Today’s Schedule (in EST!)

11:00-11:20 Introduction to CartograPlant - Dr. Jill Wegrzyn

11:20-11:40 Introduction to Data Submission with TPPS/TPPSc - Emily Grau

11:40-12:00 Introduction to Data Collection/Mobile Phenotyping with TreeSnap - Dr. Margaret Staton

12:00-12:15 Break

12:15-12:35 Behind the Scenes of CartograPlant - Environmental Layers and Data - Risharde Ramnath

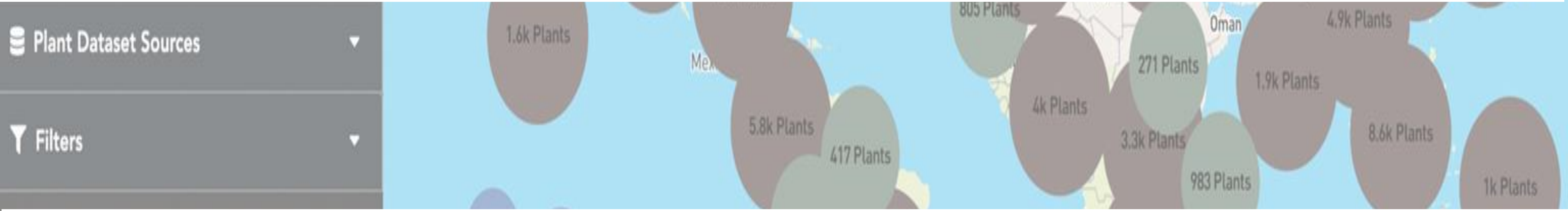
12:35-12:55 Analytics with CartograPlant (GWAS and GEA). Part 1 - Gabriel Barrett

12:55-1:15 Analytics with CartograPlant (GWAS and GEA). Part 2 – Dr. Irene Cobo-Simon

1:15-1:30 Q&A

Data Analysis with CartograPlant: Background

AG2PI WORKSHOP | JULY 12, 2023

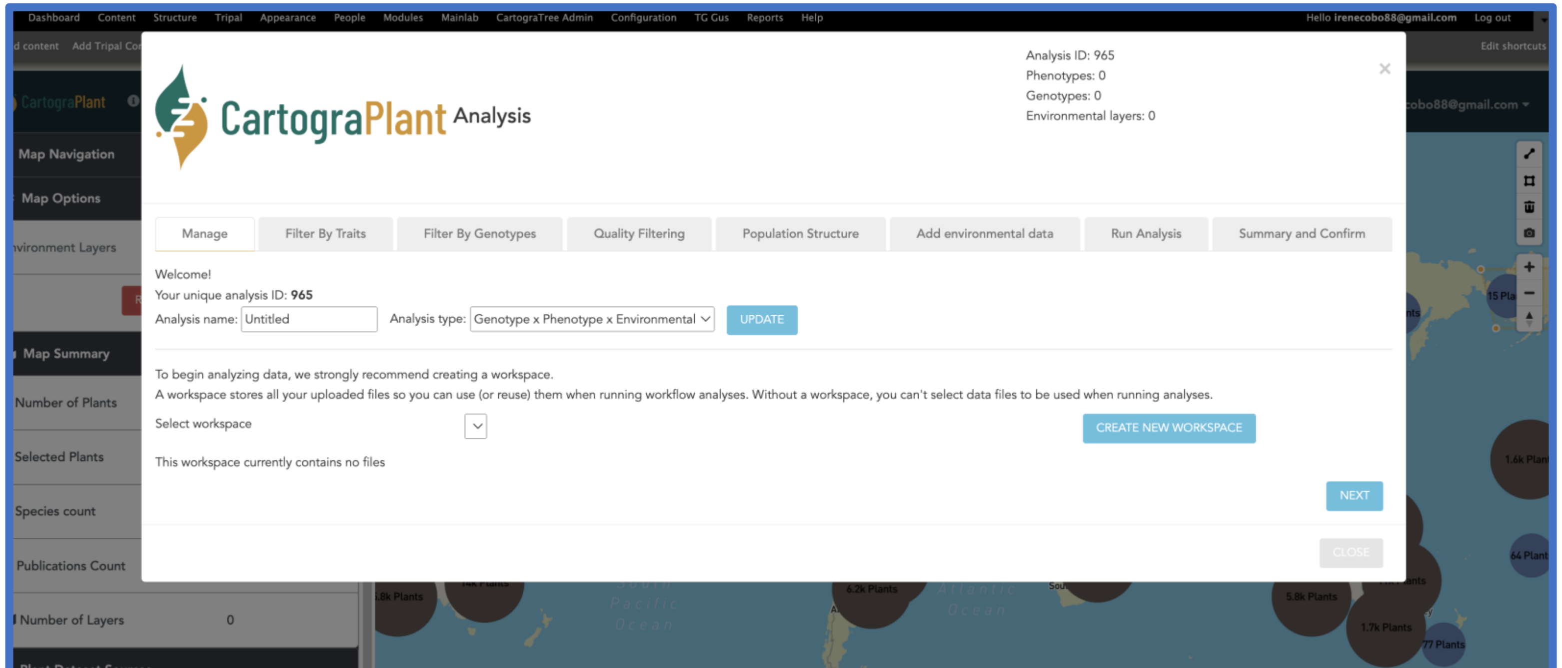


Gabriel Barrett, MSc.
MSc. degree in biology student
University of Rhode Island

UCONN
UNIVERSITY OF CONNECTICUT

AVAILABLE STEPS FOR DATA ANALYSIS IN CARTOGRAPLANT

- **Data quality filtering**
- **Genotype Imputation**
- **Genome Wide Association Analysis (GWAS: GxP, GxE)**
- **Population Structure**



Analysis ID: 965
Phenotypes: 0
Genotypes: 0
Environmental layers: 0

Welcome!
Your unique analysis ID: **965**
Analysis name: Untitled
Analysis type: Genotype x Phenotype x Environmental

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

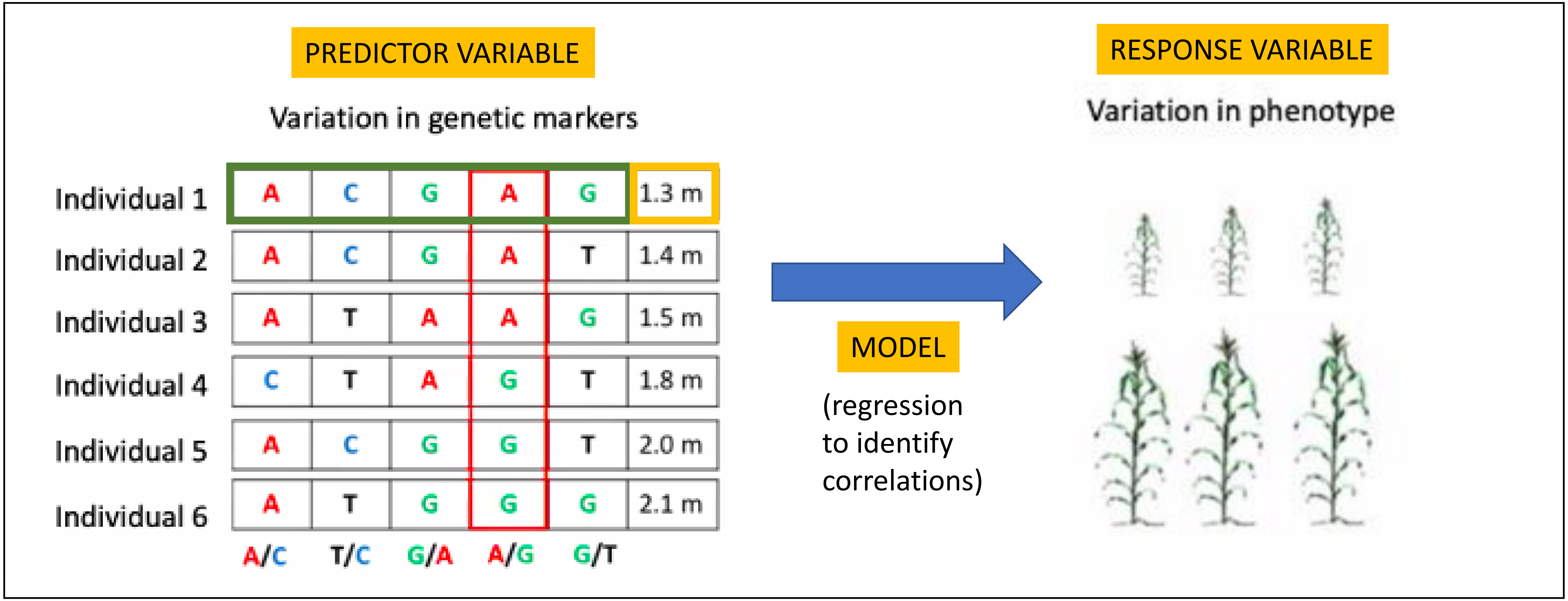
This workspace currently contains no files

HOW DOES CARTOGRAPLANT DO ANALYSIS?

- To perform data analysis, CartograPlant makes use of **Galaxy** (https://galaxyproject.org/)
- Galaxy is a **user-friendly platform** that allows the **bioinformatic analysis of high throughput data without the need of coding skills**.

WHAT IS GWAS?

- A genome-wide association study (GWAS) is an approach used in genetics research to associate specific genetic variations (usually single nucleotide polymorphisms or SNPs) with traits of interest.



PREDICTOR VARIABLE

Variation in genetic markers

Individual 1	A	C	G	A	G	1.3 m
Individual 2	A	C	G	A	T	1.4 m
Individual 3	A	T	A	A	G	1.5 m
Individual 4	C	T	A	G	T	1.8 m
Individual 5	A	C	G	G	T	2.0 m
Individual 6	A	T	G	G	G	2.1 m
	A/C	T/C	G/A	A/G	G/T	

MODEL

(regression to identify correlations)

RESPONSE VARIABLE

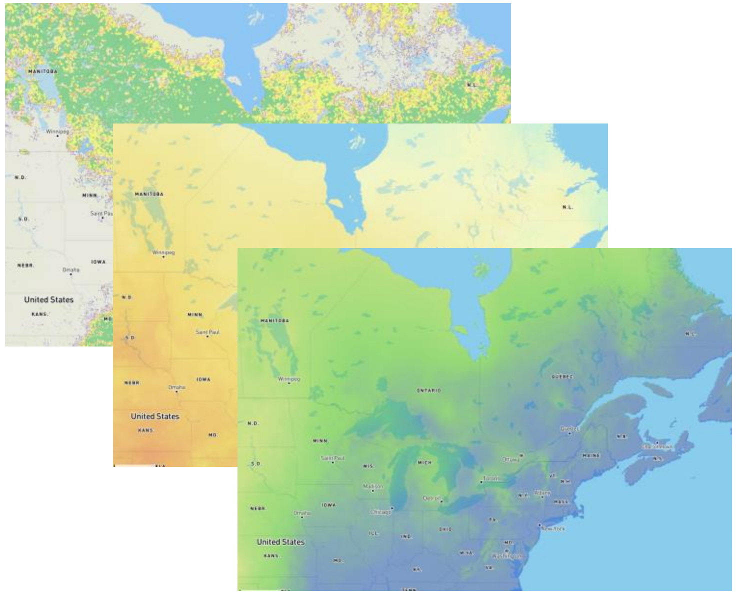
Variation in phenotype

WHAT IS LANDSCAPE GENOMICS?

- Landscape genomics (GxE) a.k.a. GEA is an approach used in genetics research to identify relationships between environmental factors and genetic variations (usually SNPs) to investigate the genetic adaptation of organisms in response to these environmental factors of interest.

PREDICTOR VARIABLE

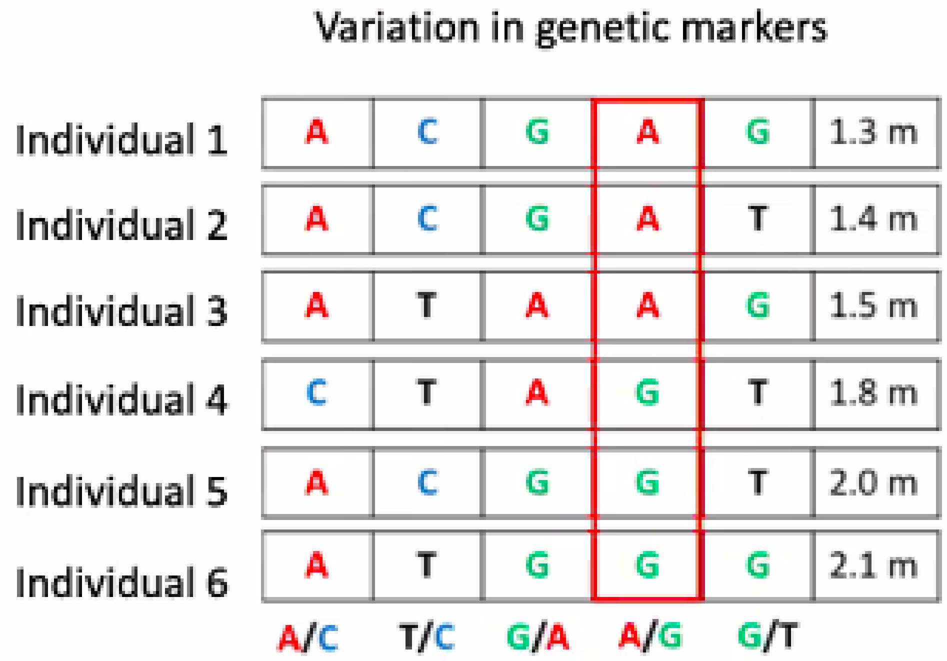
Environmental variables



MODEL

(regression to identify correlations)

