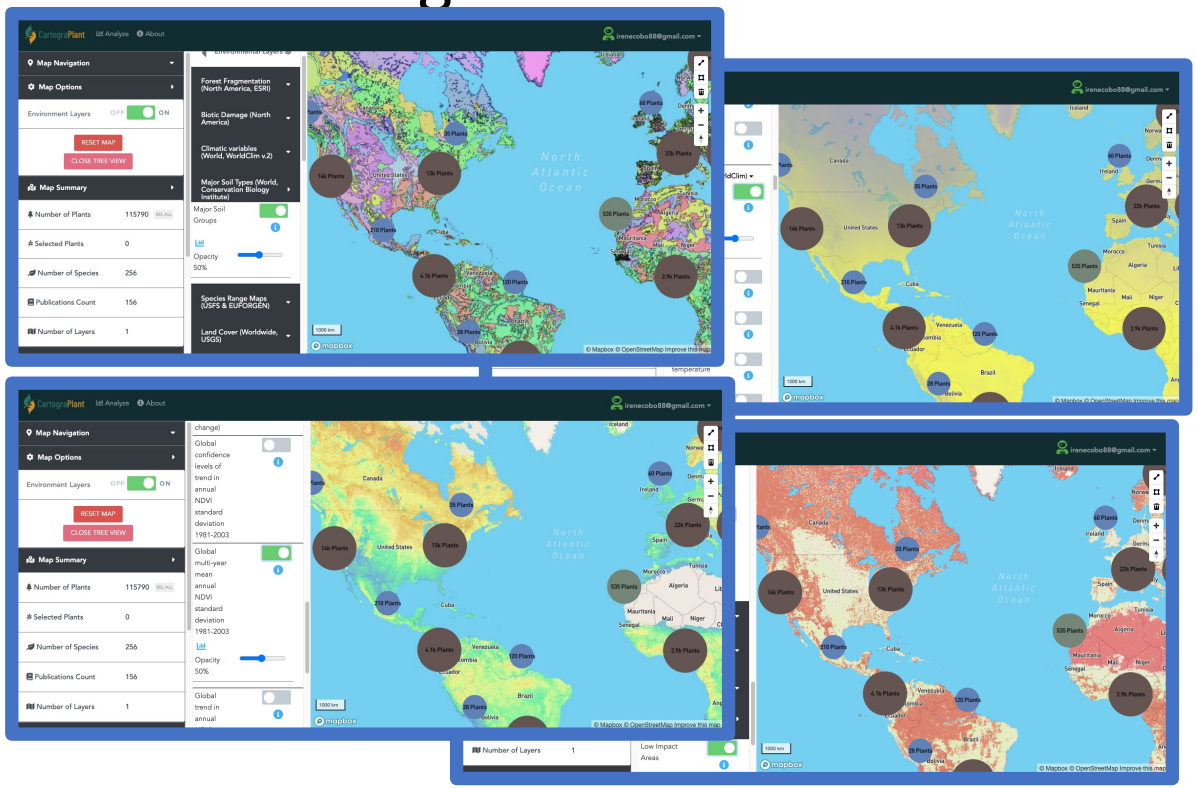
Maddie Gadomski

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**

Environmental layers

950 environmental layers are now available in CartograPlant

- CLIMATE DATA
- ECOREGIONS
- SOIL TYPE
- FOREST FRAGMENTATION – CANOPY HEIGHT
- NEON FIELD STATIONS
- SEED ZONES
- BIODIVERSITY HOTSPOTS
- WORLD FOREST ID DATA
- 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 DATA
- POPULATION DENSITY



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

DATA VISUALIZATION

LEFT PANEL

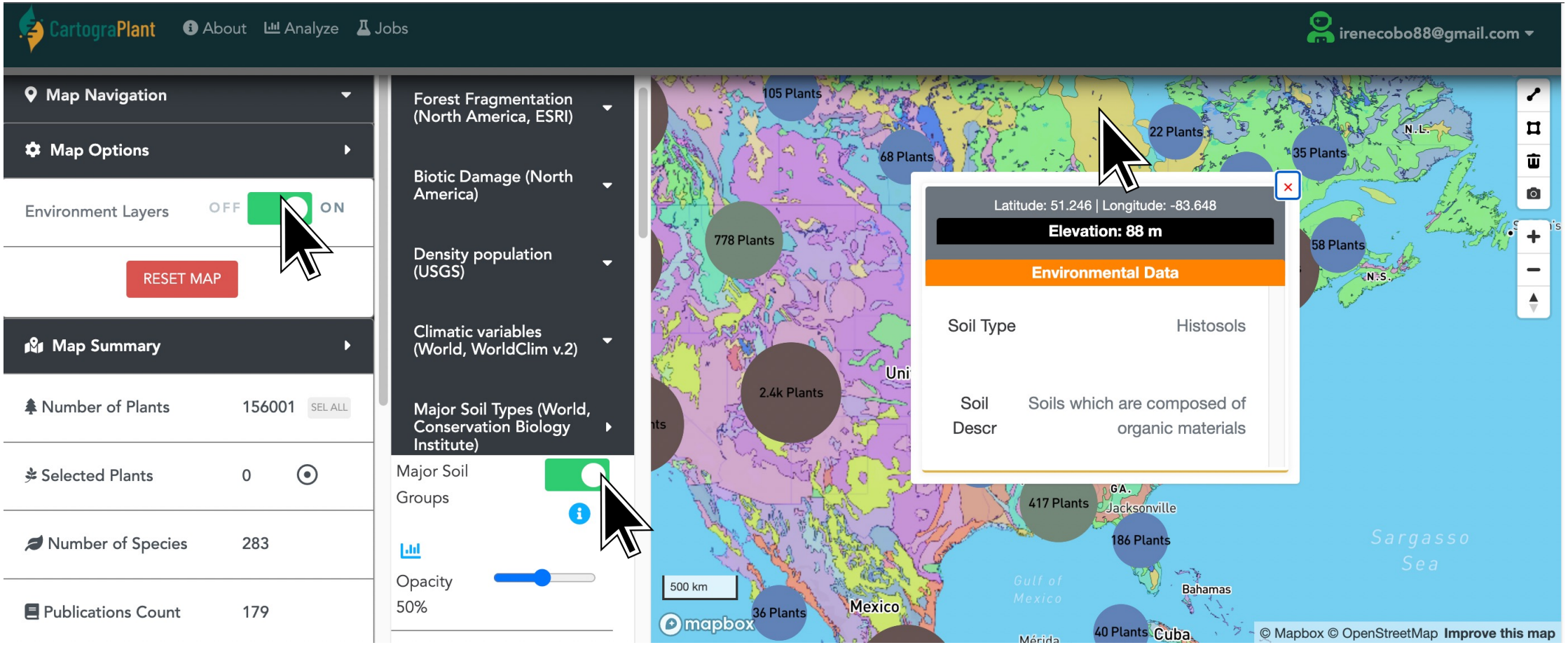
Action panel to interact with the map and plants, located to the left of the screen