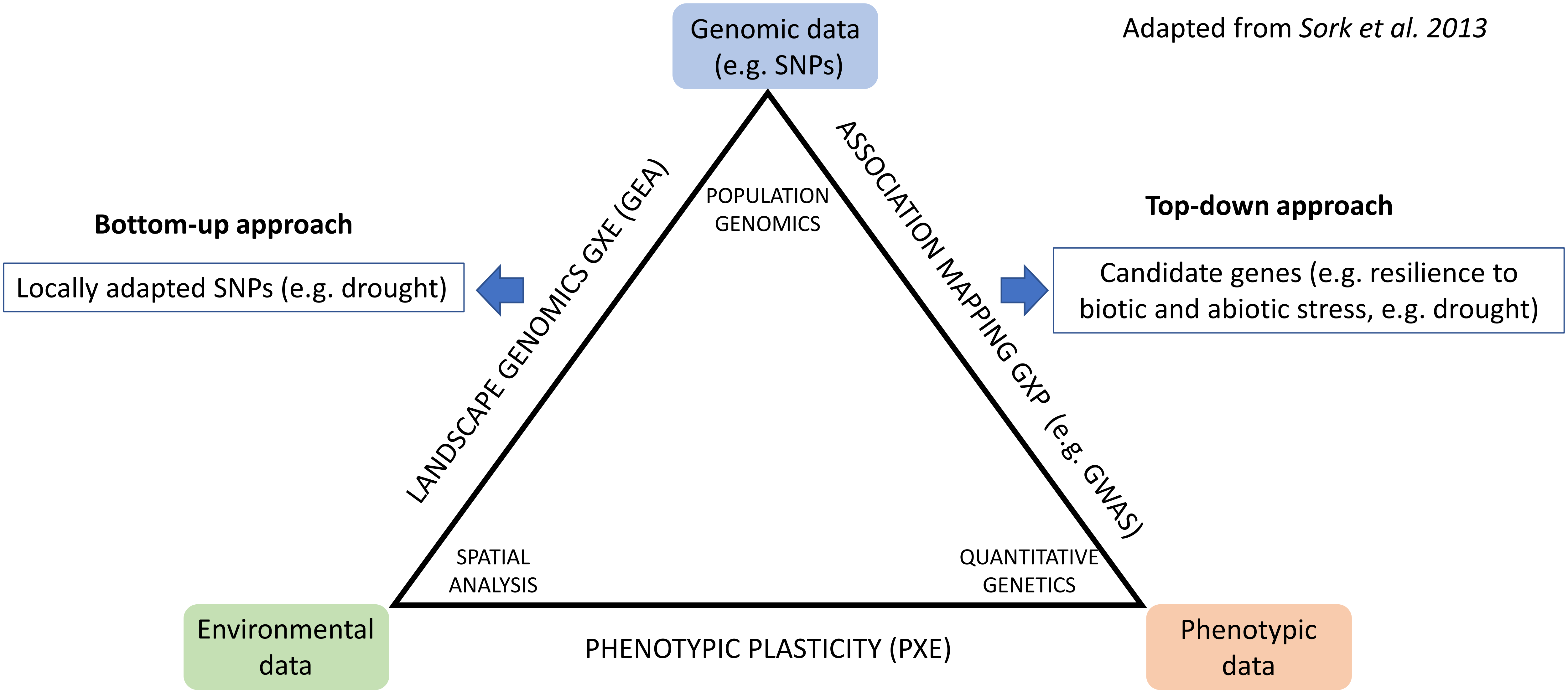
RESPONSE VARIABLE



Variation in genetic markers

Individual 1	A	C	G	A	G	1.3 m
Individual 2	A	C	G	A	T	1.4 m
Individual 3	A	T	A	A	G	1.5 m
Individual 4	C	T	A	G	T	1.8 m
Individual 5	A	C	G	G	T	2.0 m
Individual 6	A	T	G	G	G	2.1 m
	A/C	T/C	G/A	A/G	G/T	

USEFULNESS OF PERFORMING GXP AND GXE TOGETHER



Adapted from *Sork et al. 2013*

Genomic data (e.g. SNPs)

POPULATION GENOMICS

LANDSCAPE GENOMICS GXE (GEA)

ASSOCIATION MAPPING GXP (e.g. GWAS)

Bottom-up approach

Locally adapted SNPs (e.g. drought)

Top-down approach

Candidate genes (e.g. resilience to biotic and abiotic stress, e.g. drought)

SPATIAL ANALYSIS

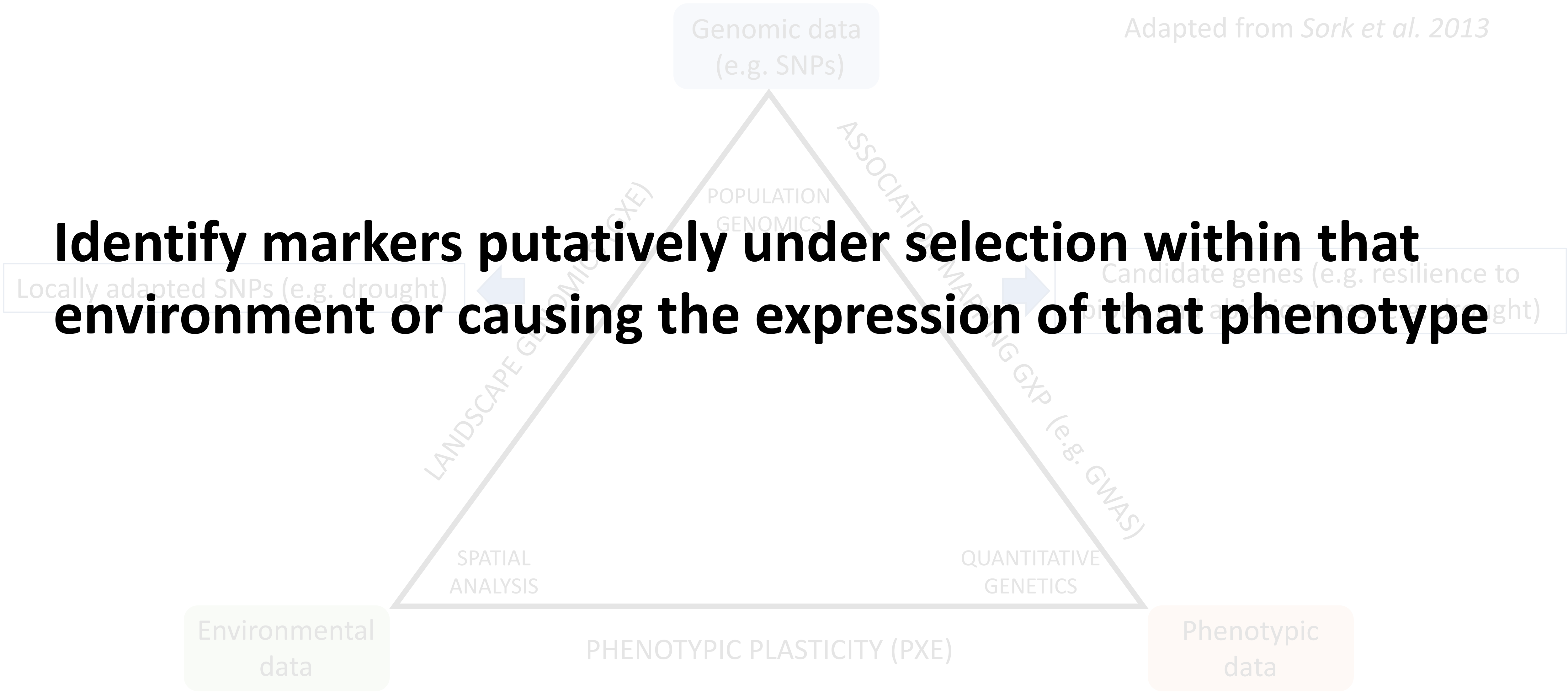
QUANTITATIVE GENETICS

Environmental data

PHENOTYPIC PLASTICITY (PXE)

Phenotypic data

USEFULNESS OF PERFORMING GXP AND GXE TOGETHER



Adapted from *Sork et al. 2013*

Genomic data (e.g. SNPs)

POPULATION GENOMICS

LANDSCAPE GENOMICS (GXE)

ASSOCIATION MAPPING GXP (e.g. GWAS)

Locally adapted SNPs (e.g. drought)

Candidate genes (e.g. resilience to biotic and abiotic stresses e.g. drought)

SPATIAL ANALYSIS

QUANTITATIVE GENETICS

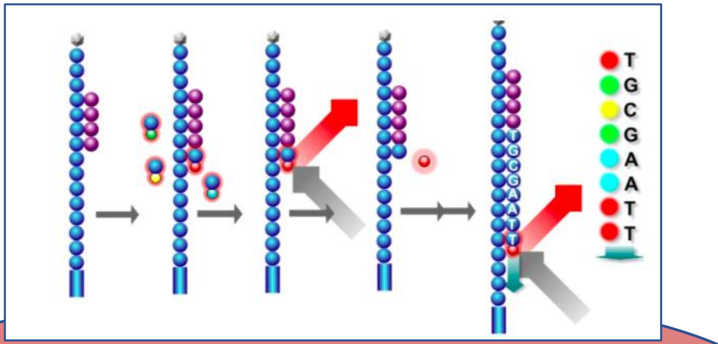
Environmental data

PHENOTYPIC PLASTICITY (PXE)

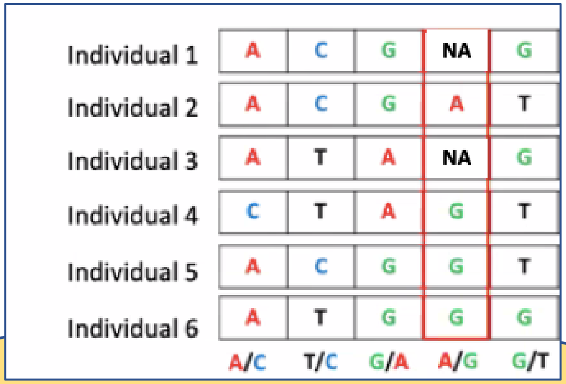
Phenotypic data

Identify markers putatively under selection within that environment or causing the expression of that phenotype

FACTORS THAT CAN AFFECT THE ACCURACY OF THE GWAS/GEA RESULTS

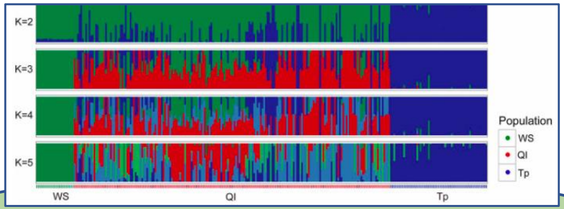


GENOTYPING ERRORS

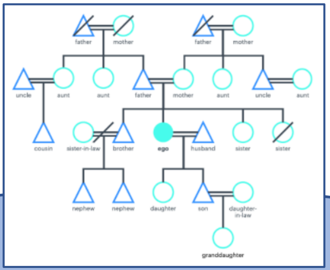


Individual 1	A	C	G	NA	G
Individual 2	A	C	G	A	T
Individual 3	A	T	A	NA	G
Individual 4	C	T	A	G	T
Individual 5	A	C	G	G	T
Individual 6	A	T	G	G	G
	A/C	T/C	G/A	A/G	G/T

MISSING SNPs

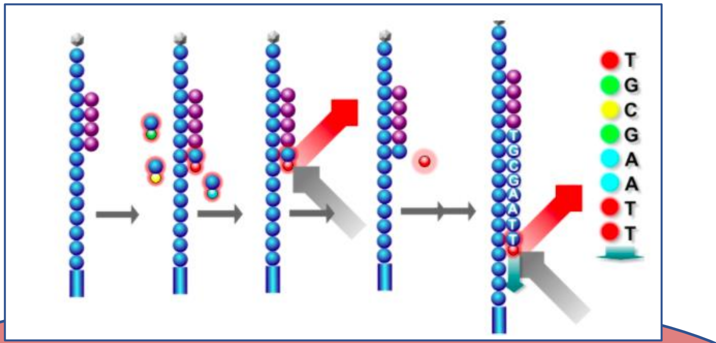


POPULATION STRUCTURE/KINSHIP

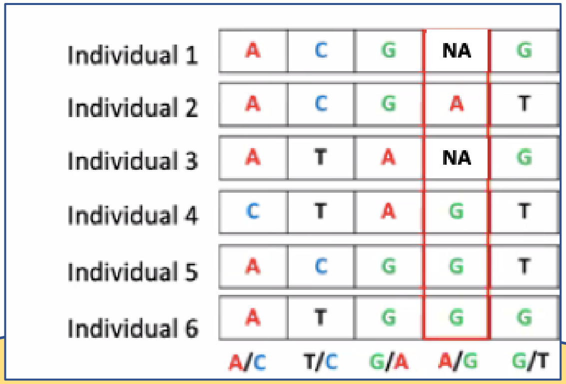


KINSHIP

FACTORS THAT CAN AFFECT THE ACCURACY OF THE GWAS/GEA RESULTS

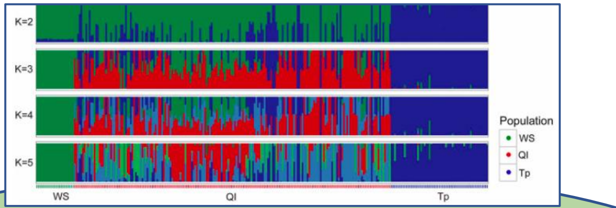


GENOTYPING ERRORS

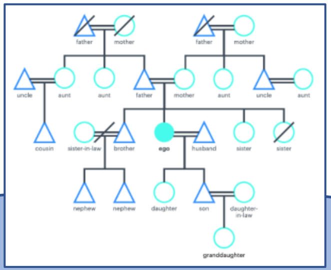


Individual 1	A	C	G	NA	G
Individual 2	A	C	G	A	T
Individual 3	A	T	A	NA	G
Individual 4	C	T	A	G	T
Individual 5	A	C	G	G	T
Individual 6	A	T	G	G	G
	A/C	T/C	G/A	A/G	G/T

MISSING SNPs



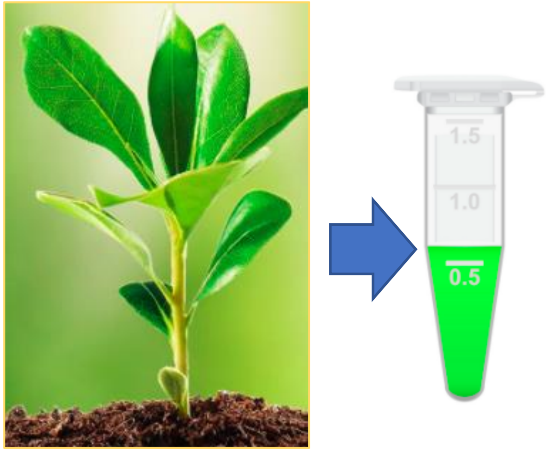
POPULATION STRUCTURE/KINSHIP

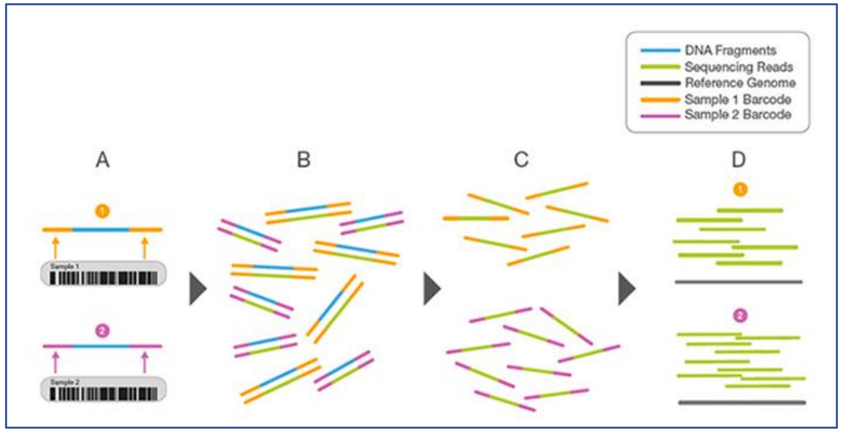


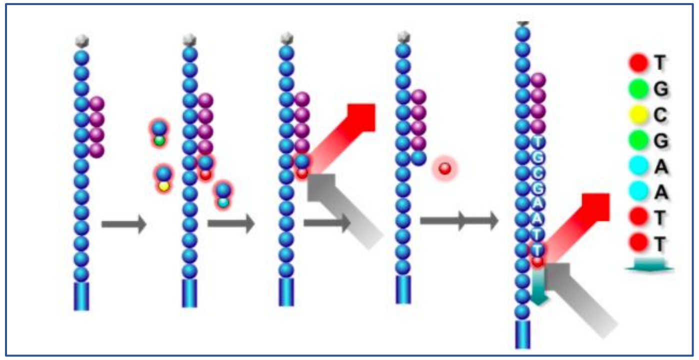
KINSHIP

GENOTYPING ERRORS

- A **genotype error** describes when the **observed genotype** does not correspond with the **true genotype**





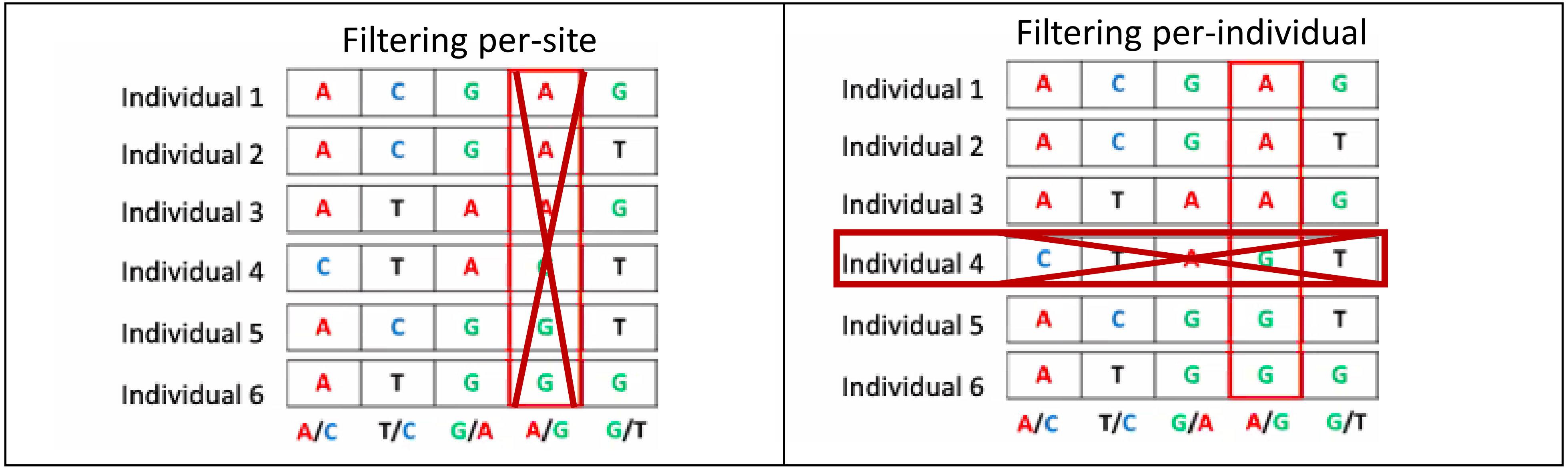


SAMPLE PREPARATION → LIBRARY PREPARATION → SEQUENCING

GENOTYPING ERRORS

SOLUTION:

- **Filtering Data Based on Quality Metrics** (prior to analysis)
- **Remove markers and individuals:**
  - Low quality scores
  - High levels of missing genotypes
  - Minimum genotype counts (minor allele count)



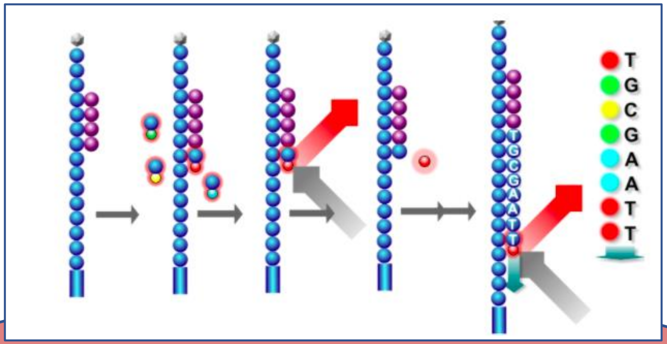
Filtering per-site

Individual 1	A	C	G	A	G
Individual 2	A	C	G	A	T
Individual 3	A	T	A	A	G
Individual 4	C	T	A	G	T
Individual 5	A	C	G	G	T
Individual 6	A	T	G	G	G
	A/C	T/C	G/A	A/G	G/T

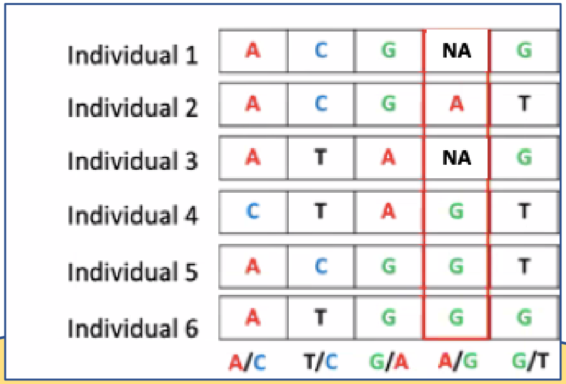
Filtering per-individual

Individual 1	A	C	G	A	G
Individual 2	A	C	G	A	T
Individual 3	A	T	A	A	G
Individual 4	C	T	A	G	T
Individual 5	A	C	G	G	T
Individual 6	A	T	G	G	G
	A/C	T/C	G/A	A/G	G/T

FACTORS THAT CAN AFFECT THE ACCURACY OF THE GWAS/GEA RESULTS

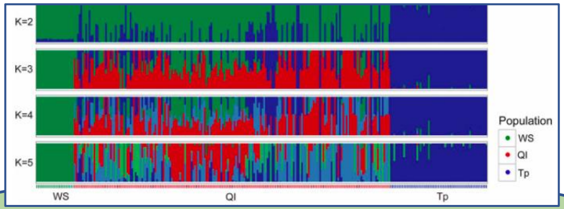


GENOTYPING ERRORS

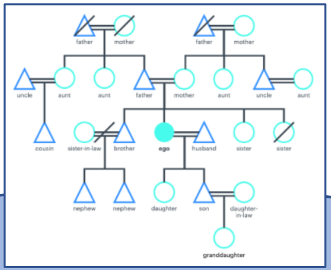


Individual 1	A	C	G	NA	G
Individual 2	A	C	G	A	T
Individual 3	A	T	A	NA	G
Individual 4	C	T	A	G	T
Individual 5	A	C	G	G	T
Individual 6	A	T	G	G	G
	A/C	T/C	G/A	A/G	G/T

MISSING SNPs



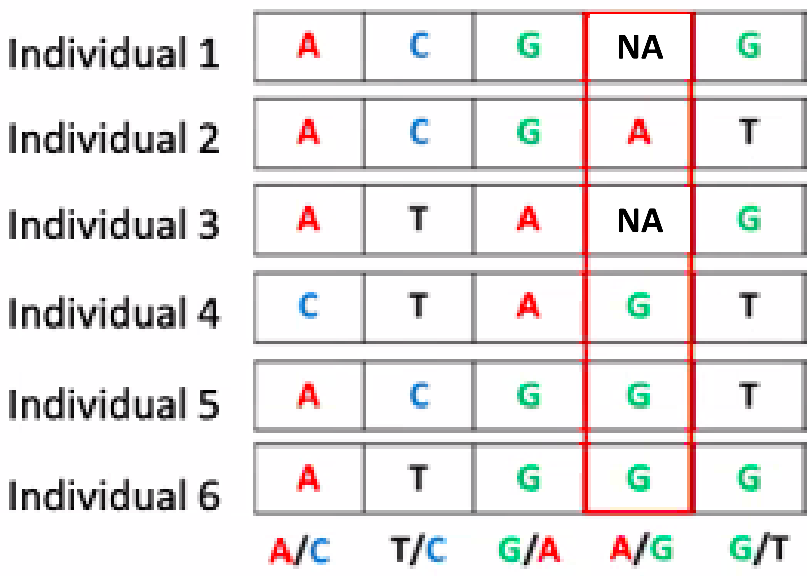
POPULATION STRUCTURE/KINSHIP



KINSHIP

MISSING SNPs

- Even after performing data quality filtering, **missing SNPs may remain in our dataset.**
- Missing SNPs can **reduce the statistical power of our GWAS/GEA**, especially in the case of low density SNP panels.



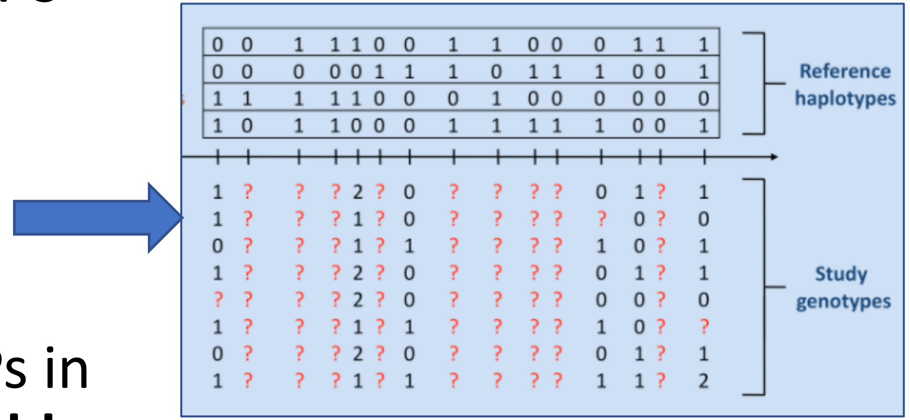
Individual 1	A	C	G	NA	G
Individual 2	A	C	G	A	T
Individual 3	A	T	A	NA	G
Individual 4	C	T	A	G	T
Individual 5	A	C	G	G	T
Individual 6	A	T	G	G	G
	A/C	T/C	G/A	A/G	G/T

MISSING SNPs

SOLUTION

Imputation of lacking SNPs

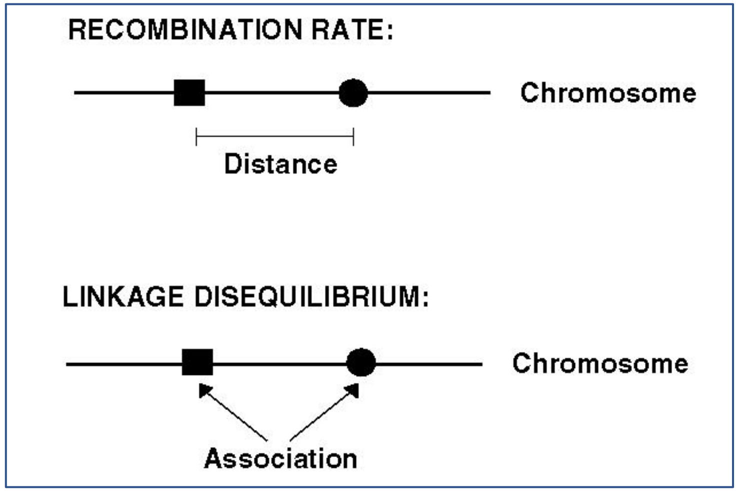
- This imputation or inference of the lacking SNPs can be performed by different methods:
  - **Using a Reference panel of SNPs** (same species, similar population): **Non-suitable for non-model species**
  - **Based on Linkage disequilibrium**: Looking for SNPs in linkage disequilibrium with the lacking SNPs: **suitable for non-model species**

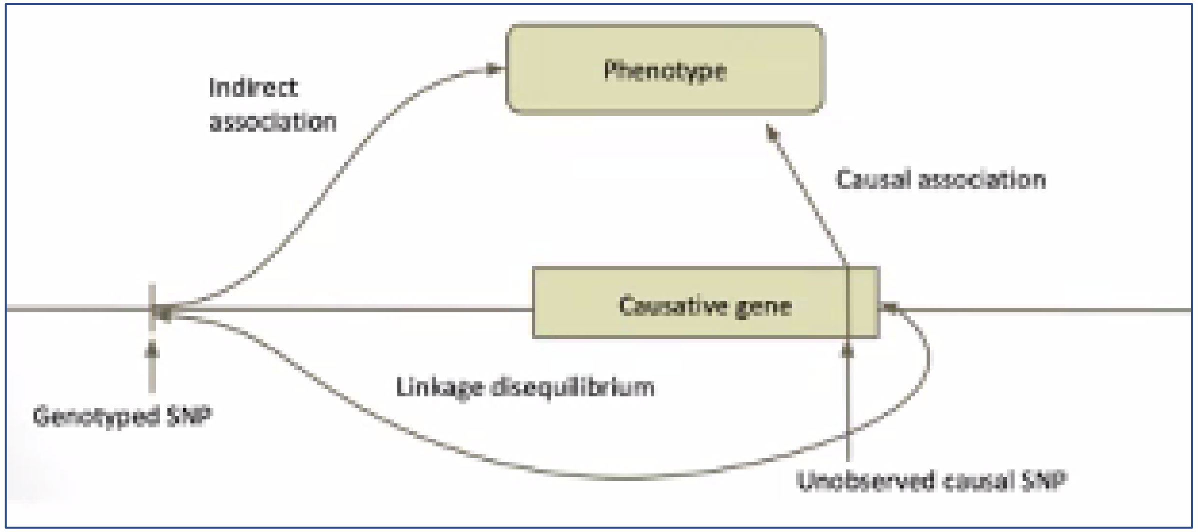


Reference haplotypes	0	0	1	1	1	0	0	1	1	0	0	0	1	1	1
	0	0	0	0	0	1	1	1	0	1	1	1	0	0	1
	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0
	1	0	1	1	0	0	0	1	1	1	1	1	0	0	1
Study genotypes	1	?	?	?	2	?	0	?	?	?	?	0	1	?	1
	1	?	?	?	1	?	0	?	?	?	?	?	0	?	0
	0	?	?	?	1	?	1	?	?	?	?	1	0	?	1
	1	?	?	?	2	?	0	?	?	?	?	0	1	?	1
	?	?	?	?	2	?	0	?	?	?	?	0	0	?	0
	1	?	?	?	1	?	1	?	?	?	?	1	0	?	?
	0	?	?	?	2	?	0	?	?	?	?	0	1	?	1
	1	?	?	?	1	?	1	?	?	?	?	1	1	?	2

WHAT IS LINKAGE DISEQUILIBRIUM (LD)?