RIGHT PANEL

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

DATA VISUALIZATION

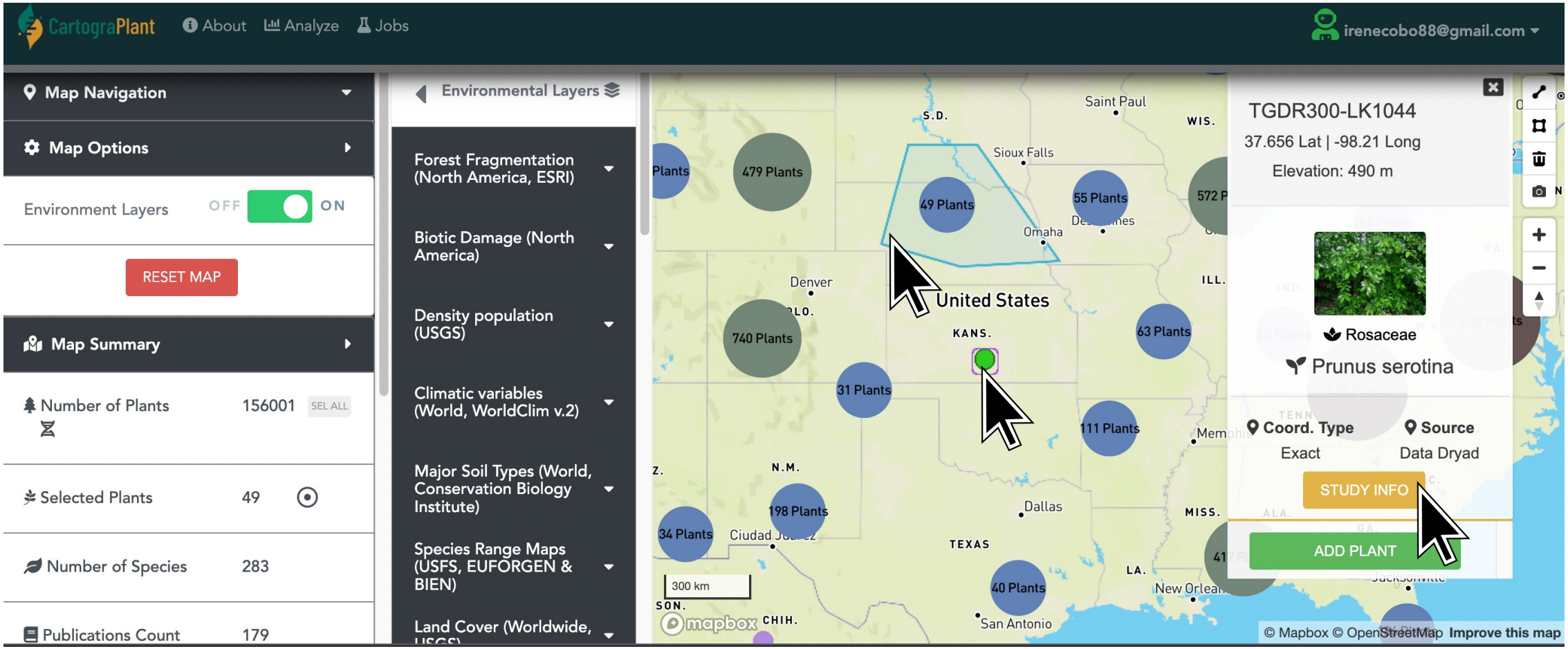
https://cartograplant.org/



BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA VISUALIZATION

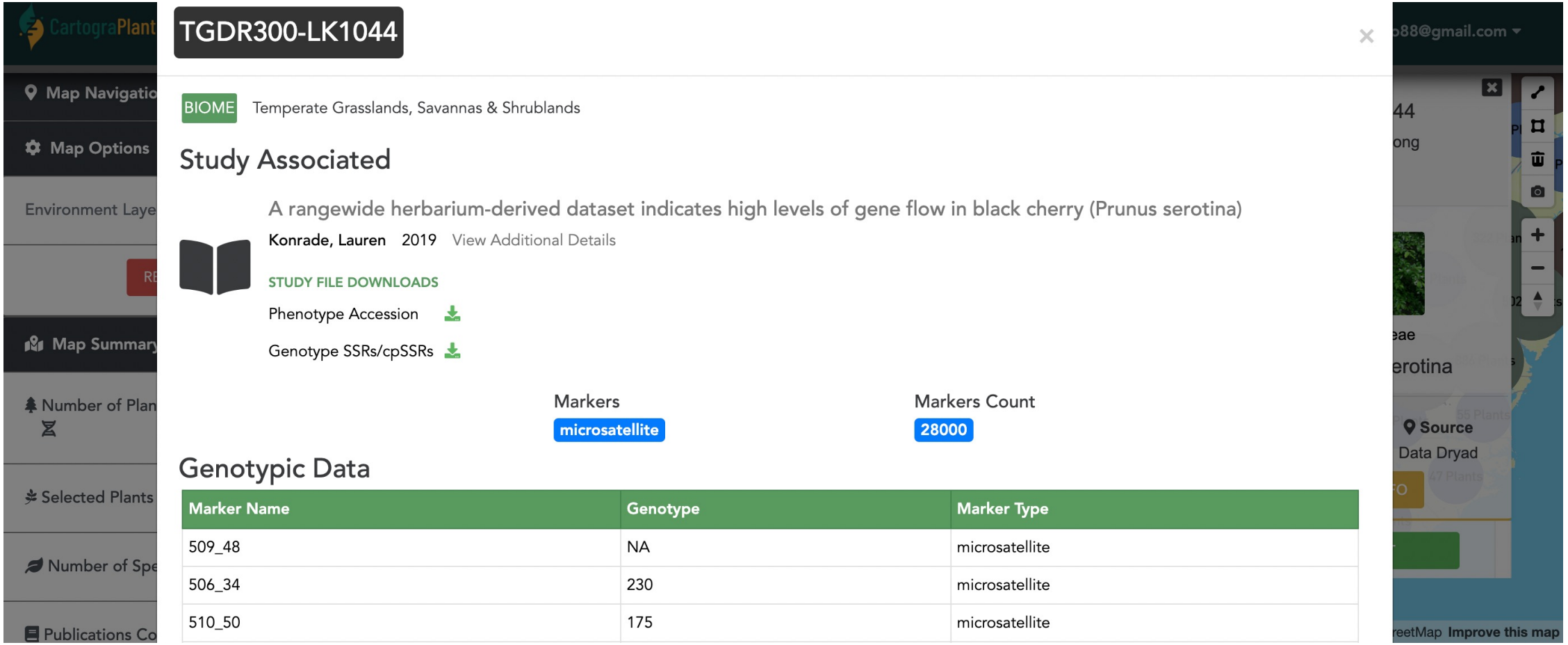
https://cartograplant.org/



BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA VISUALIZATION

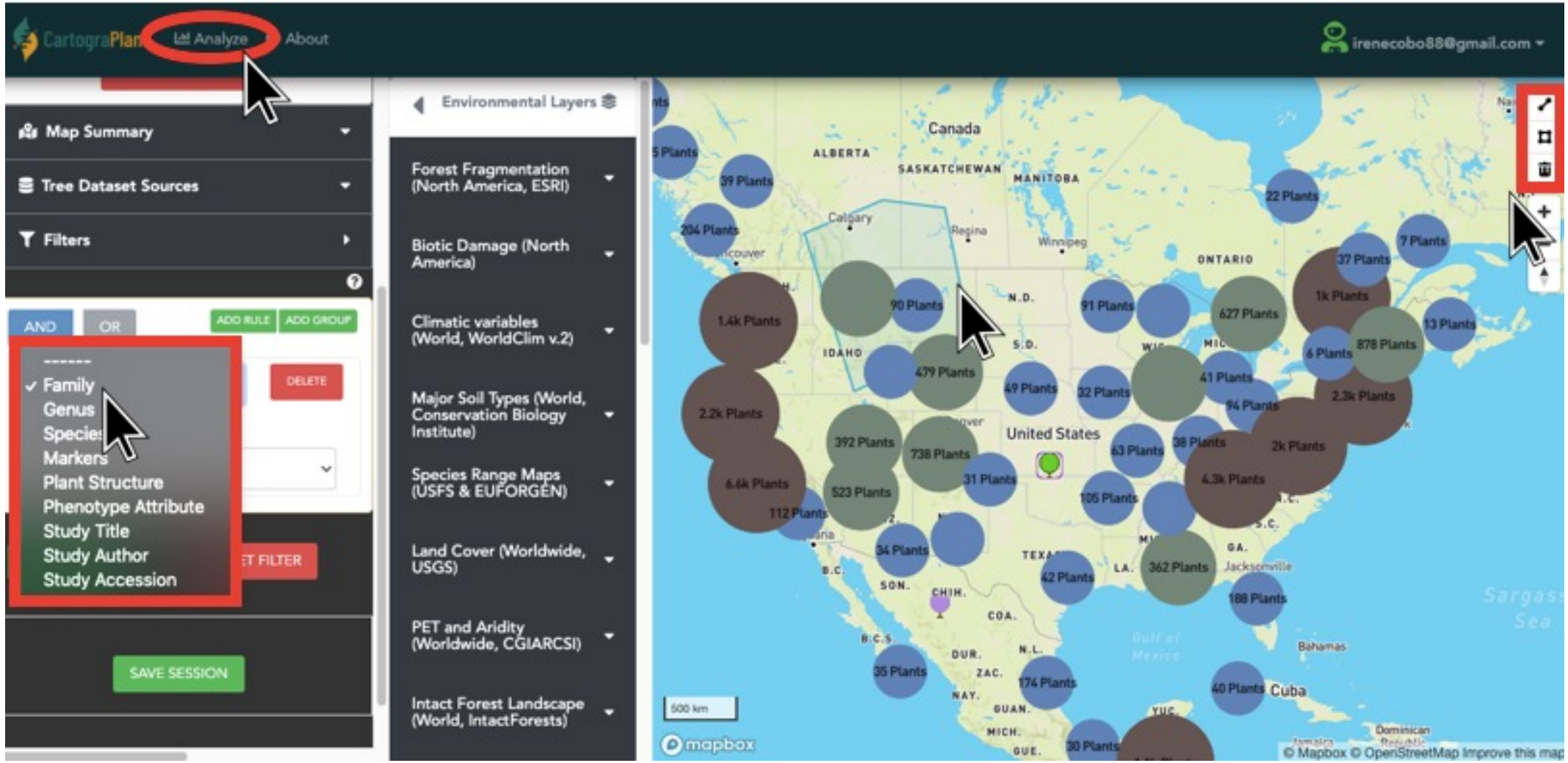
https://cartograplant.org/



BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA VISUALIZATION

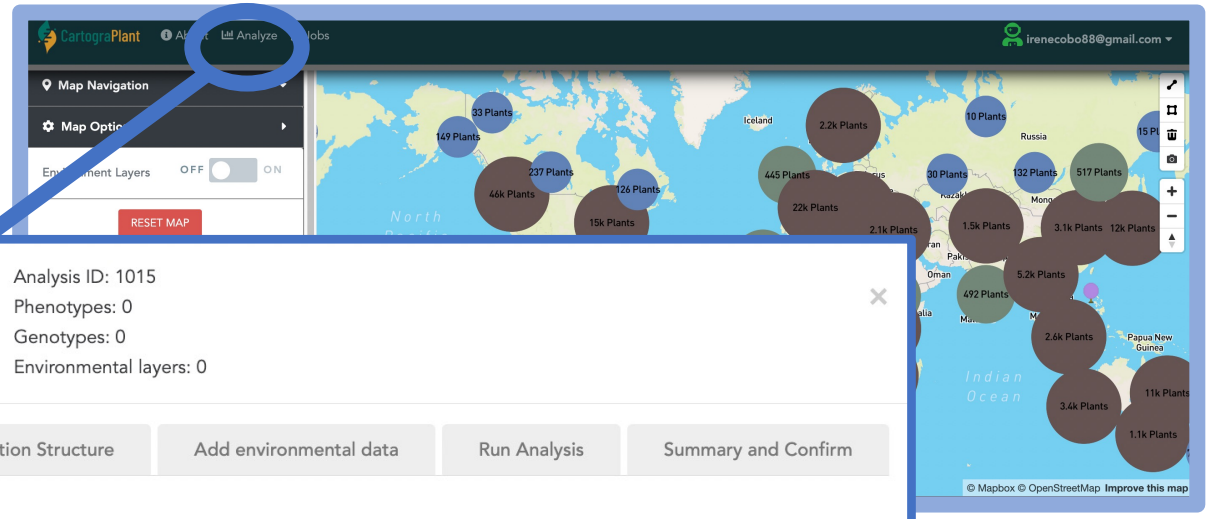
https://cartograplant.org/



BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA ANALYSIS

https://cartograplant.org/



CartograPlant

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: Untitled Analysis type: Genotype x Phenotype x Environmental 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 Test CREATE NEW WORKSPACE