- **Non-random association between two or more loci in the population** (two or more loci that tend to be inherited together)
- **The closer two markers are, the higher LD** (due to recombination)
- It is also the basis of the **identification of causative/locally adapted genes in GWAS/GEA**





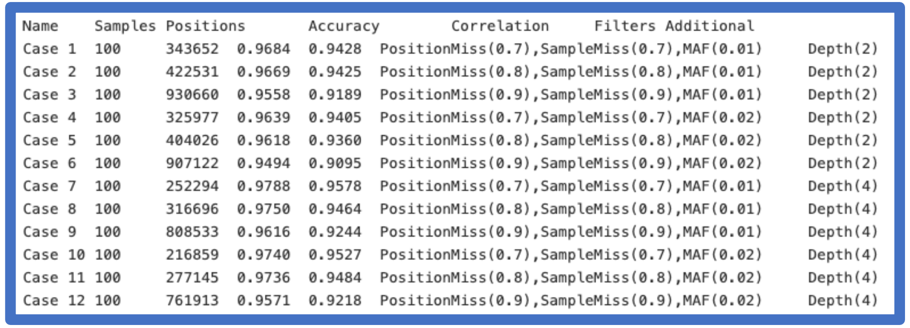
MISSING SNPs

SOLUTION

LinkImputeR

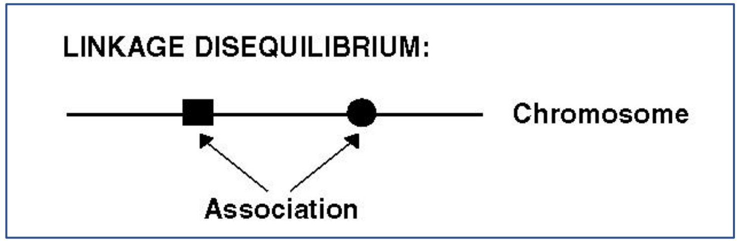
- **Tests** for different combinations of SNP data **quality filtering thresholds** (aka “Cases”).
- **Objective**: **Maximize the quantity and quality** of the resulting SNPs, while maintaining **accuracy**.
- **Imputation** for **non-model organisms**
  - No requires ordered markers or a reference panel
  - **LD-kNN Imputation, based on Linkage Disequilibrium** (LD)

CASES

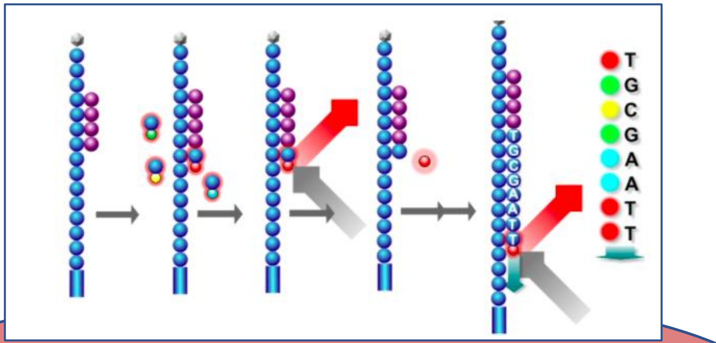


```
Name     Samples Positions          Accuracy          Correlation      Filters Additional
Case 1   100       343652  0.9684  0.9428  PositionMiss(0.7),SampleMiss(0.7),MAF(0.01)      Depth(2)
Case 2   100       422531  0.9669  0.9425  PositionMiss(0.8),SampleMiss(0.8),MAF(0.01)      Depth(2)
Case 3   100       930660  0.9558  0.9189  PositionMiss(0.9),SampleMiss(0.9),MAF(0.01)      Depth(2)
Case 4   100       325977  0.9639  0.9405  PositionMiss(0.7),SampleMiss(0.7),MAF(0.02)      Depth(2)
Case 5   100       404026  0.9618  0.9360  PositionMiss(0.8),SampleMiss(0.8),MAF(0.02)      Depth(2)
Case 6   100       907122  0.9494  0.9095  PositionMiss(0.9),SampleMiss(0.9),MAF(0.02)      Depth(2)
Case 7   100       252294  0.9788  0.9578  PositionMiss(0.7),SampleMiss(0.7),MAF(0.01)      Depth(4)
Case 8   100       316696  0.9750  0.9464  PositionMiss(0.8),SampleMiss(0.8),MAF(0.01)      Depth(4)
Case 9   100       808533  0.9616  0.9244  PositionMiss(0.9),SampleMiss(0.9),MAF(0.01)      Depth(4)
Case 10 100       216859  0.9740  0.9527  PositionMiss(0.7),SampleMiss(0.7),MAF(0.02)      Depth(4)
Case 11 100       277145  0.9736  0.9484  PositionMiss(0.8),SampleMiss(0.8),MAF(0.02)      Depth(4)
Case 12 100       761913  0.9571  0.9218  PositionMiss(0.9),SampleMiss(0.9),MAF(0.02)      Depth(4)
```

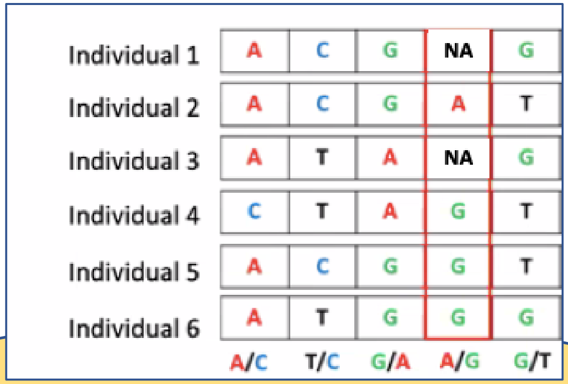
IMPUTATION



FACTORS THAT CAN AFFECT THE ACCURACY OF THE GWAS/GEA RESULTS

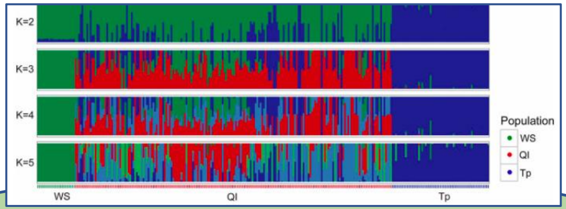


GENOTYPING ERRORS

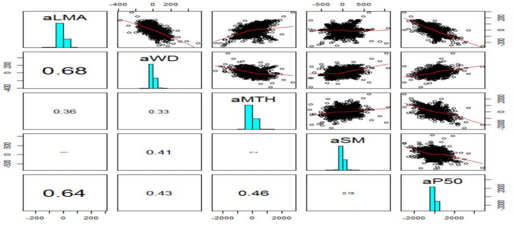


Individual 1	A	C	G	NA	G
Individual 2	A	C	G	A	T
Individual 3	A	T	A	NA	G
Individual 4	C	T	A	G	T
Individual 5	A	C	G	G	T
Individual 6	A	T	G	G	G
	A/C	T/C	G/A	A/G	G/T

MISSING SNPs



POPULATION STRUCTURE/KINSHIP

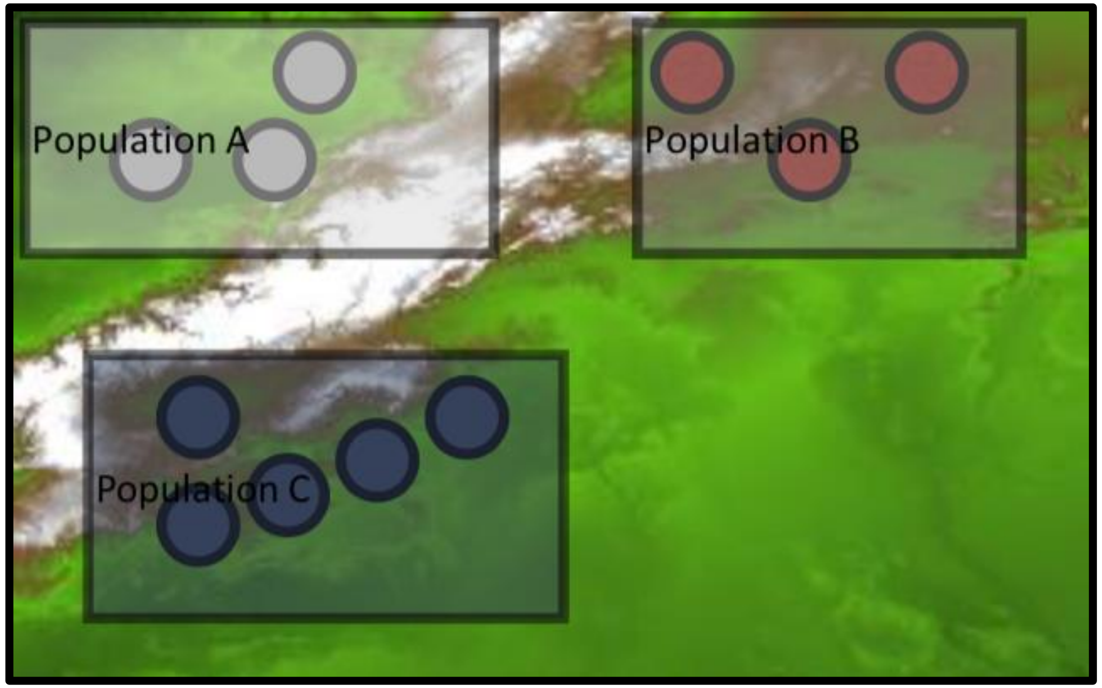


MULTICOLLINEARITY

POPULATION STRUCTURE/KINSHIP CORRECTION

SOLUTION:

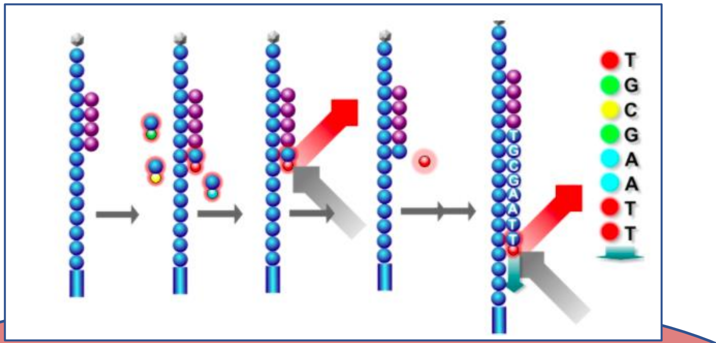
- Account for **confounding variables** in the model
- **Objective of GWAS/GEA**: To identify **alleles associated with** the **traits/environmental variables of interest**.
- Reduce **spurious associations or false positives** in GWAS/GEA
  - **Population structure:** patterns caused by by isolation by distance, genetic drift, kinship etc.
- These hidden sample structures can act as **confounding variables**, leading to false positives or spurious associations



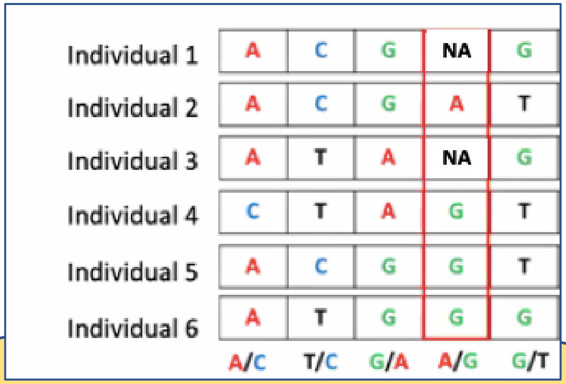
Example of confounding variable

Sometimes the population structure **mimic** the environmental variable of interest

FACTORS THAT CAN AFFECT THE ACCURACY OF THE GWAS/GEA RESULTS

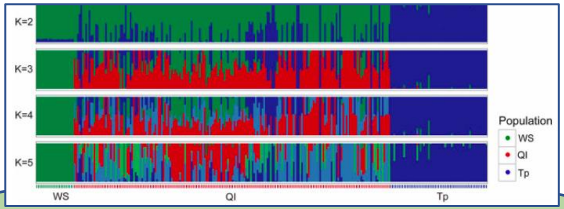


GENOTYPING ERRORS

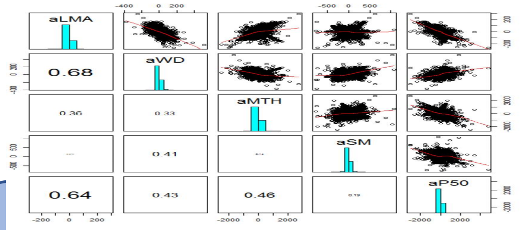


Individual 1	A	C	G	NA	G
Individual 2	A	C	G	A	T
Individual 3	A	T	A	NA	G
Individual 4	C	T	A	G	T
Individual 5	A	C	G	G	T
Individual 6	A	T	G	G	G
	A/C	T/C	G/A	A/G	G/T