MANAGE WORKSPACE FILES

This workspace currently contains no files

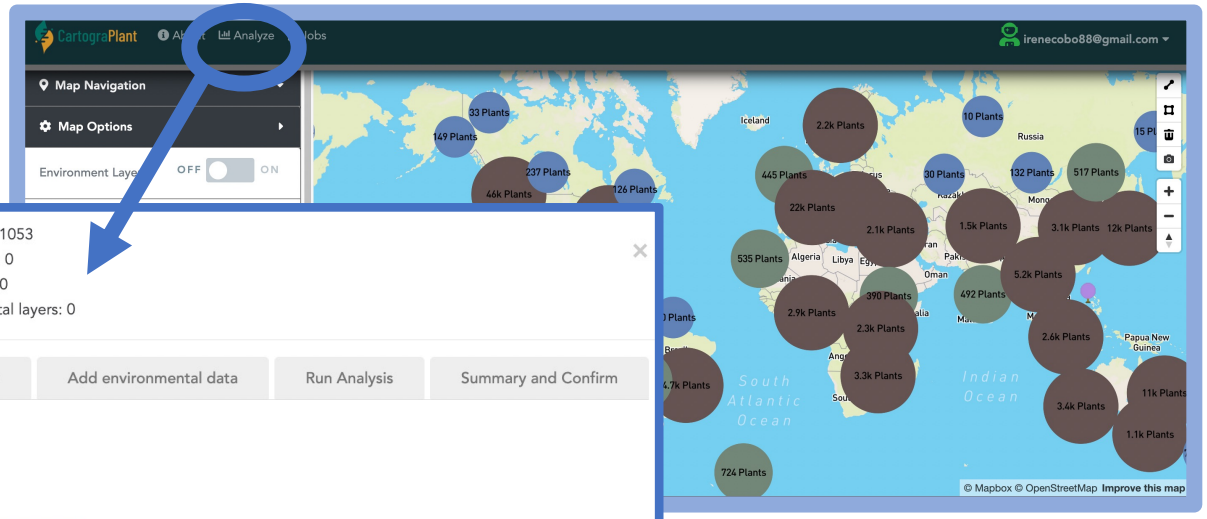
NEXT

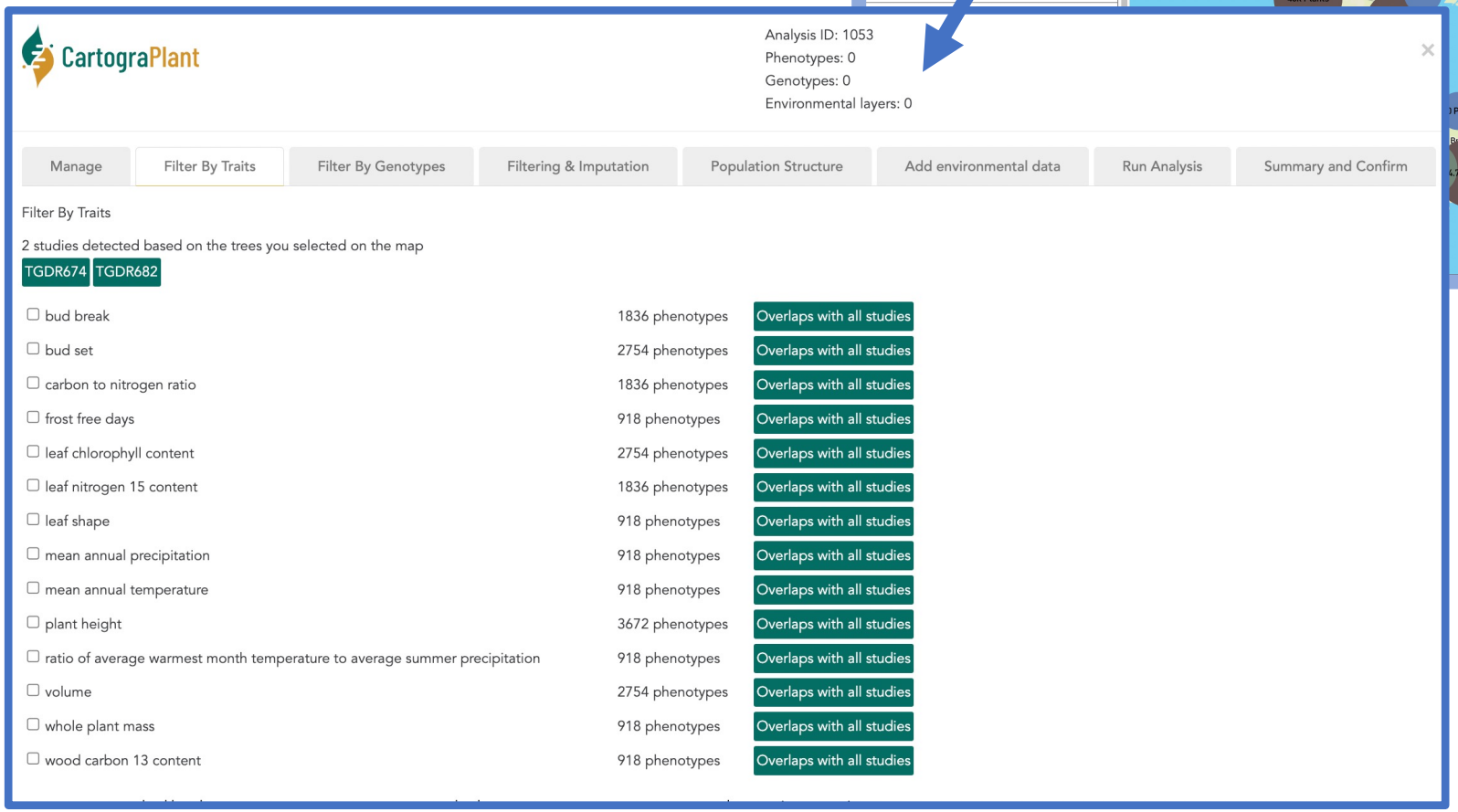
CLOSE

BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA ANALYSIS

https://cartograplant.org/

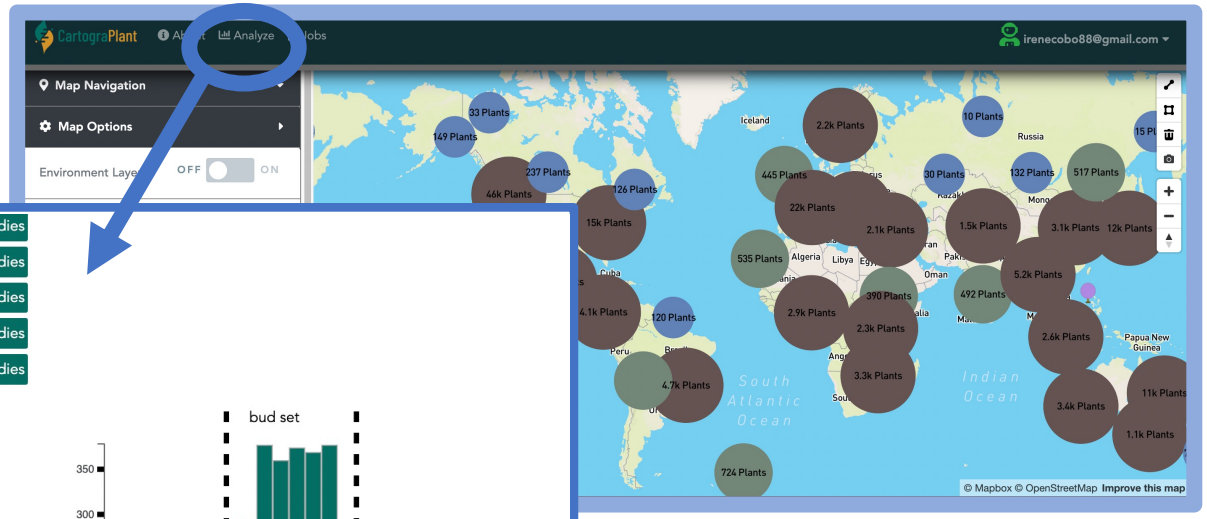


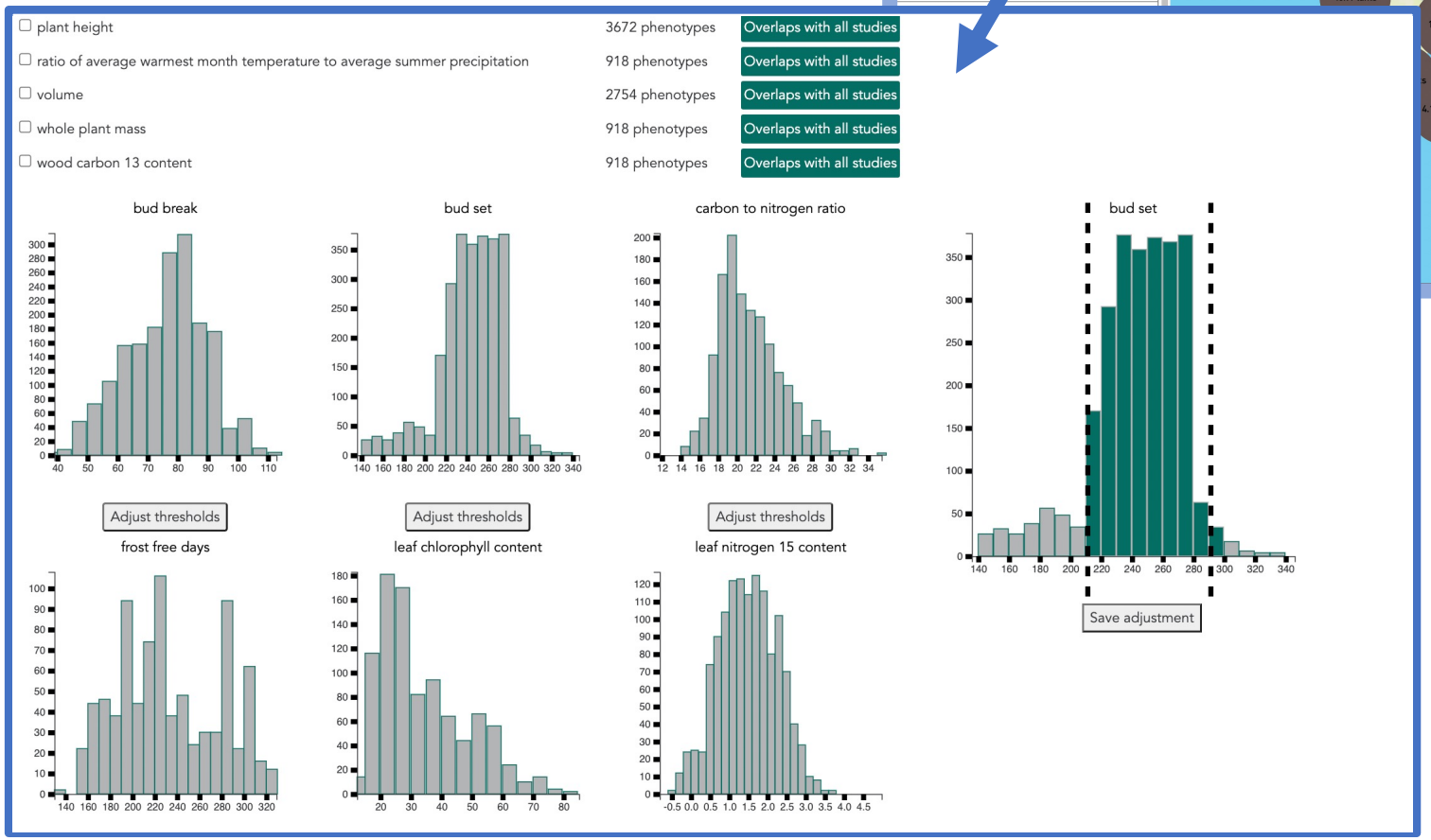


BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA ANALYSIS

https://cartograplant.org/

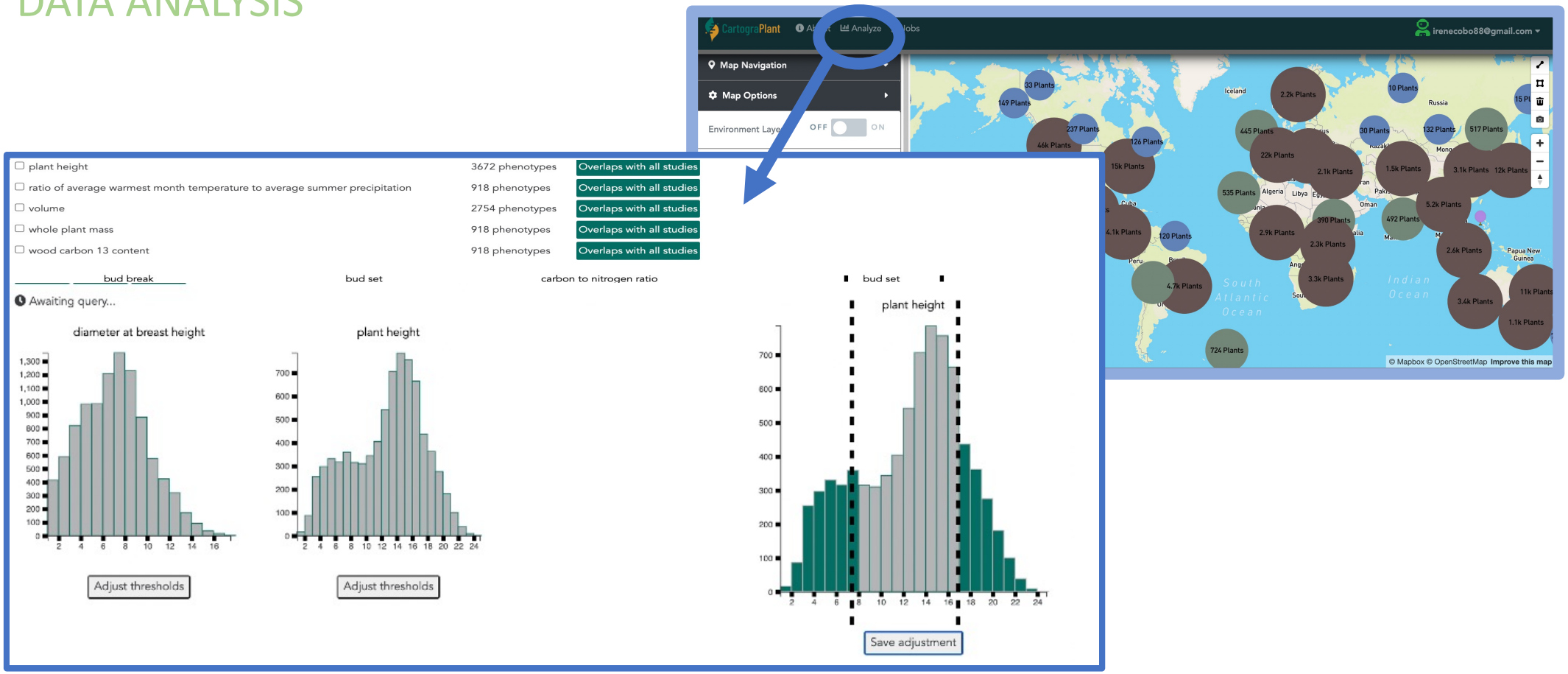




BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA ANALYSIS

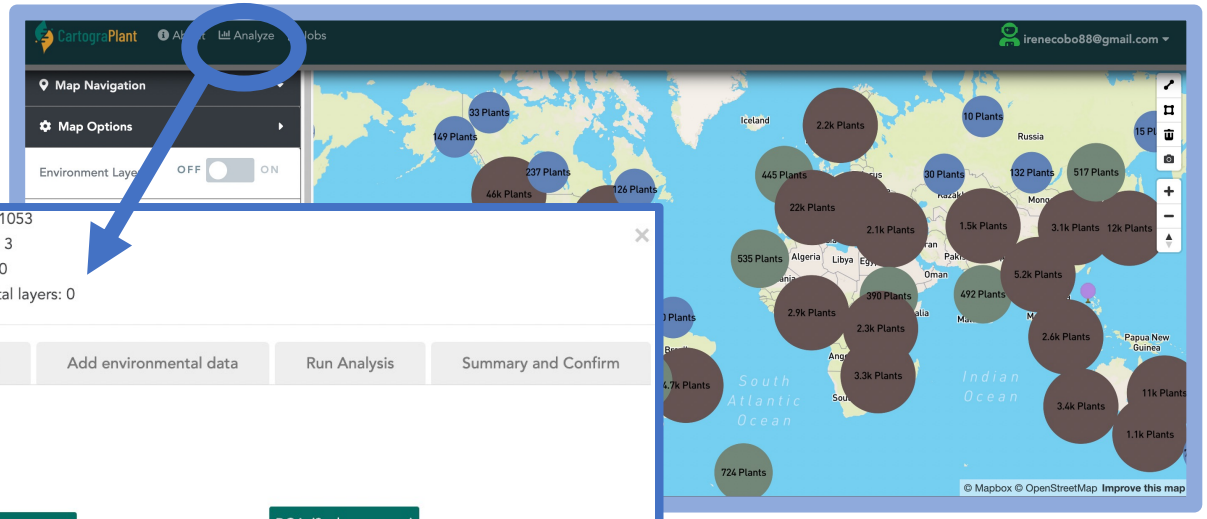
https://cartograplant.org/



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

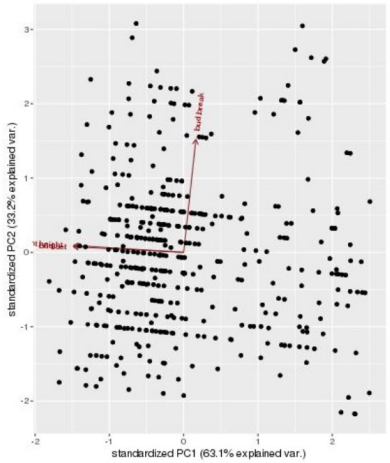
Filter By Traits

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

TGDR674 TGDR682

☑ bud break	1836 phenotypes	Overlaps with all studies
☑ bud set	2411 phenotypes	Overlaps with all studies
☐ carbon to nitrogen ratio	1836 phenotypes	Overlaps with all studies
☐ frost free days	918 phenotypes	Overlaps with all studies
☐ leaf chlorophyll content	2754 phenotypes	Overlaps with all studies
☐ leaf nitrogen 15 content	1836 phenotypes	Overlaps with all studies
☐ leaf shape	918 phenotypes	Overlaps with all studies
☐ mean annual precipitation	918 phenotypes	Overlaps with all studies
☐ mean annual temperature	918 phenotypes	Overlaps with all studies
☑ plant height	3672 phenotypes	Overlaps with all studies
☐ ratio of average warmest month temperature to average summer precipitation	918 phenotypes	Overlaps with all studies
☐ volume	2754 phenotypes	Overlaps with all studies
☐ whole plant mass	918 phenotypes	Overlaps with all studies
☐ wood carbon 13 content	918 phenotypes	Overlaps with all studies