MISSING SNPs



POPULATION STRUCTURE/KINSHIP



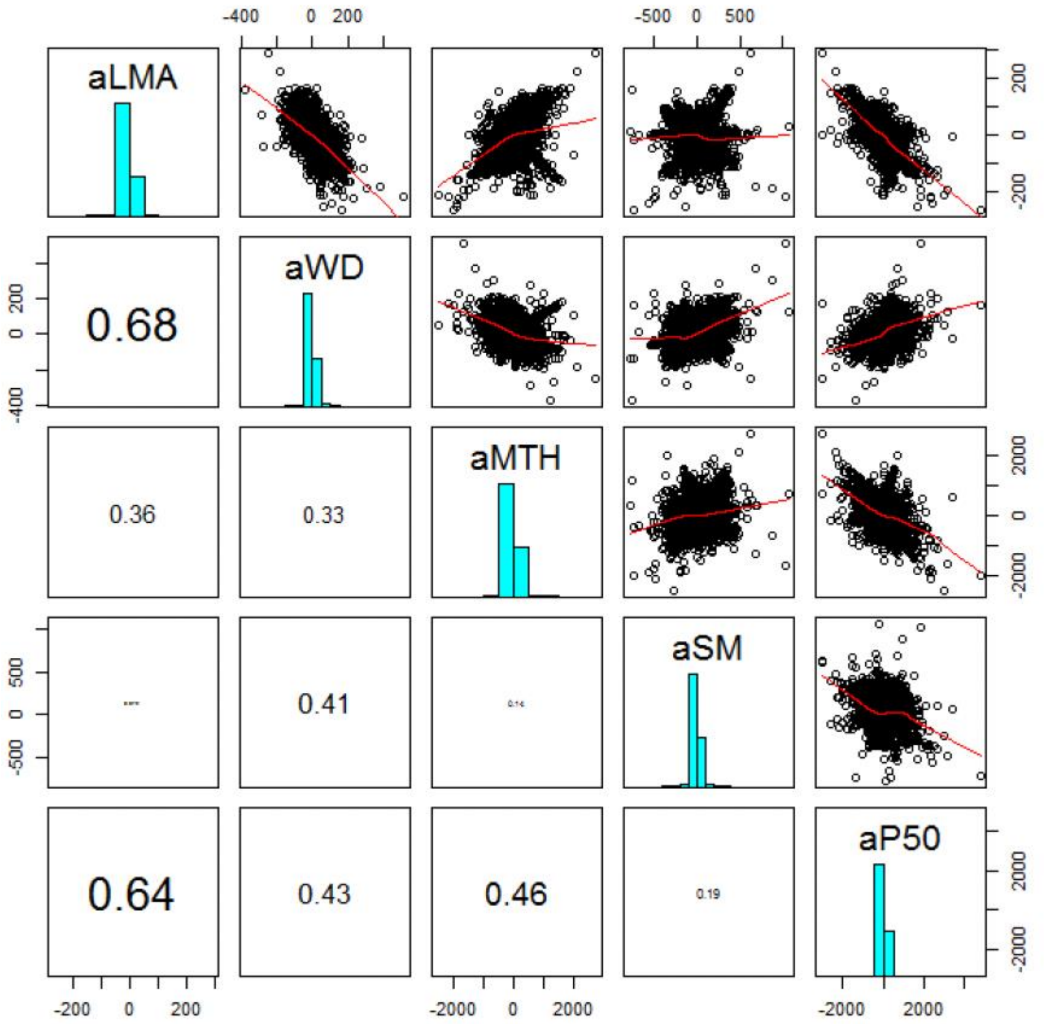
MULTICOLLINEARITY

OTHER GxE ISSUES: MULTICOLLINEARITY

- **Collinearity** refers to a linear relationship between two explanatory variables (environmental variables in our case).

SOLUTION

- **Variable selection is a critical preliminary step** in ecological modelling, including landscape genomics, **to avoid overfitting and collinearity**
- To help with this variable reduction decisions, CartograPlant offers **Pearson correlation analyses of the environmental variables of interest, plotted in pairwise scatterplot matrices**

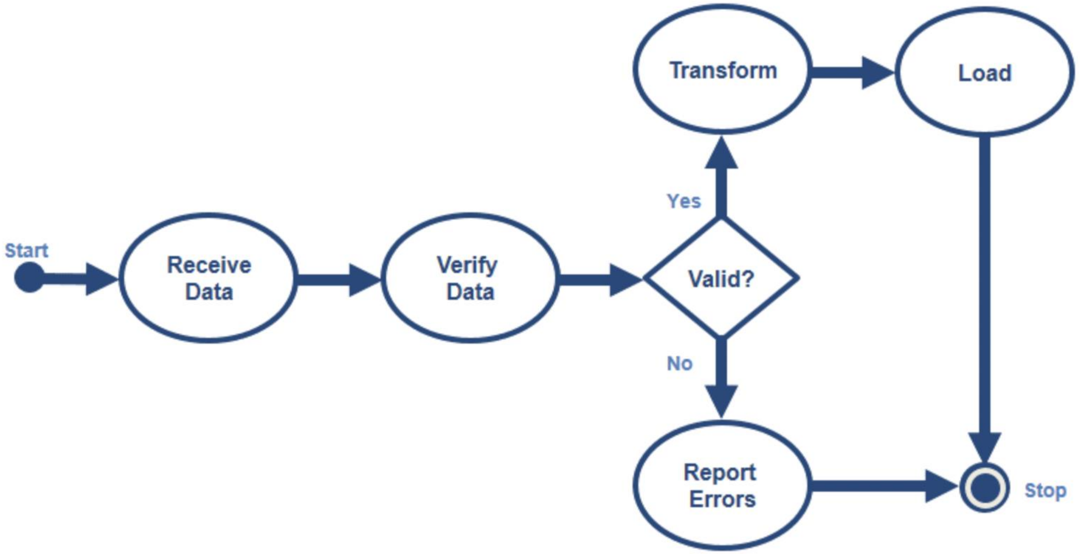


CARTOGRAPLANT ANALYTIC WORKFLOWS OBJECTIVES

- Genome Wide Association Analysis (GWAS, G X P)

- Landscape genomics (G X E)

- Mega/meta-analysis


Start

Receive Data

Verify Data

Valid?

Yes

Transform

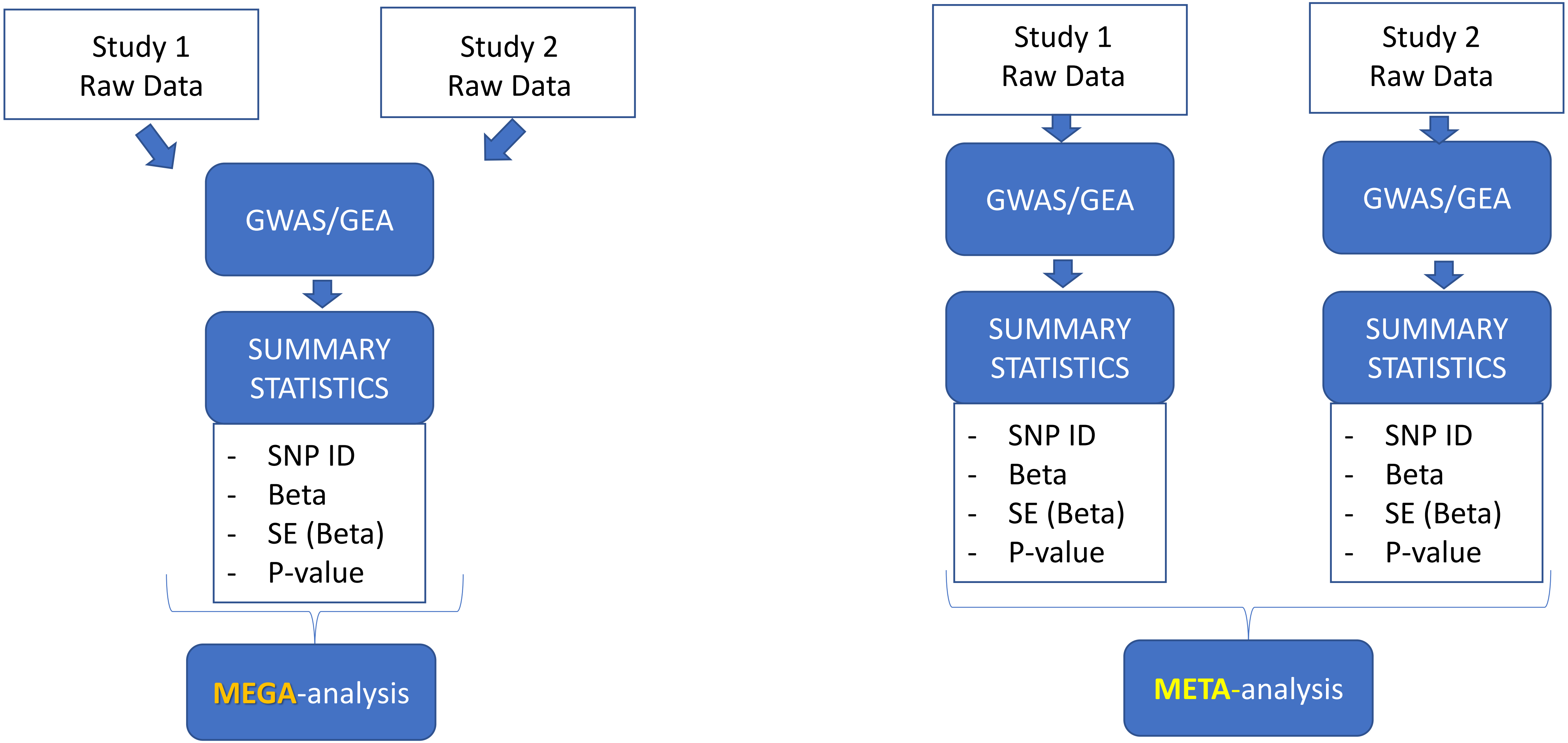
Load

No

Report Errors

Stop

WHAT ARE META-ANALYSIS AND MEGA-ANALYSIS?



Study 1
Raw Data

Study 2
Raw Data

GWAS/GEA

SUMMARY STATISTICS

- SNP ID
- Beta
- SE (Beta)
- P-value

MEGA-analysis

Study 1
Raw Data

GWAS/GEA

SUMMARY STATISTICS

- SNP ID
- Beta
- SE (Beta)
- P-value

Study 2
Raw Data

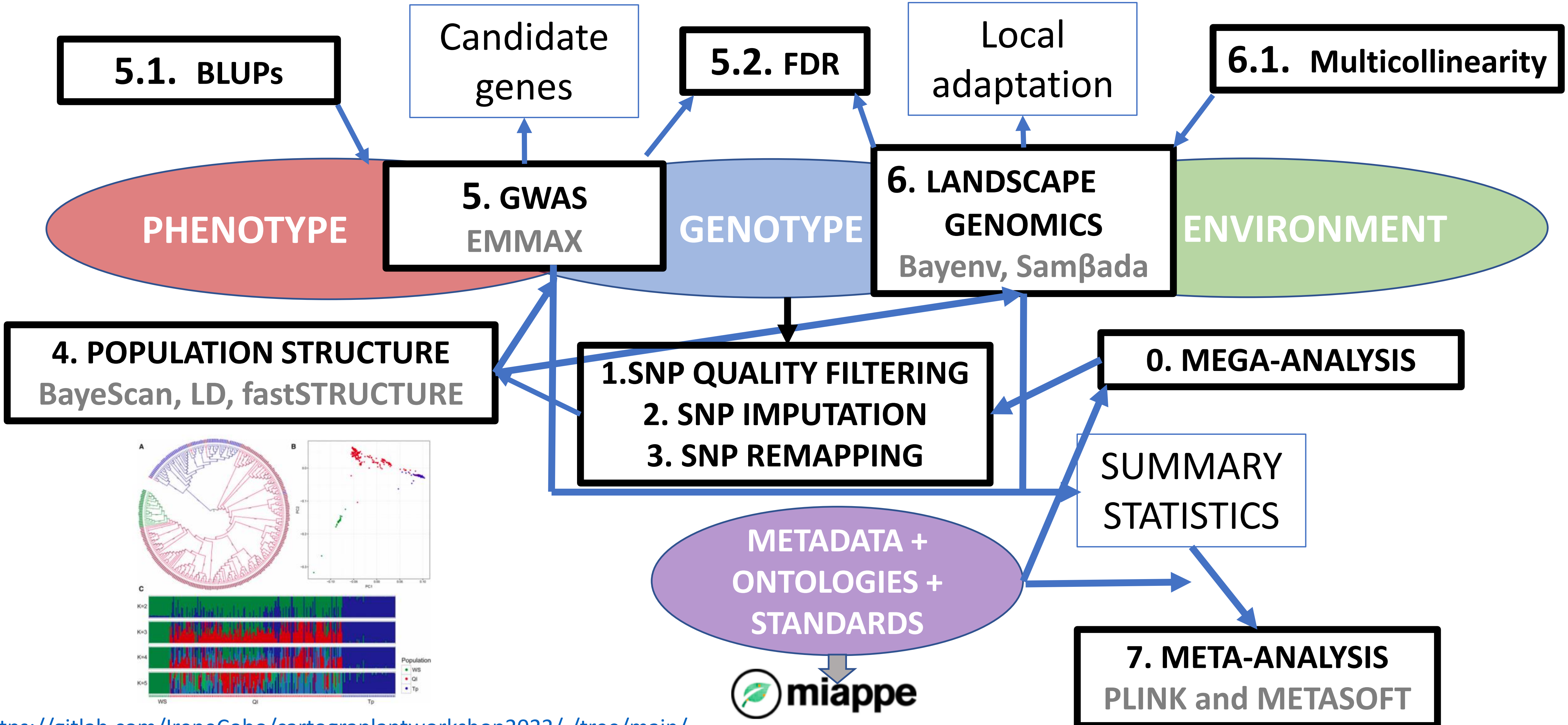
GWAS/GEA

SUMMARY STATISTICS

- SNP ID
- Beta
- SE (Beta)
- P-value

META-analysis

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

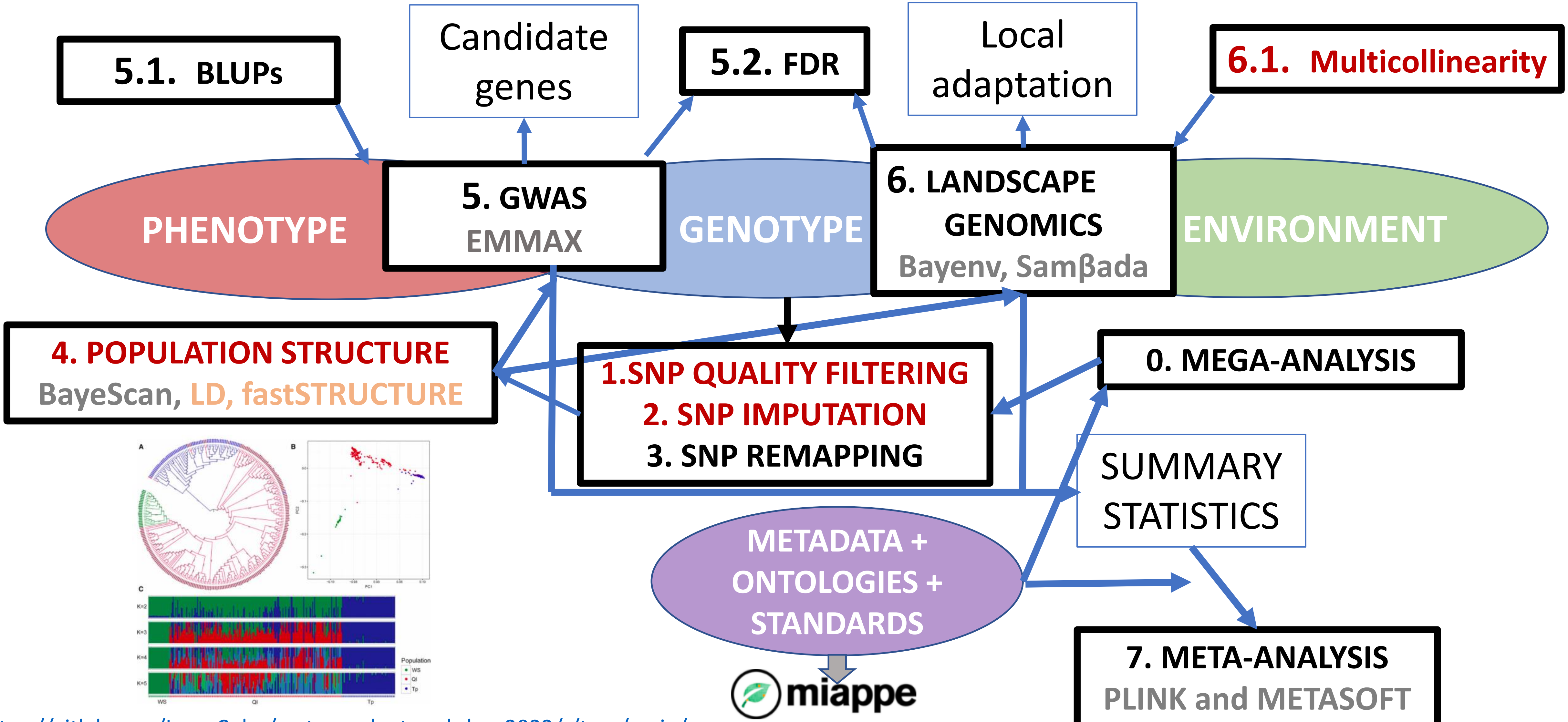
METADATA + ONTOLOGIES + STANDARDS

miappe

7. META-ANALYSIS
PLINK and METASOFT

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

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

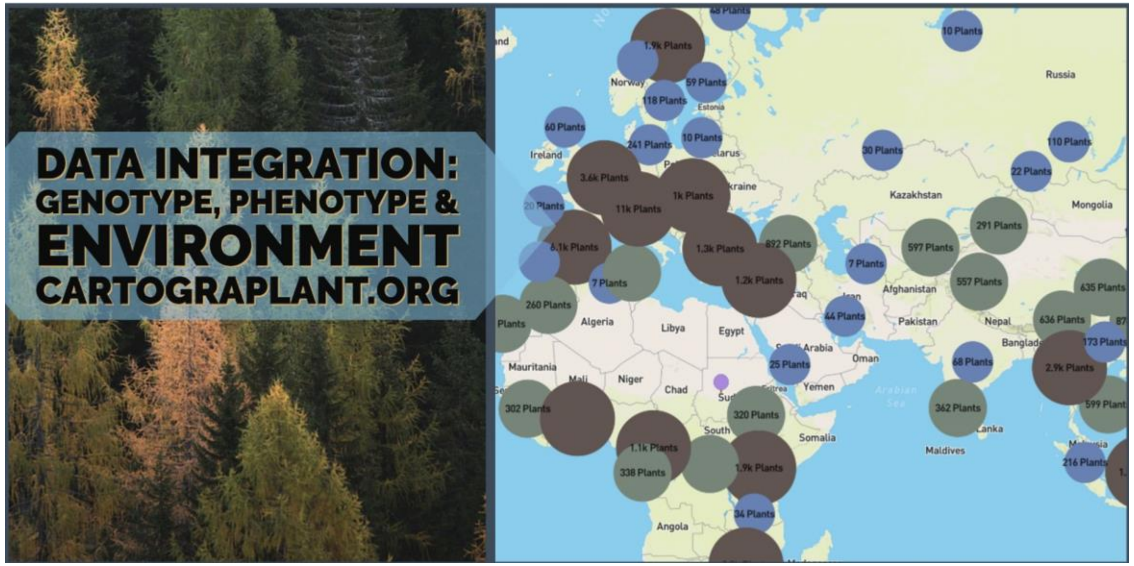
miappe

7. META-ANALYSIS
PLINK and METASOFT

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

CONCLUSIONS

- Cartograplant provides **flexible analytic workflows**
- **Native support** for **mega** and **meta-analysis**
- Is currently being **maintained** by a team of software developers
  - Constantly **adding new features** and updates to stay **up to date** with **current research standards**
- A **centralized platform** to integrate, visualize and analyze high-throughput biological data for plants



Environment Layers OFF ON

Number of Plants 239363

Selected Plants 0

Species count 424

Publications Count 307

Number of Layers 0

Today’s Schedule (in EST!)

11:00-11:20 Introduction to CartograPlant - Dr. Jill Wegrzyn

11:20-11:40 Introduction to Data Submission with TPPS/TPPSc - Emily Grau

11:40-12:00 Introduction to Data Collection/Mobile Phenotyping with TreeSnap - Dr. Margaret Staton

12:00-12:15 Break

12:15-12:35 Behind the Scenes of CartograPlant - Environmental Layers and Data - Risharde Ramnath

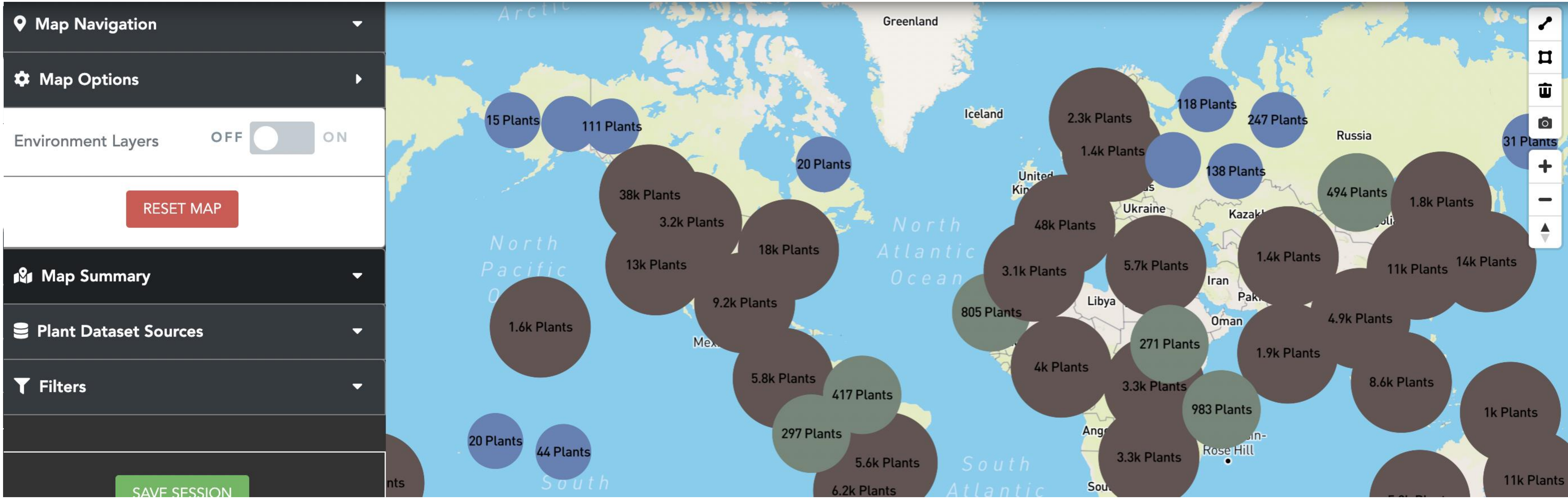
12:35-12:55 Analytics with CartograPlant (GWAS and GEA). Part 1 - Gabriel Barrett

12:55-1:15 Analytics with CartograPlant (GWAS and GEA). Part 2 – Dr. Irene Cobo-Simon

1:15-1:30 Q&A

AG2PI WORKSHOP | JULY 12, 2023

Analytics with CartograPlant (GWAS and GEA): Demo



Irene Cobo Simón, PhD.

Postdoctoral Research Associate/Postdoctoral Fellow

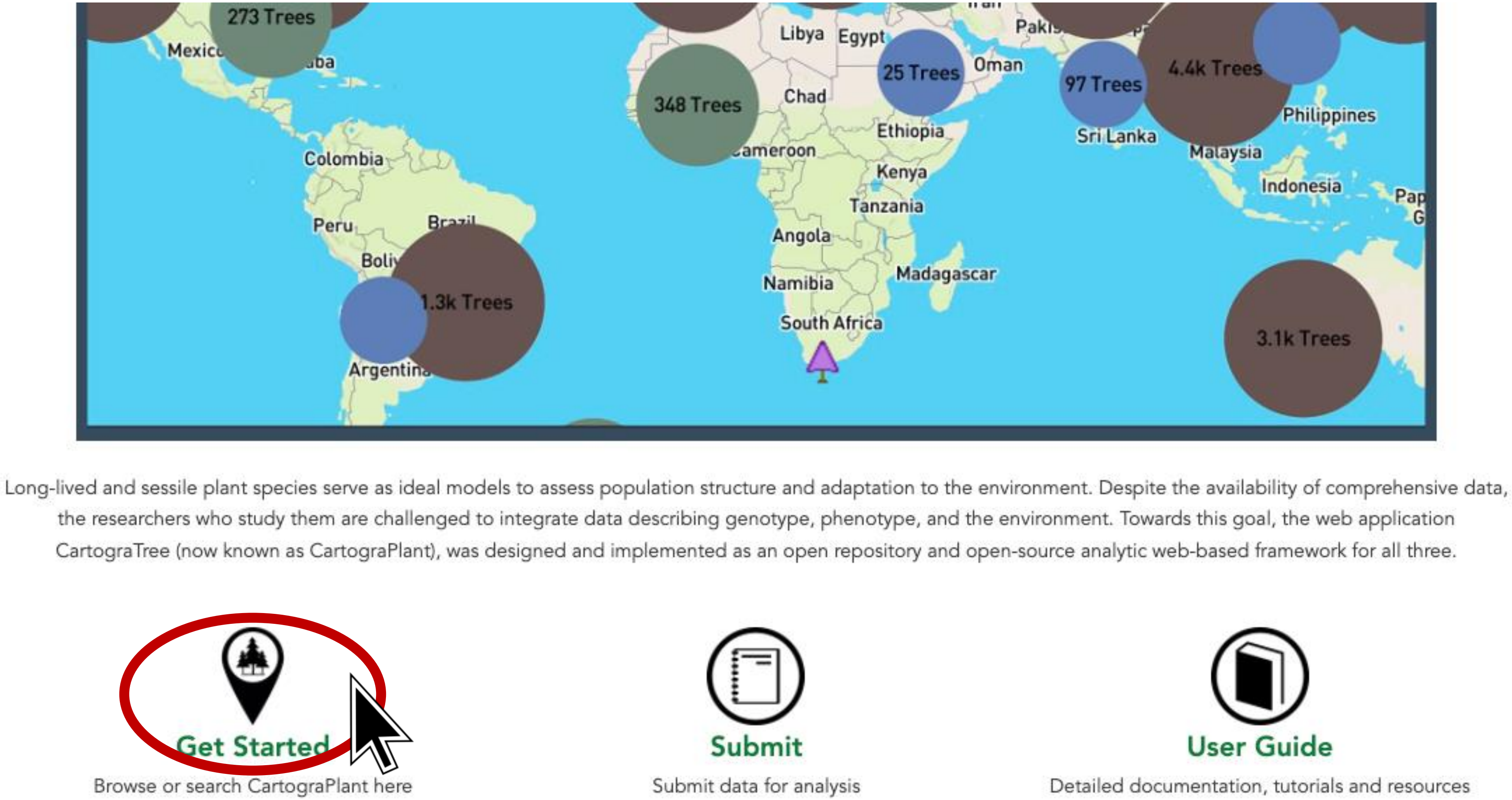
University of Connecticut/Institute of Forest Science (ICIFOR-INIA, CSIC), Spain

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

INIA
Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria

UCONN
UNIVERSITY OF CONNECTICUT

CARTOGRAPLANT WEBSITE



Long-lived and sessile plant species serve as ideal models to assess population structure and adaptation to the environment. Despite the availability of comprehensive data, the researchers who study them are challenged to integrate data describing genotype, phenotype, and the environment. Towards this goal, the web application CartograTree (now known as CartograPlant), was designed and implemented as an open repository and open-source analytic web-based framework for all three.