PCA (3 phenotypes)

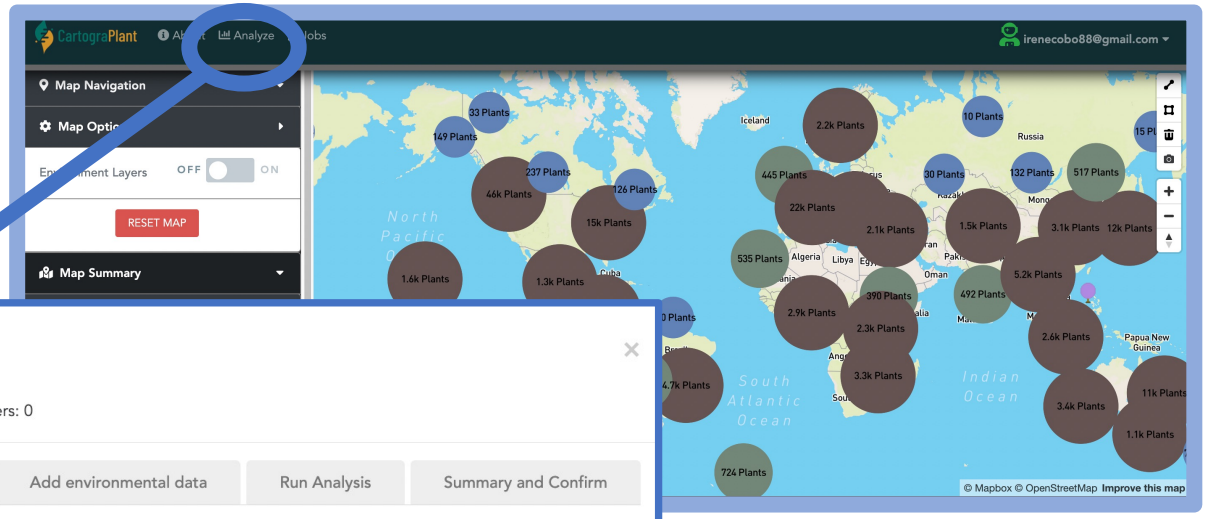


artogratree#analysis-filter-snp bud set carbon to nitrogen ratio bud set

BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA ANALYSIS

https://cartograplant.org/



CartograPlant

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

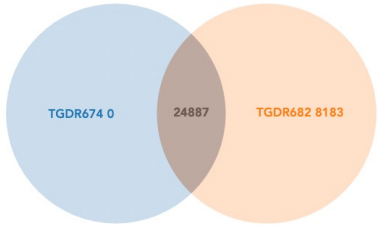
Manage | Filter By Traits | Filter By Genotypes | Filtering & Imputation | Population Structure | Add environmental data | Run Analysis | Summary and Confirm