Browse or search CartograPlant here

Submit data for analysis

Detailed documentation, tutorials and resources

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/

INTERFACE AND DATA VISUALIZATION

Environment Layers OFF ON

Number of Plants	239363	SEL ALL
Selected Plants	0	⊙
Species count	424	
Publications Count	307	
Number of Layers	0	

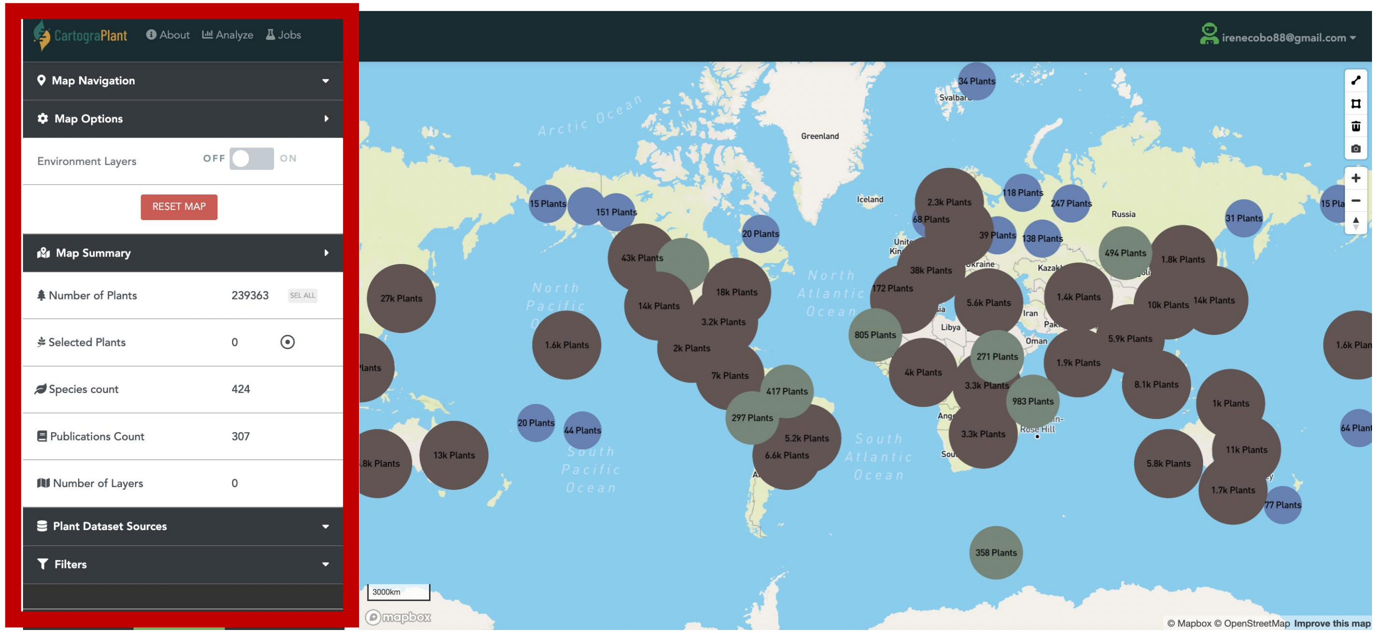
https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/

INTERFACE AND DATA VISUALIZATION

LEFT PANEL

The **action panel** from which users can interact with the map and the plants. It is located at the left of the screen



Environment Layers OFF ON

Number of Plants 239363

Selected Plants 0

Species count 424

Publications Count 307

Number of Layers 0

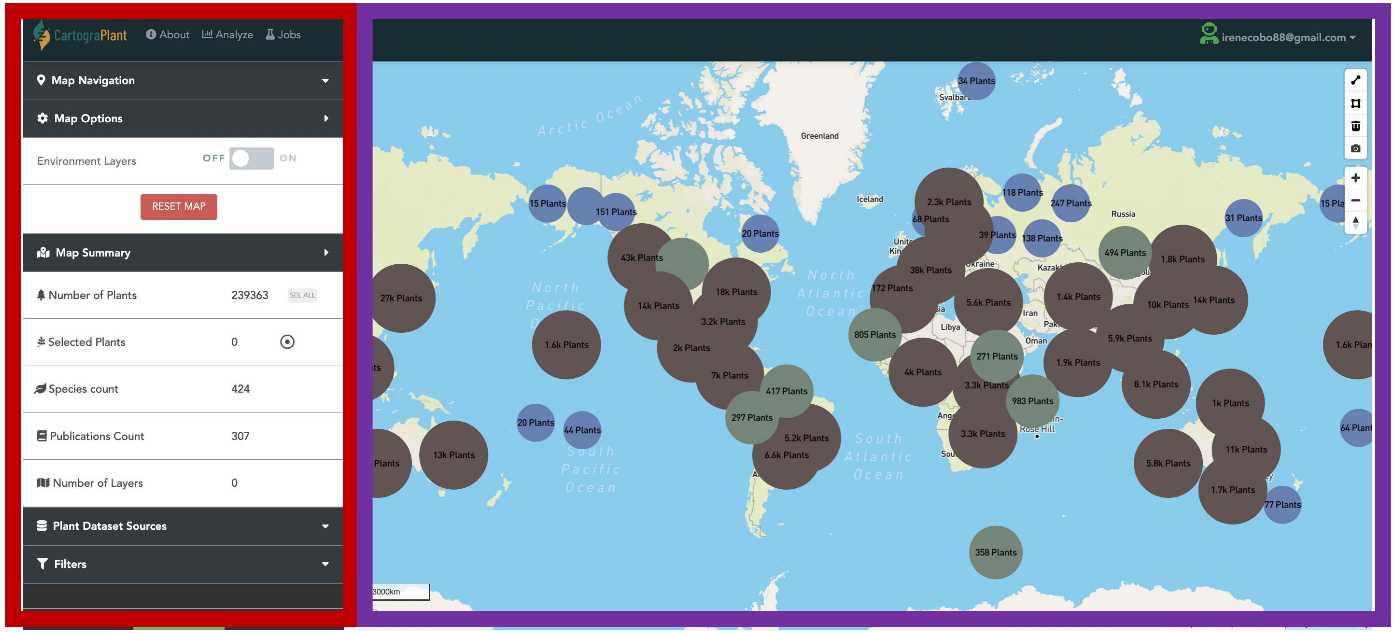
https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/

INTERFACE AND DATA VISUALIZATION

LEFT PANEL

The **action panel** from which users can interact with the map and the plants. It is located at the left of the screen



Environment Layers OFF ON

RESET MAP

Number of Plants 239363 SEL ALL

Selected Plants 0

Species count 424

Publications Count 307

Number of Layers 0

RIGHT PANEL

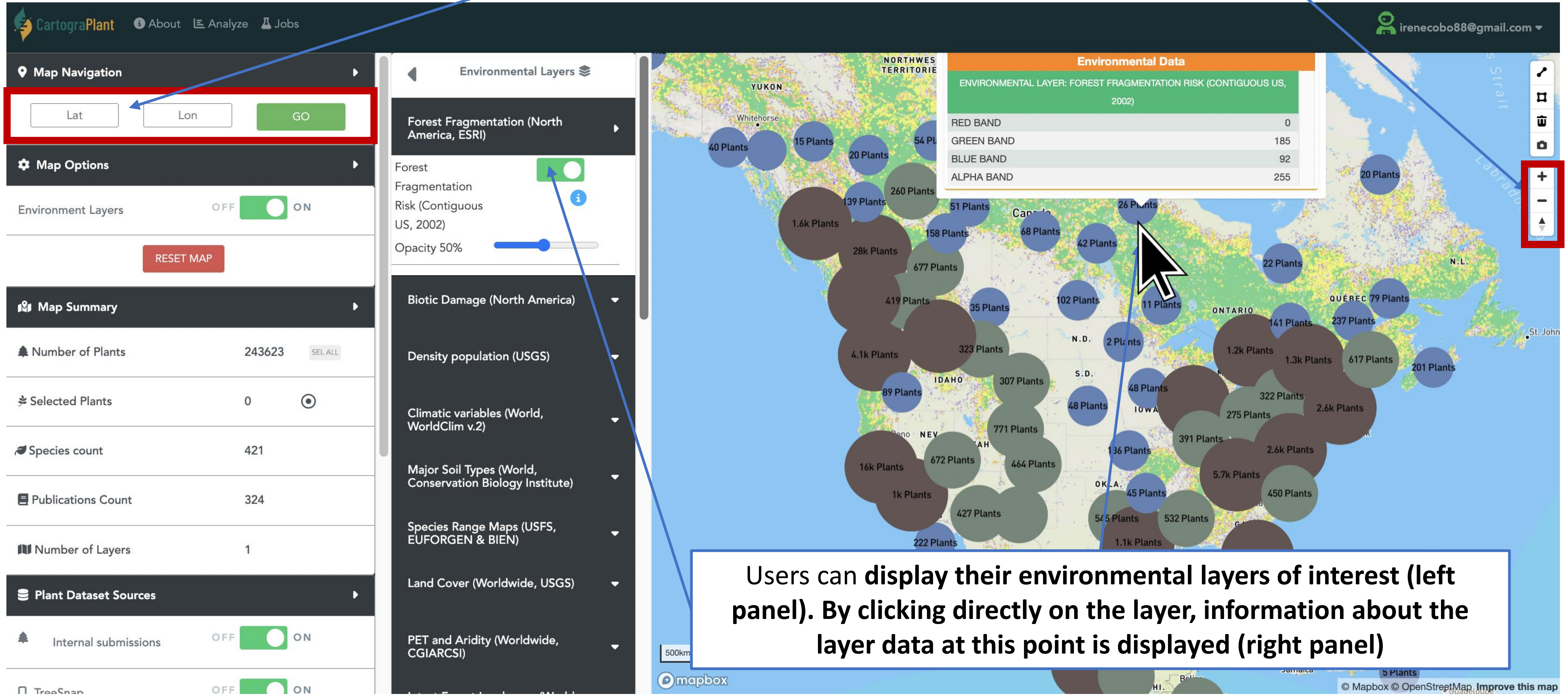
An interactive map, showing the plants and the environmental layers that were selected on the left panel

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/

INTERFACE AND DATA VISUALIZATION

Users can **navigate** on the map by using the **zoom buttons** or by **typing the coordinates of interest**



Environmental Data

ENVIRONMENTAL LAYER: FOREST FRAGMENTATION RISK (CONTIGUOUS US, 2002)

RED BAND	0
GREEN BAND	185
BLUE BAND	92
ALPHA BAND	255

Map Summary

Number of Plants	243623
Selected Plants	0
Species count	421
Publications Count	324
Number of Layers	1

Forest Fragmentation Risk (Contiguous US, 2002): ON

Opacity 50%

Environment Layers: ON

Internal submissions: ON

Users can **display their environmental layers of interest (left panel). By clicking directly on the layer, information about the layer data at this point is displayed (right panel)**

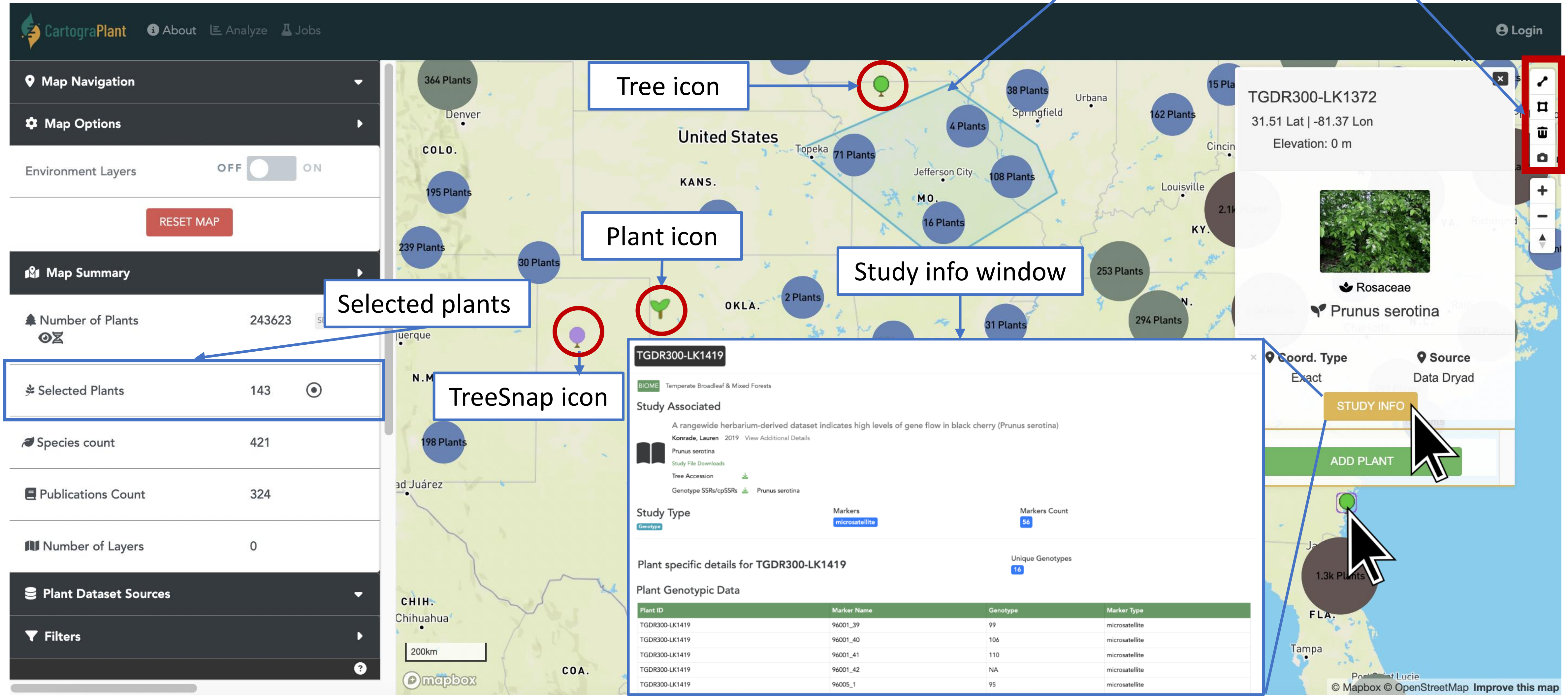
https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/

INTERFACE AND DATA VISUALIZATION

Polygon selecting plants

Polygon drawing button



Tree icon

Plant icon

Study info window

Selected plants

TreeSnap icon

Number of Plants	243623
Selected Plants	143
Species count	421
Publications Count	324
Number of Layers	0

TGDR300-LK1372

31.51 Lat | -81.37 Lon

Elevation: 0 m

Rosaceae

Prunus serotina

Coord. Type: Exact

Source: Data Dryad

STUDY INFO

ADD PLANT

TGDR300-LK1419

BIOME Temperate Broadleaf & Mixed Forests

Study Associated

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

Konrade, Lauren 2019 View Additional Details

Prunus serotina

Study File Downloads

Tree Accession

Genotype SSRs/cpSSRs Prunus serotina

Study Type: Genotype

Markers: microsatellite

Markers Count: 56

Plant specific details for TGDR300-LK1419

Unique Genotypes: 16

Plant Genotypic Data

Plant ID	Marker Name	Genotype	Marker Type
TGDR300-LK1419	96001_39	99	microsatellite
TGDR300-LK1419	96001_40	106	microsatellite
TGDR300-LK1419	96001_41	110	microsatellite
TGDR300-LK1419	96001_42	NA	microsatellite
TGDR300-LK1419	96005_1	95	microsatellite