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

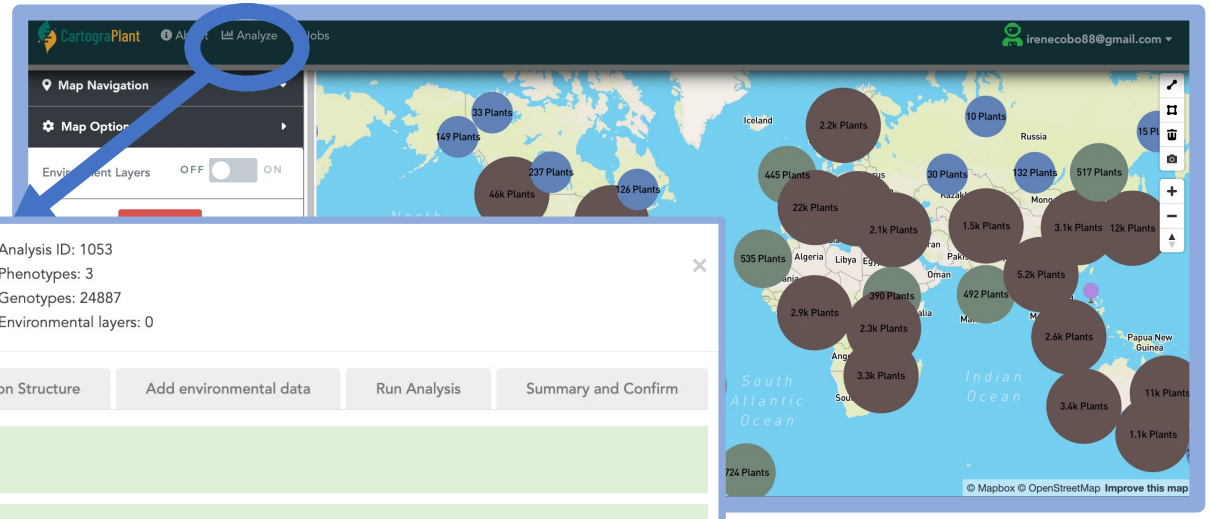
NEXT

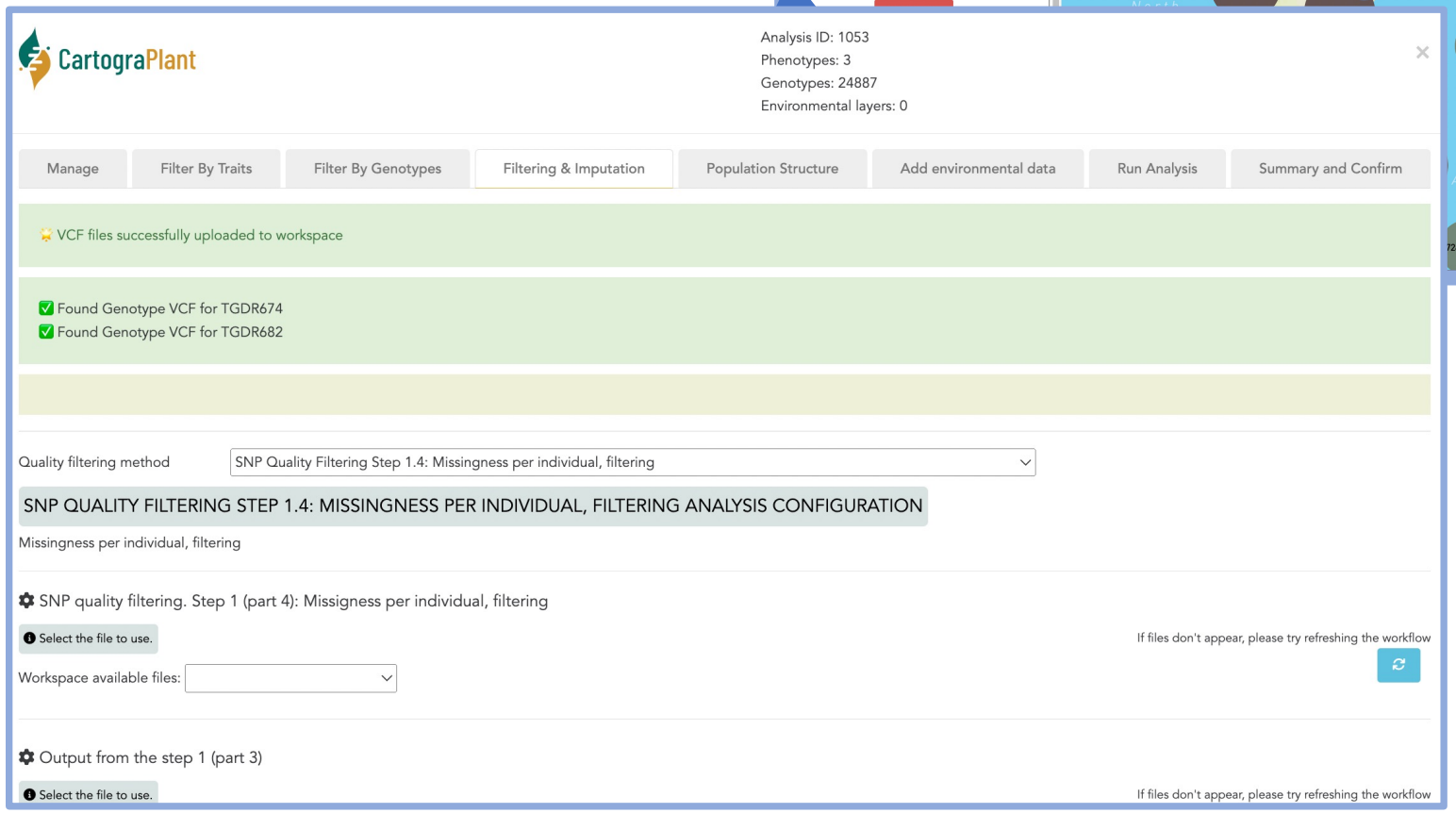
CLOSE

BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA ANALYSIS

https://cartograplant.org/

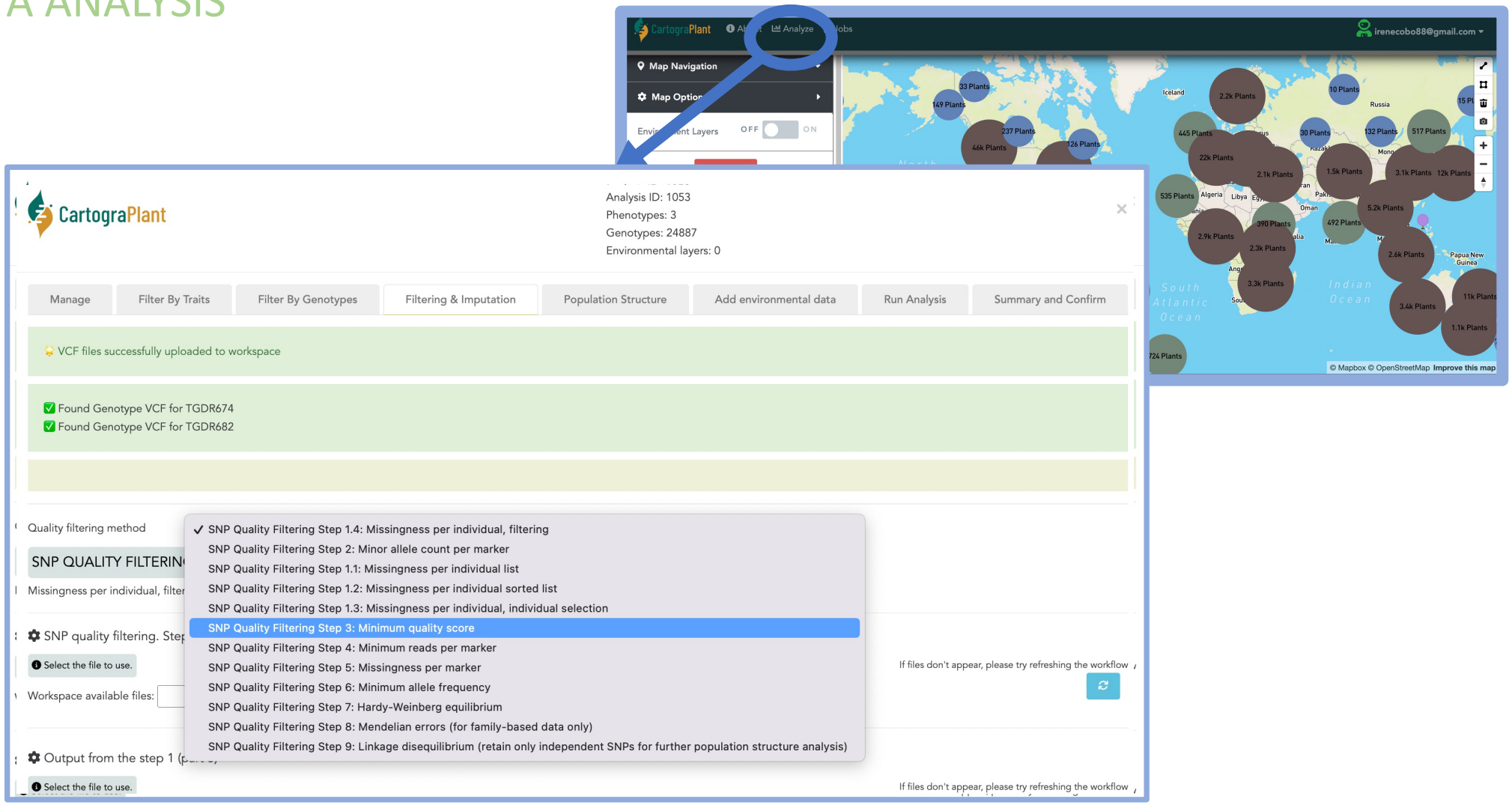




BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA ANALYSIS

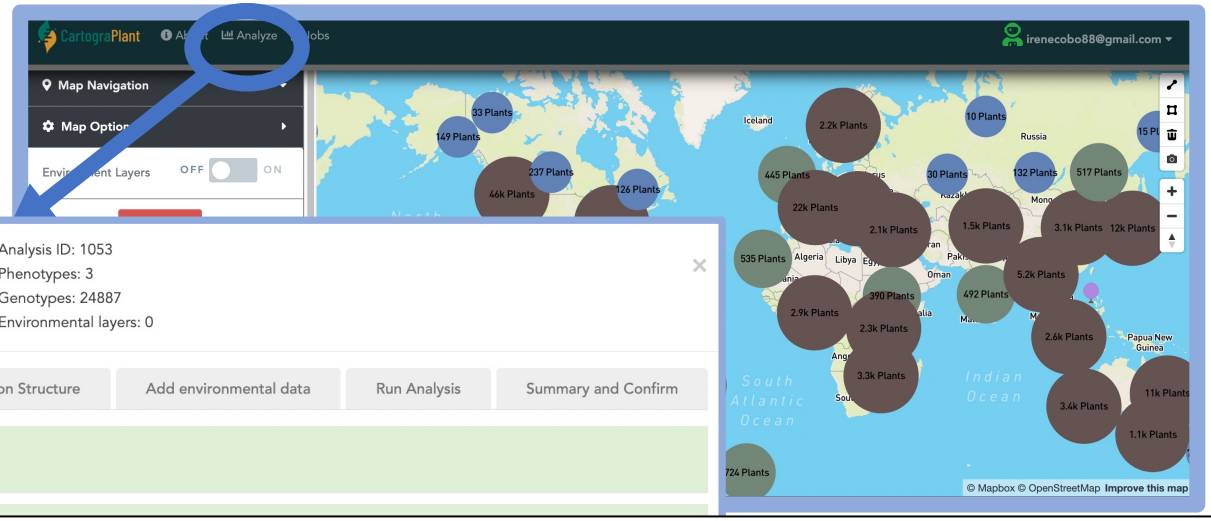
https://cartograplant.org/

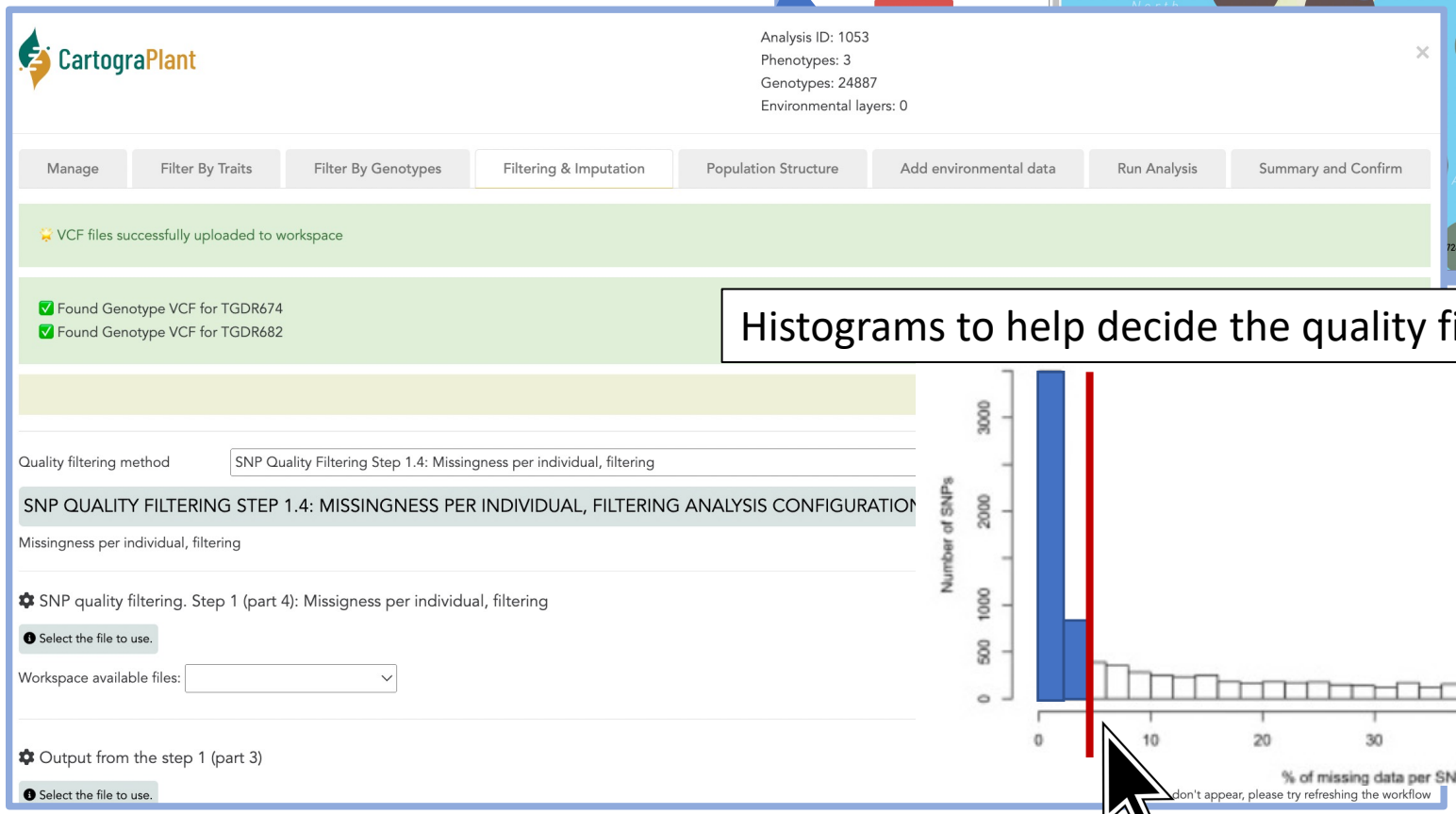


BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA ANALYSIS

https://cartograplant.org/



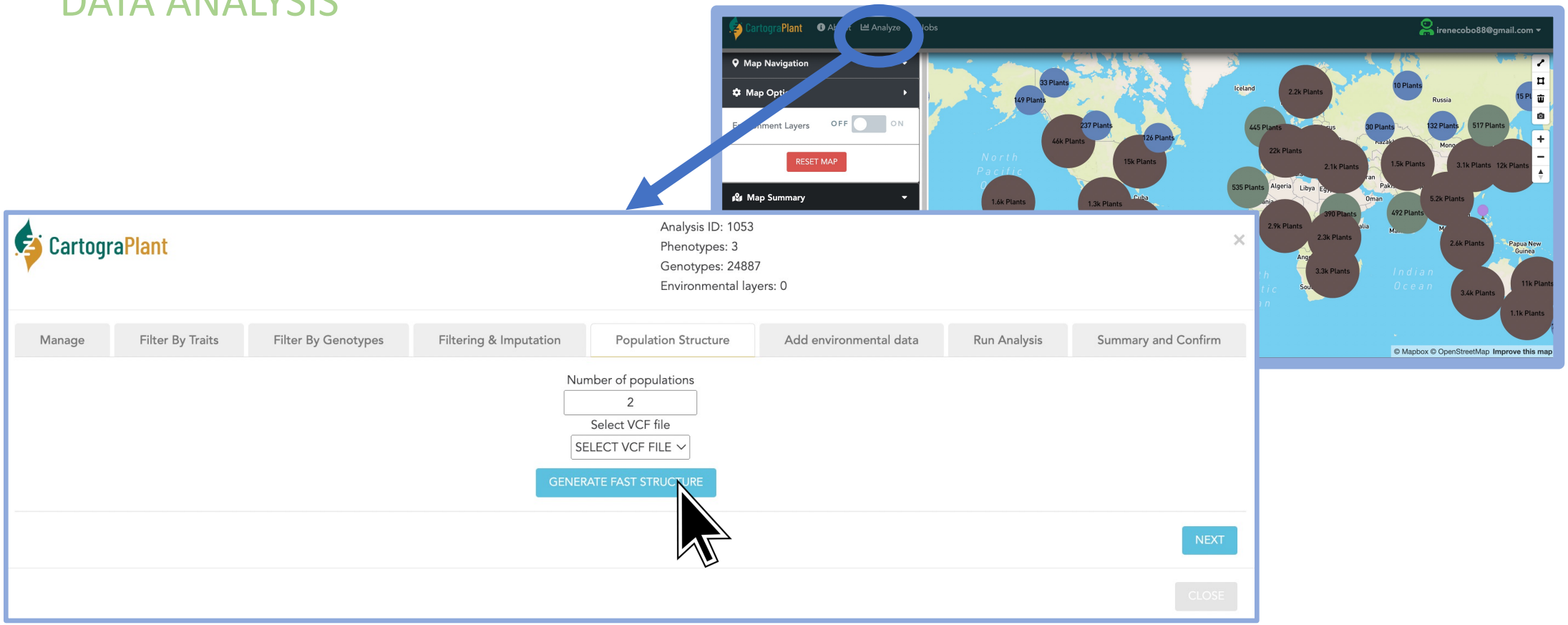


Histograms to help decide the quality filtering thresholds available soon!

BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA ANALYSIS

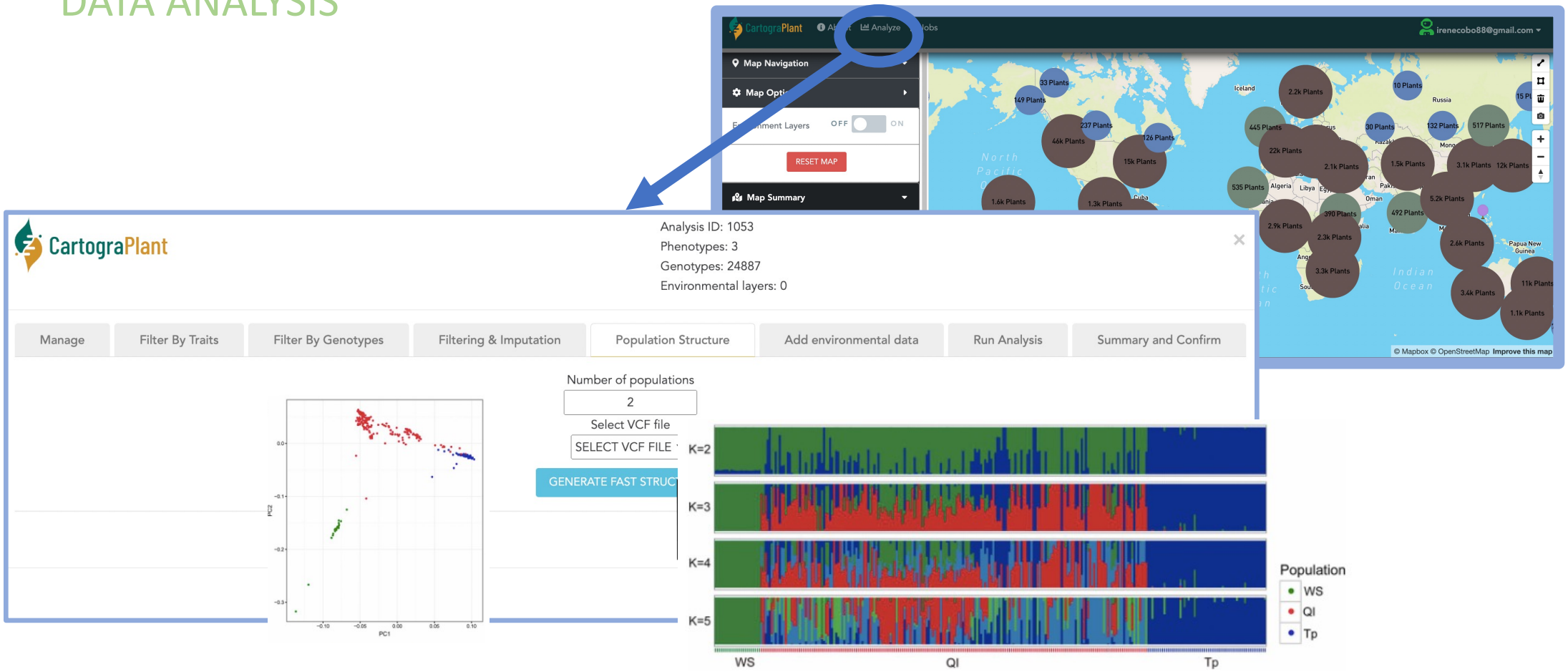
https://cartograplant.org/



BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA ANALYSIS

https://cartograplant.org/

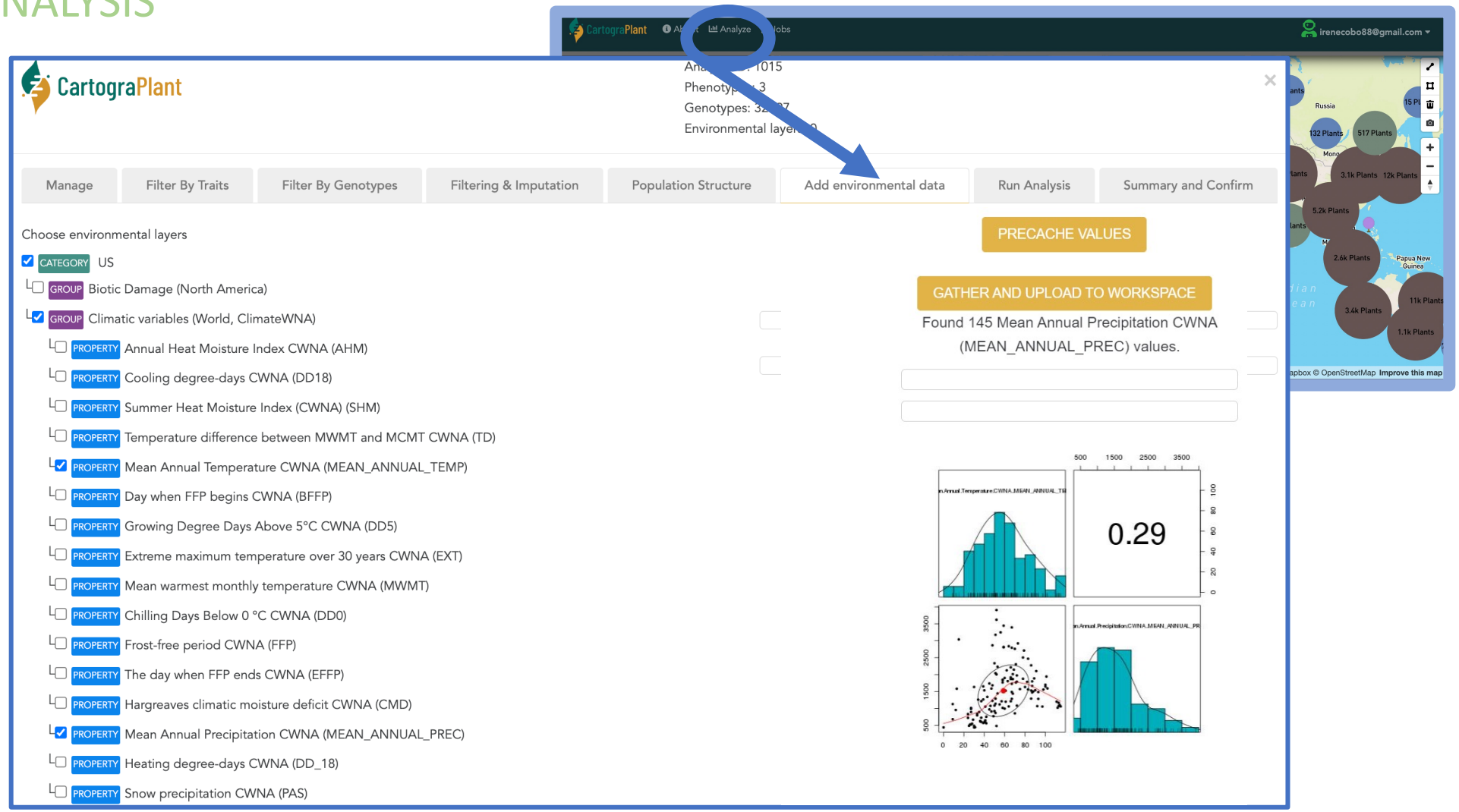


Population structure calculation (PCA and DAPC) and visualization (fastSTRUCTURE, PCA and DAPC) available soon!!!

BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA ANALYSIS

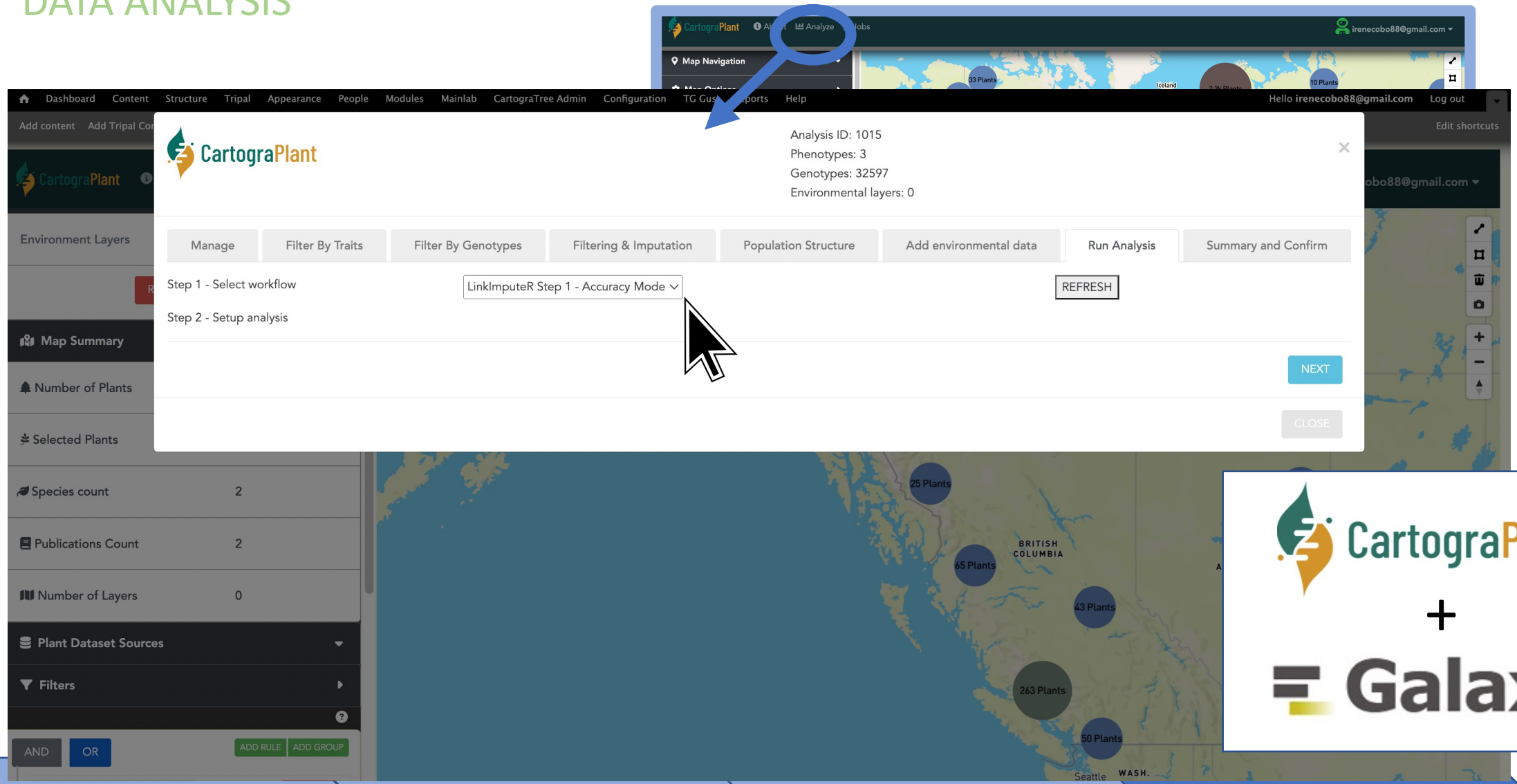
https://cartograplant.org/



BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA ANALYSIS

https://cartograplant.org/

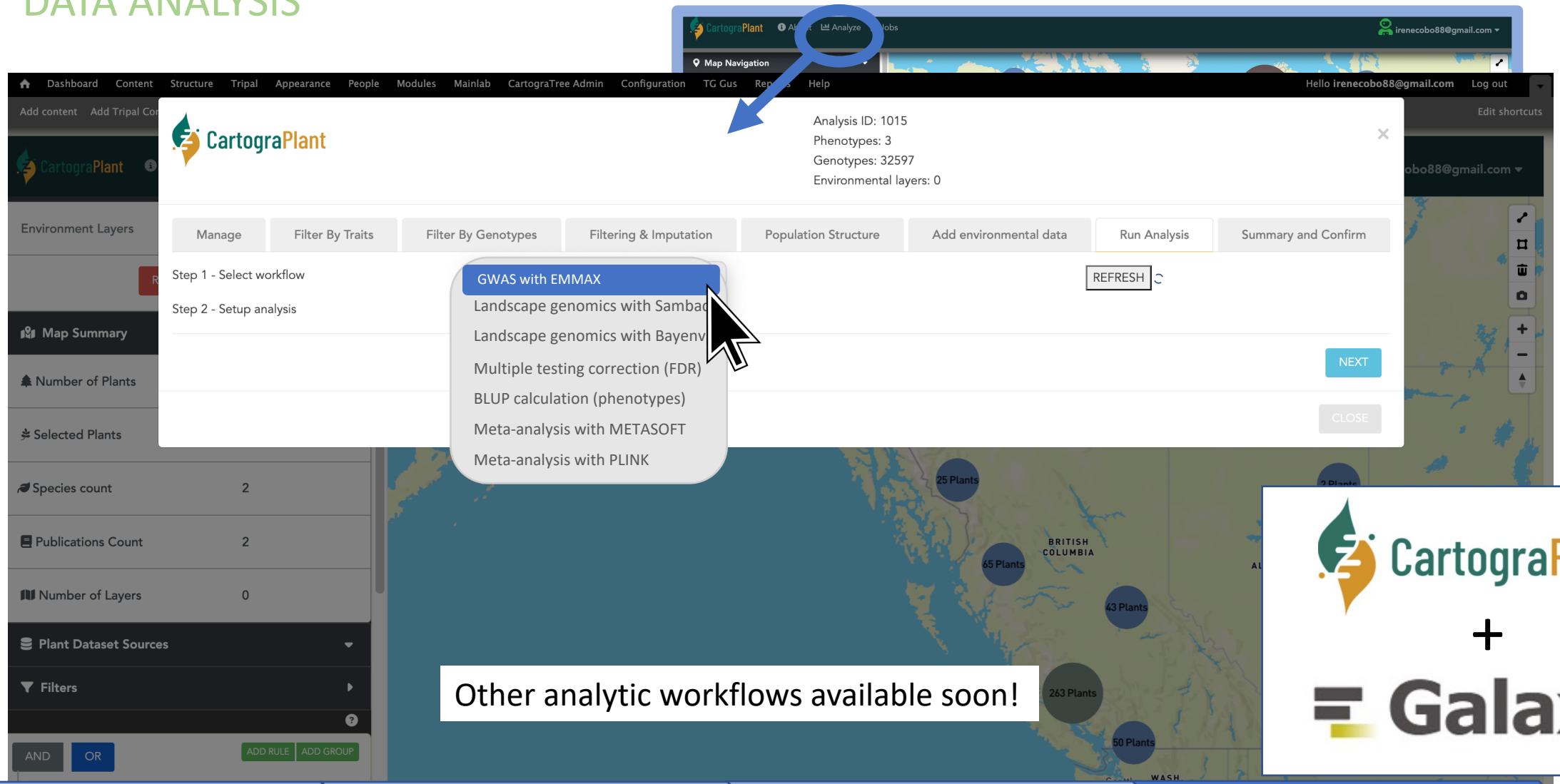


CartograPlant + Galaxy

BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA ANALYSIS

https://cartograplant.org/



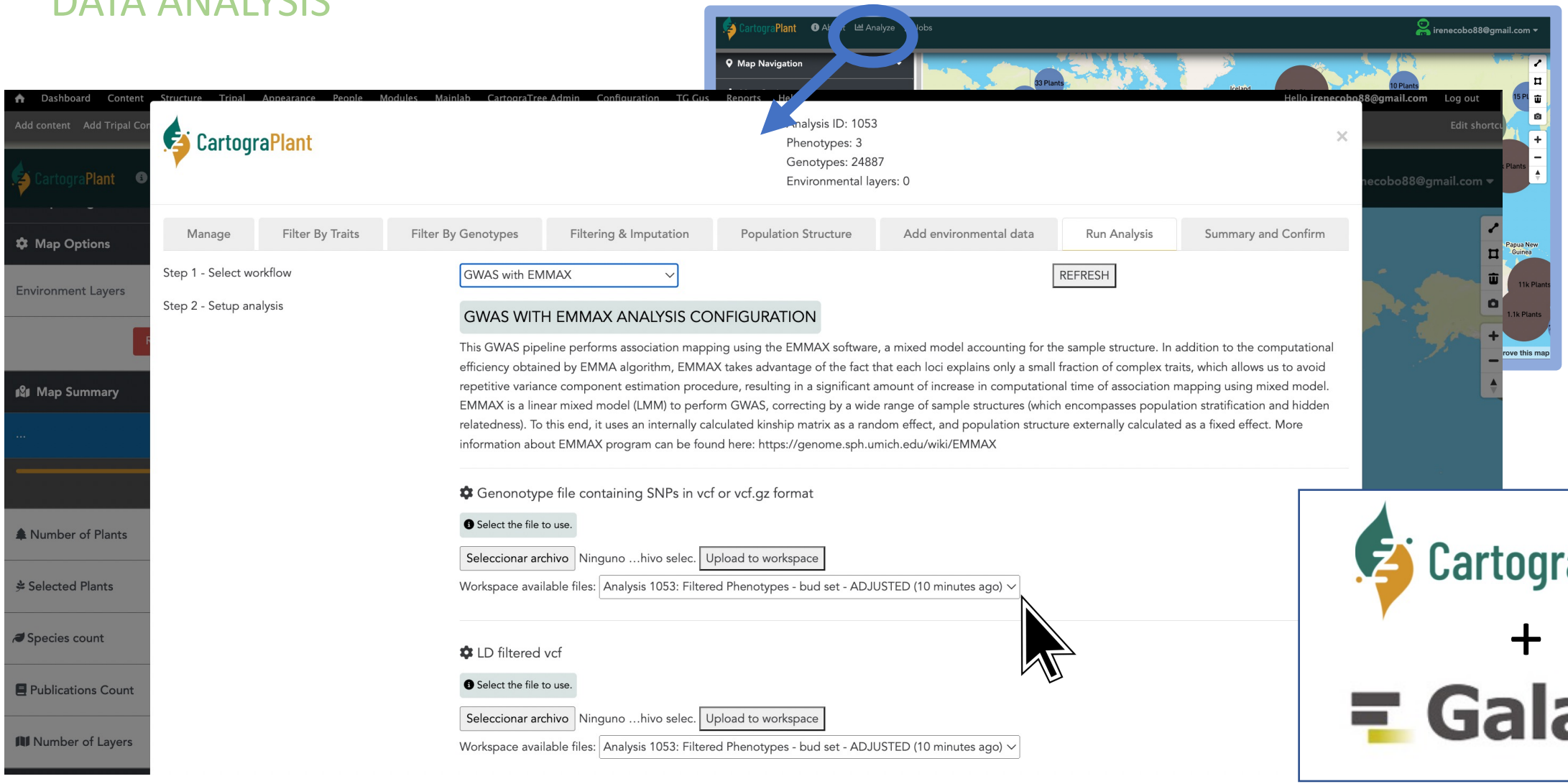
Other analytic workflows available soon!

CartograPlant + Galaxy

BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

DATA ANALYSIS

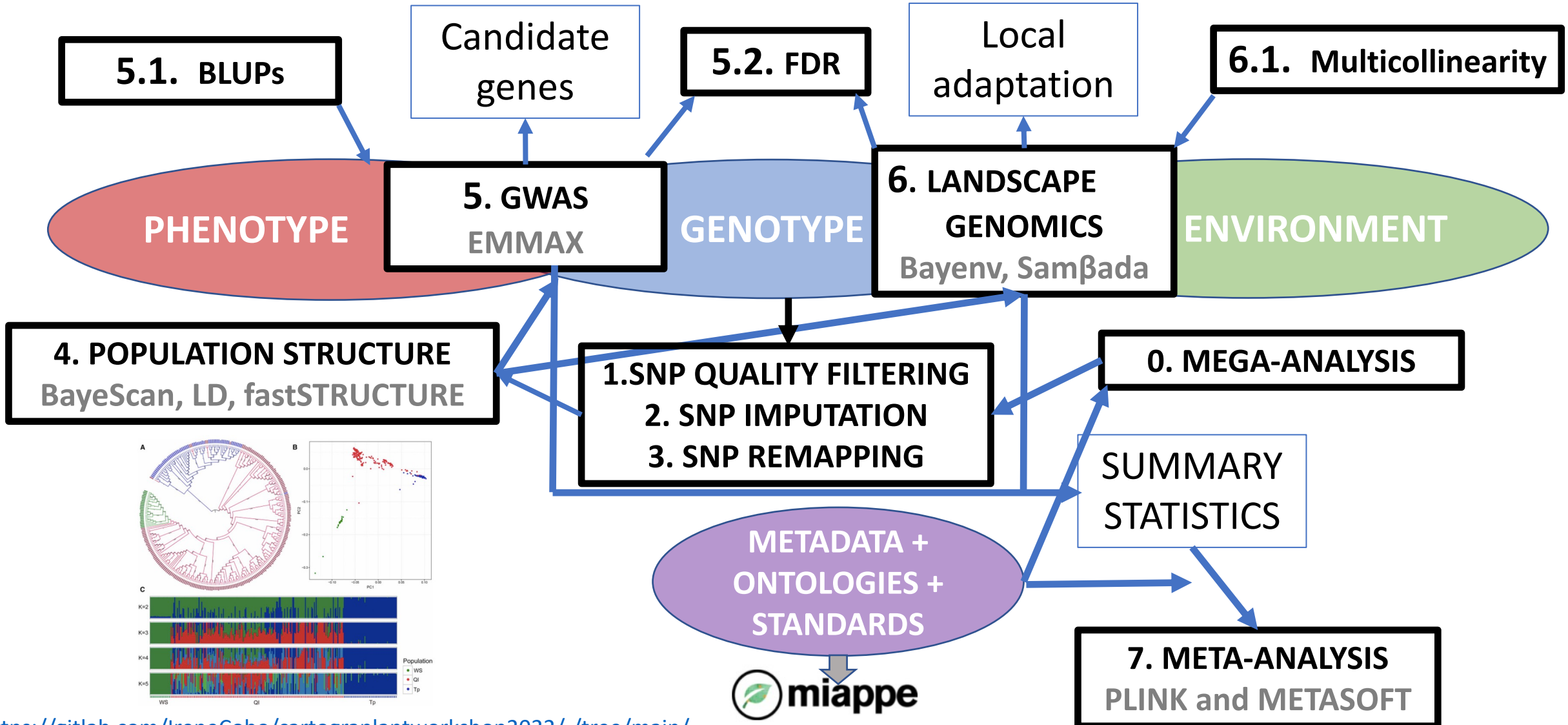
https://cartograplant.org/



CartograPlant + Galaxy

BACKGROUND | DATA INTEGRATION | DATA VISUALIZATION | DATA ANALYSIS

CARTOGRAPLANT WORKFLOWS OVERVIEW



https://gitlab.com/IreneCobo/cartograplantworkshop2022/-/tree/main/

CARTOGRAPLANT WORKFLOWS OVERVIEW

5.1. BLUPs

Candidate genes

5.2. FDR

Local adaptation

6.1. Multicollinearity

PHENOTYPE

5. GWAS
EMMAX

GENOTYPE

6. LANDSCAPE GENOMICS
Bayenv, Samβada

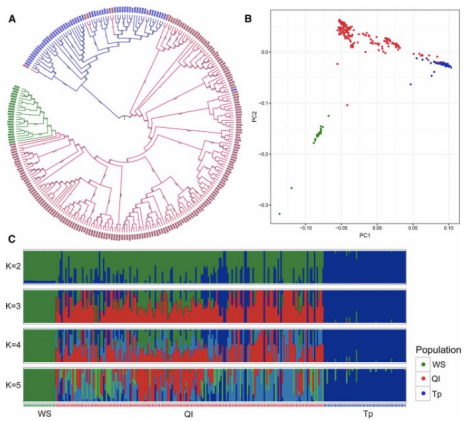
ENVIRONMENT

4. POPULATION STRUCTURE
BayeScan, LD, fastSTRUCTURE

1.SNP QUALITY FILTERING
2. SNP IMPUTATION
3. SNP REMAPPING

0. MEGA-ANALYSIS

SUMMARY STATISTICS



METADATA + ONTOLOGIES + STANDARDS

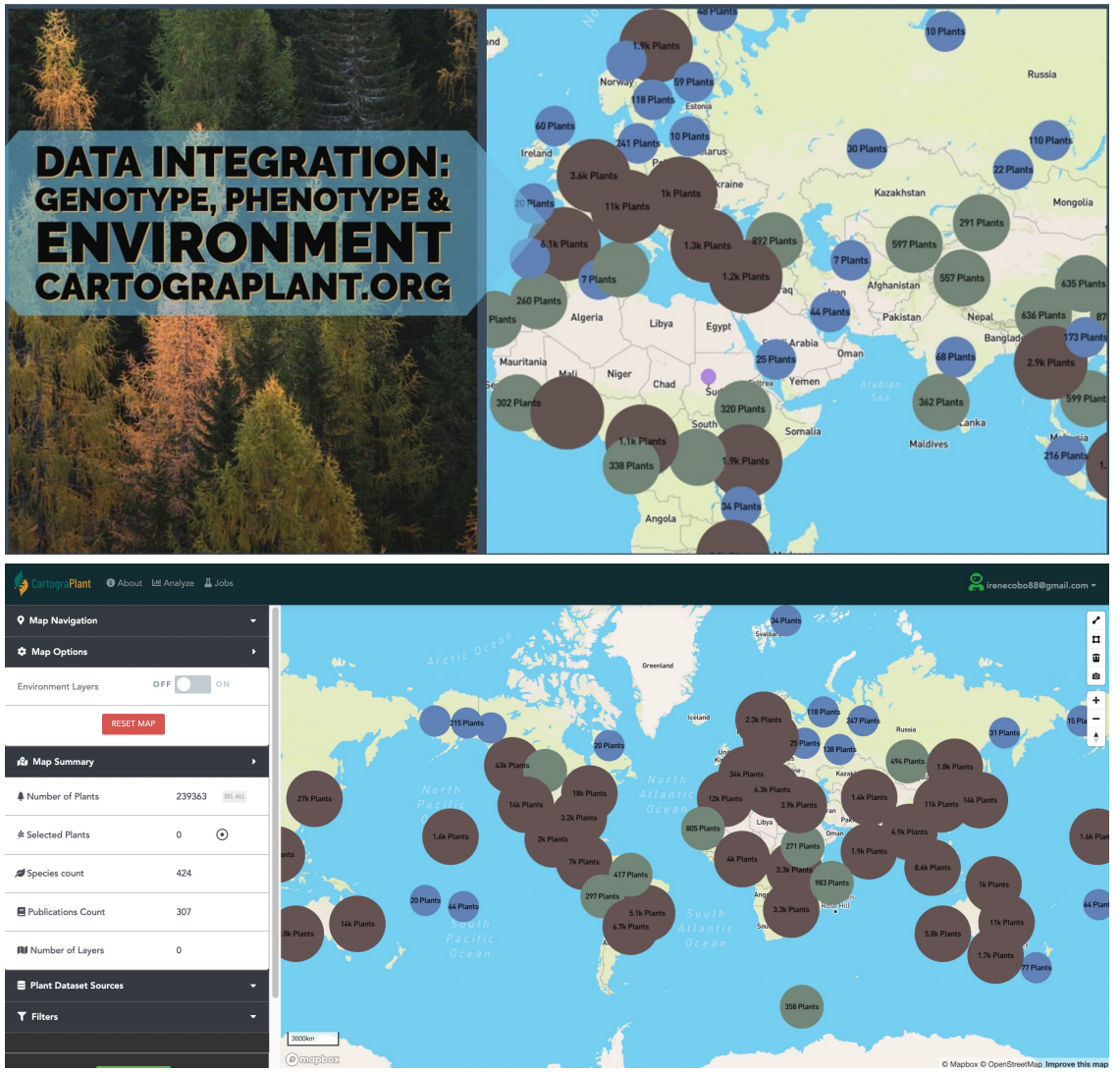
miappe

7. META-ANALYSIS
PLINK and METASOFT

https://gitlab.com/IreneCobo/cartograplantworkshop2022/-/tree/main/

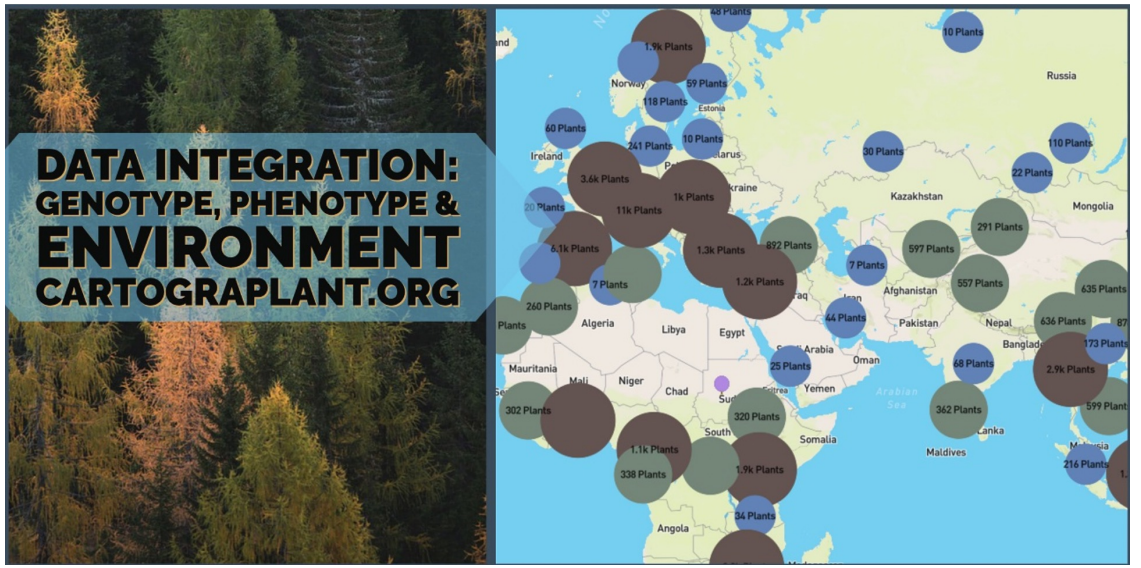
CONCLUSIONS

- **CartograPlant** is intended to serve as a **community resource for Plant Molecular Ecology.**



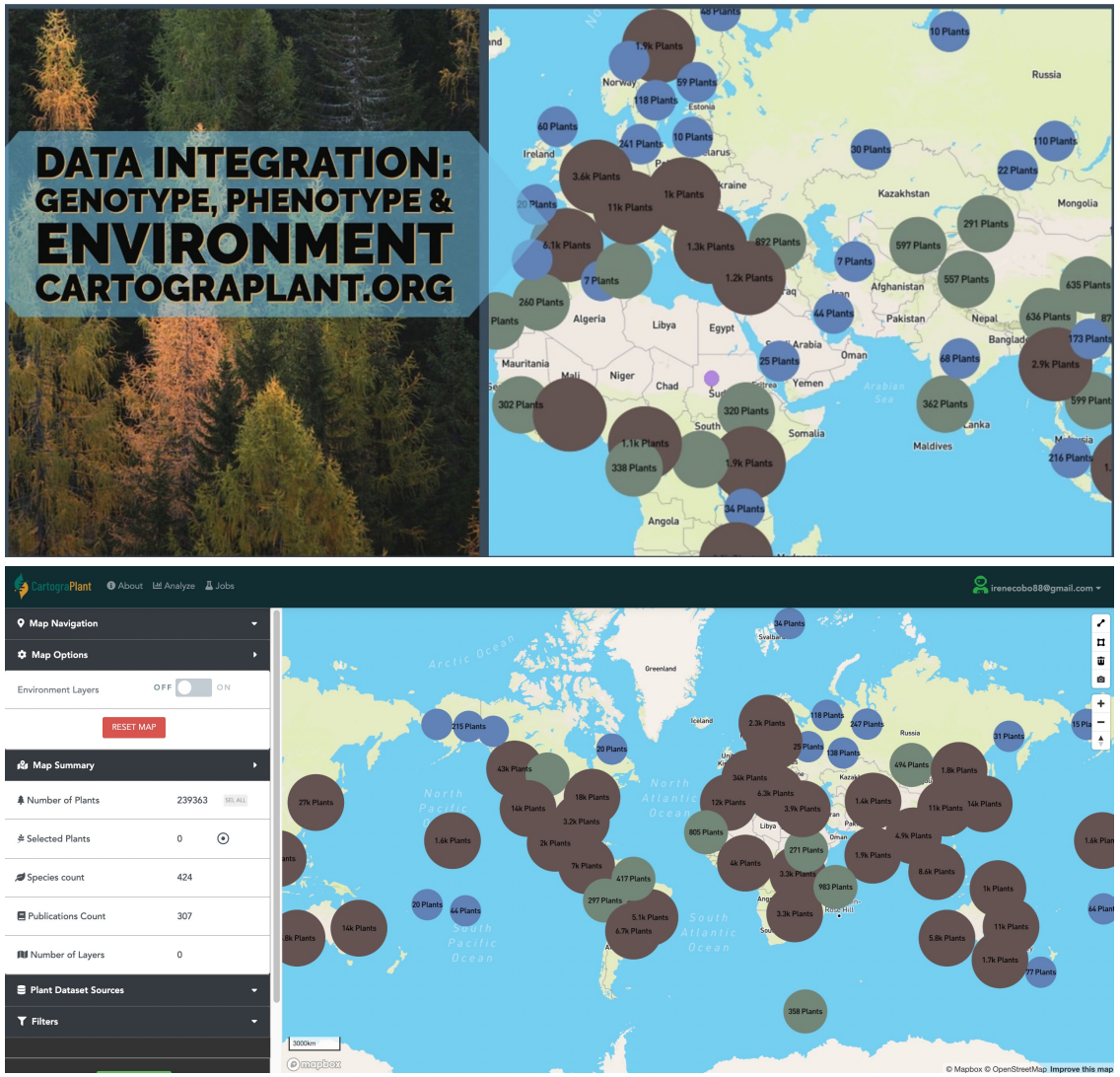
CONCLUSIONS

- **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.**



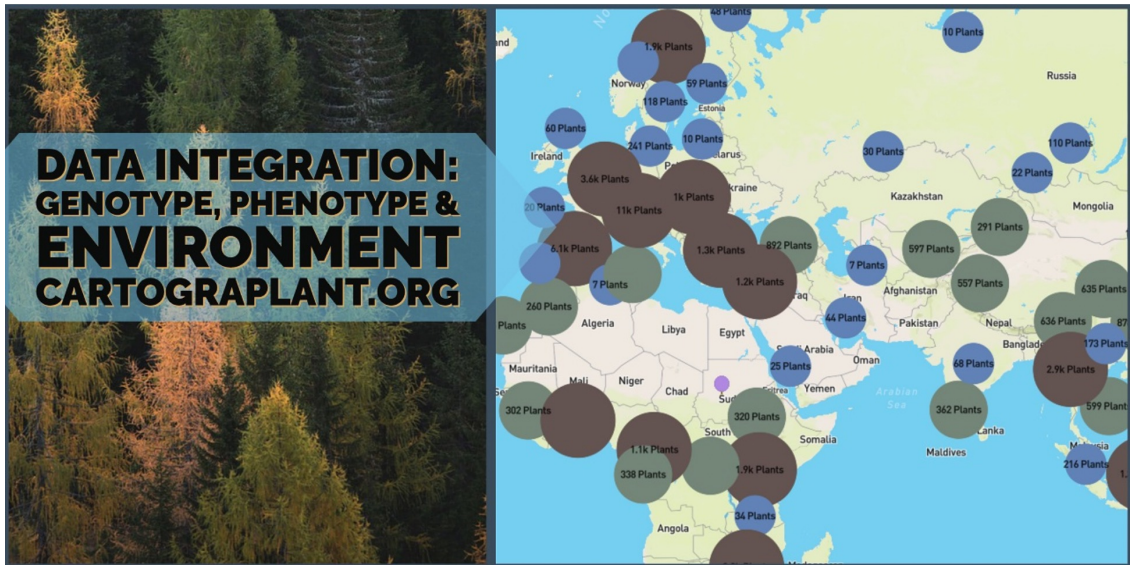
CONCLUSIONS

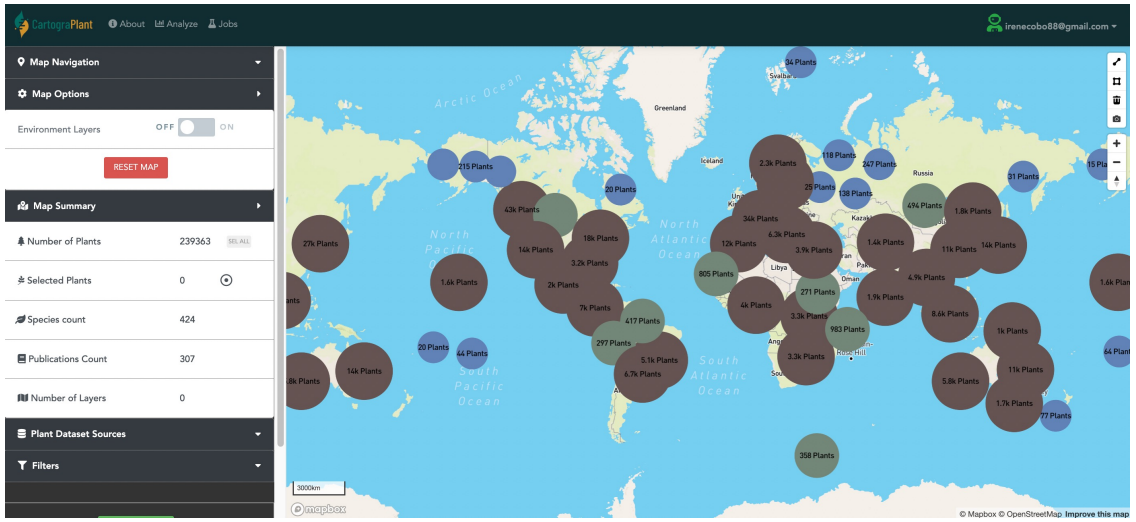
- **CartograPlant** is intended to serve as a **community resource for Plant Molecular Ecology.**
- **These flexible analytic workflows** allow to analyze a diversity of data types (e.g. SNPs, SSR) and experimental designs (e.g. natural populations, common garden) and facilitate a **diversity of biological 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 envrironmental) from different studies, thanks to the metadata collection using ontologies and standards.



CONCLUSIONS

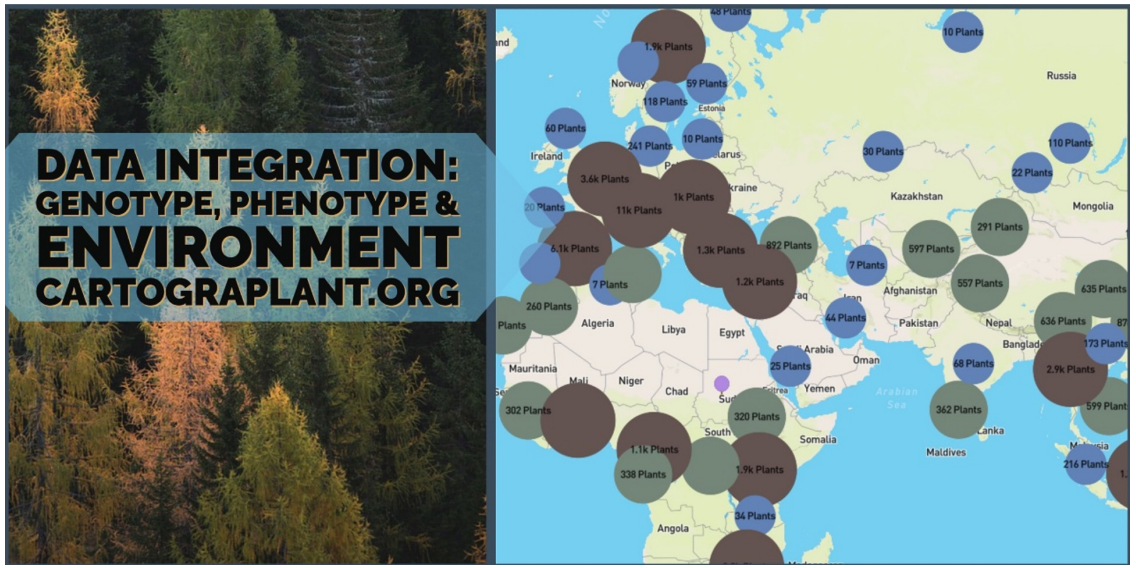
- **CartograPlant** is intended to serve as a **community resource for Plant Molecular Ecology.**
- **These flexible analytic workflows** allow to analyze a diversity of data types (e.g. SNPs, SSR) and experimental designs (e.g. natural populations, common garden) and facilitate a **diversity of biological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

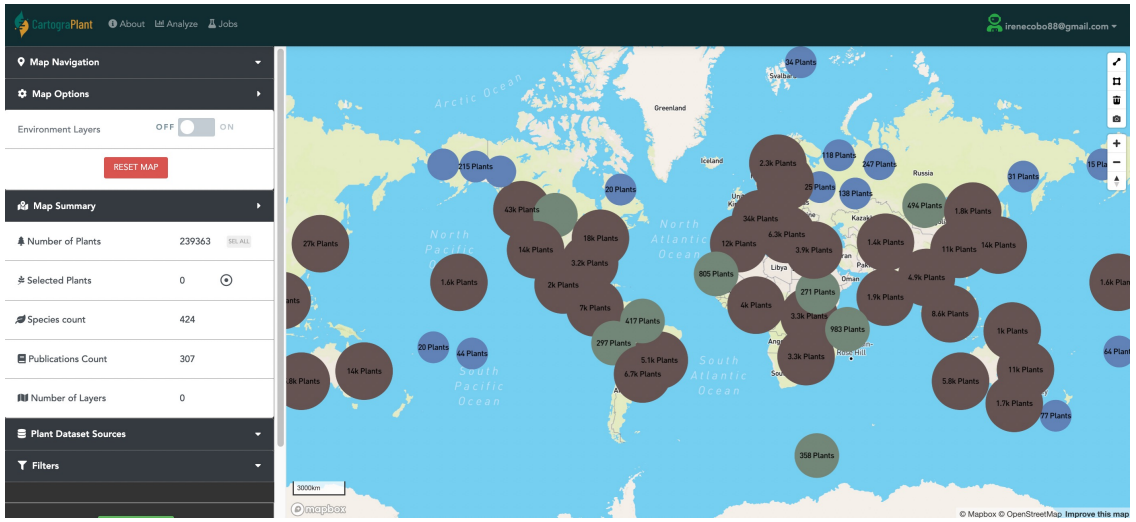




CONCLUSIONS

- **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 envrironmental) 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/