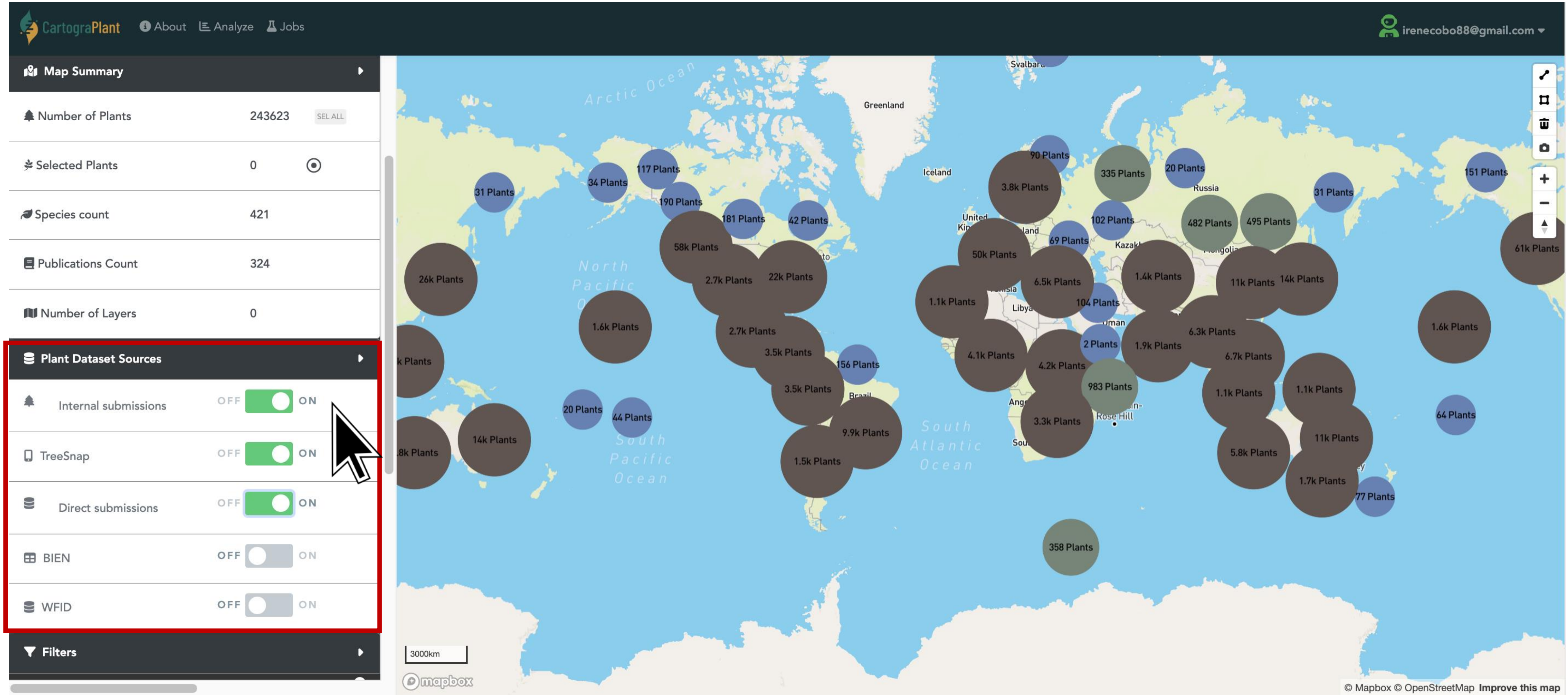
https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/

INTERFACE AND DATA VISUALIZATION

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



Map Summary

Number of Plants	243623	SEL ALL
Selected Plants	0	◉
Species count	421	
Publications Count	324	
Number of Layers	0	

Plant Dataset Sources

Internal submissions	OFF ● ON
TreeSnap	OFF ● ON
Direct submissions	OFF ● ON
BIEN	● OFF ON
WFID	● OFF ON

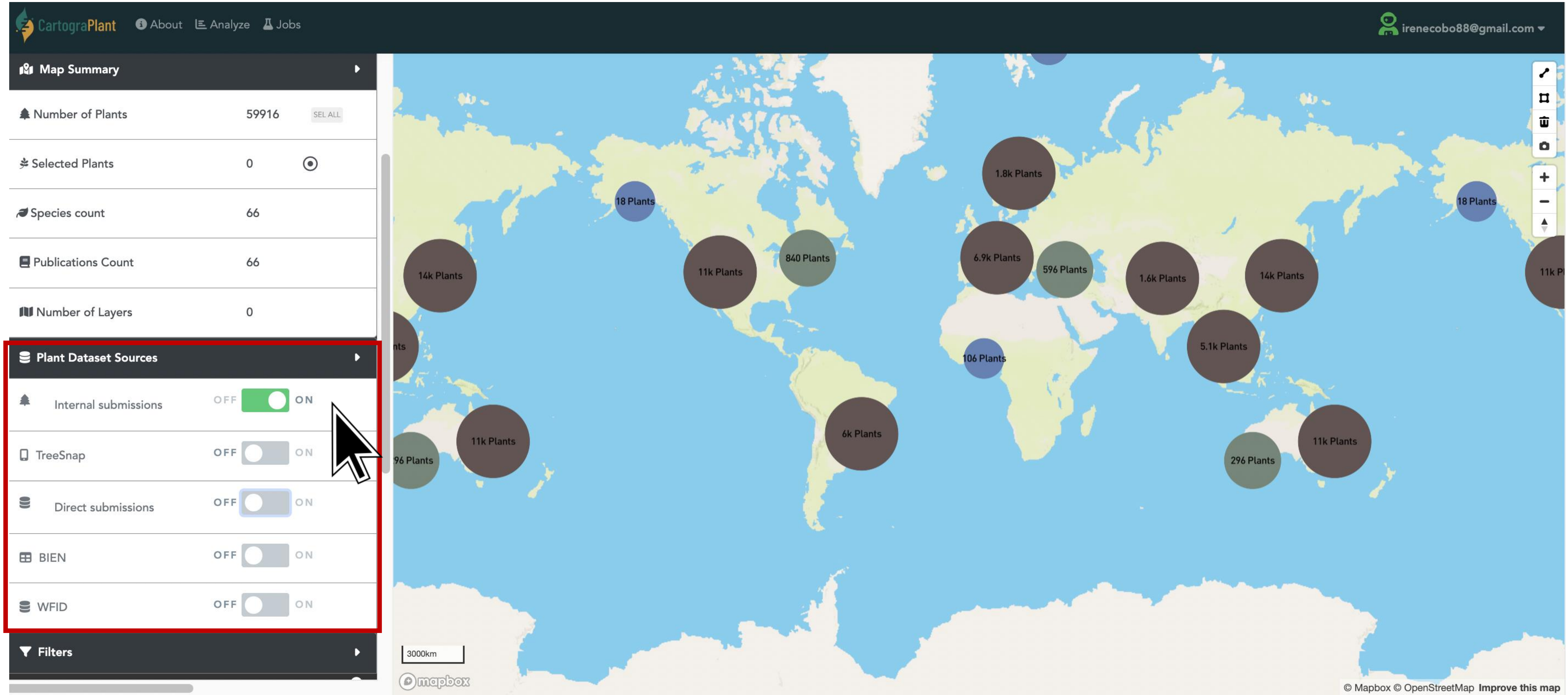
Filters

DATA SELECTION FOR ANALYSIS

INTERFACE AND DATA VISUALIZATION

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



Map Summary

Number of Plants	59916	SEL ALL
Selected Plants	0	◉
Species count	66	
Publications Count	66	
Number of Layers	0	

Plant Dataset Sources

Internal submissions	OFF ● ON
TreeSnap	● OFF ON
Direct submissions	● OFF ON
BIEN	● OFF ON
WFID	● OFF ON

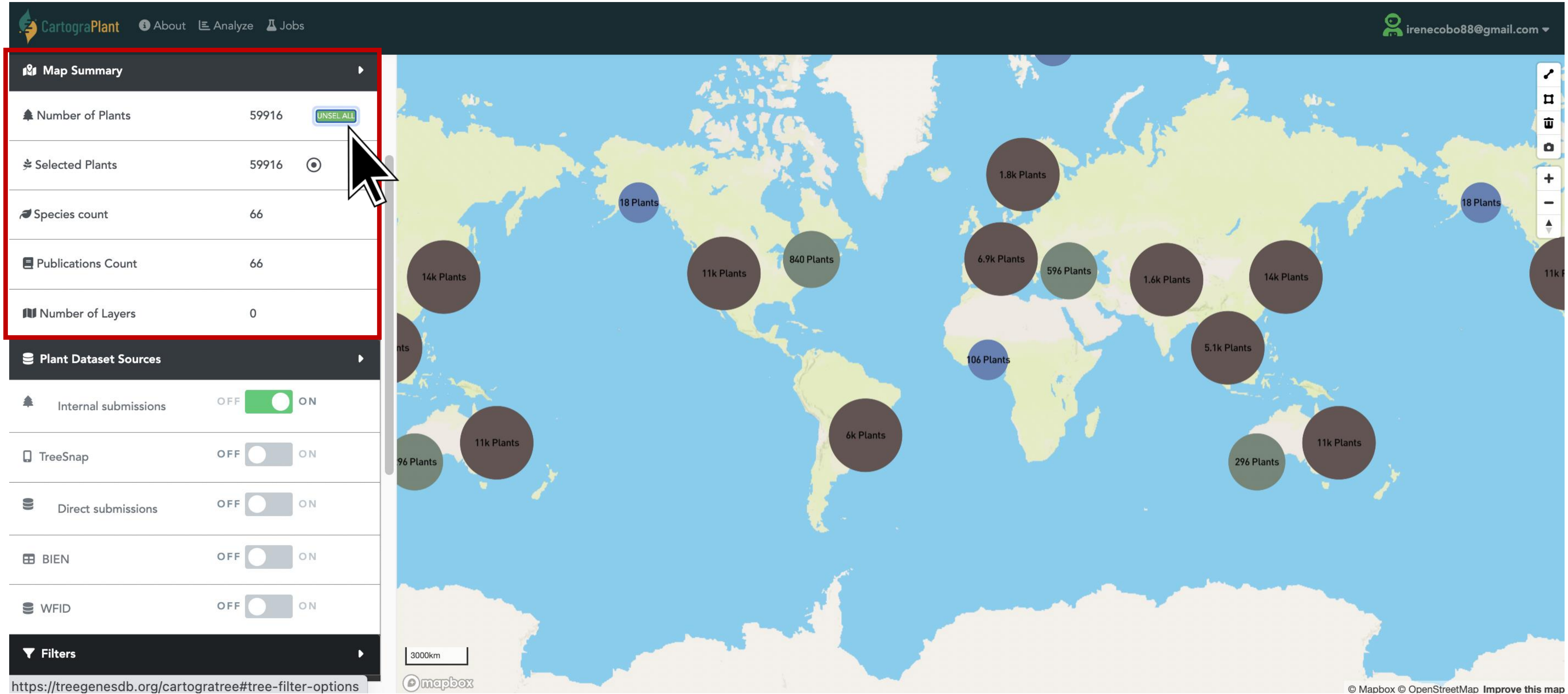
Filters

DATA SELECTION FOR ANALYSIS

INTERFACE AND DATA VISUALIZATION

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



Map Summary

Number of Plants	59916
Selected Plants	59916
Species count	66
Publications Count	66
Number of Layers	0

Plant Dataset Sources

Internal submissions	ON
TreeSnap	OFF
Direct submissions	OFF
BIEN	OFF
WFID	OFF

Filters

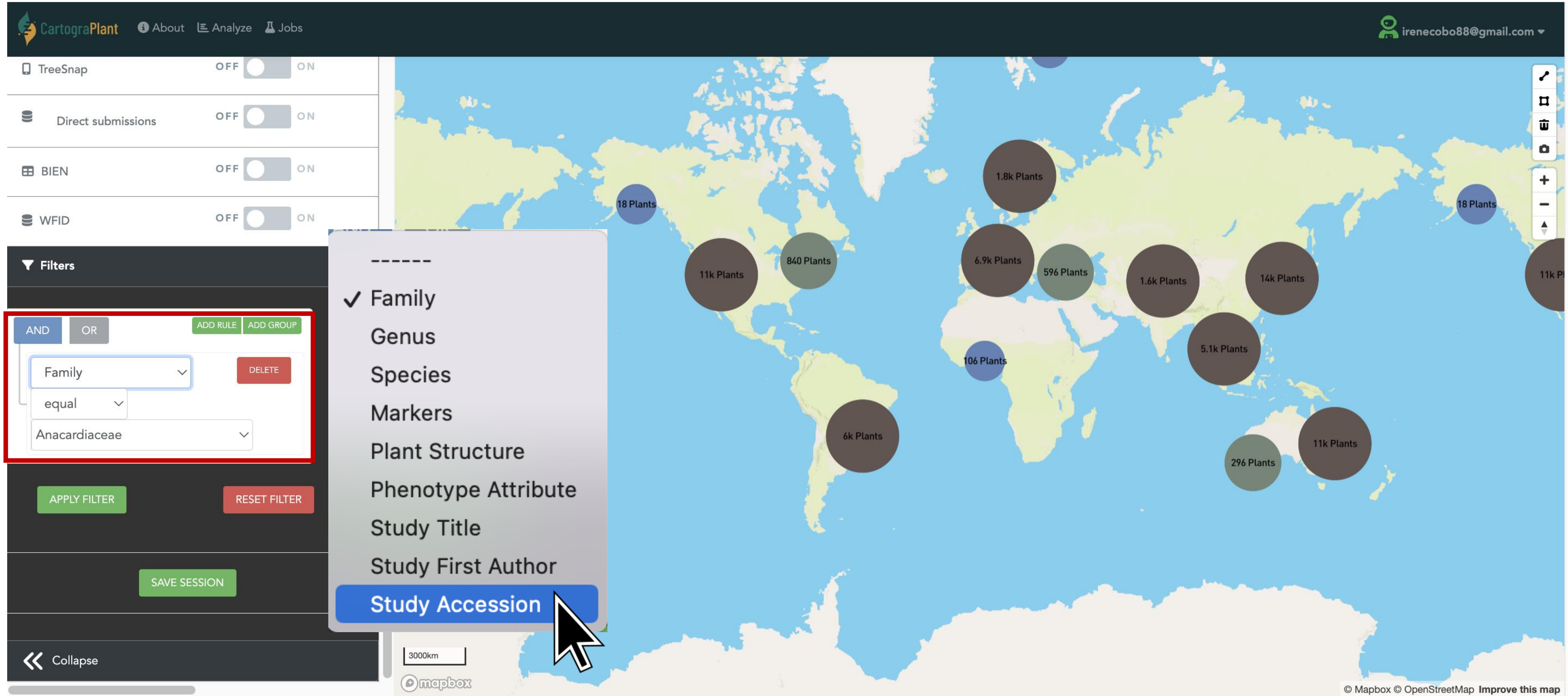
https://treegenesdb.org/cartogratree#tree-filter-options

DATA SELECTION FOR ANALYSIS

INTERFACE AND DATA VISUALIZATION

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



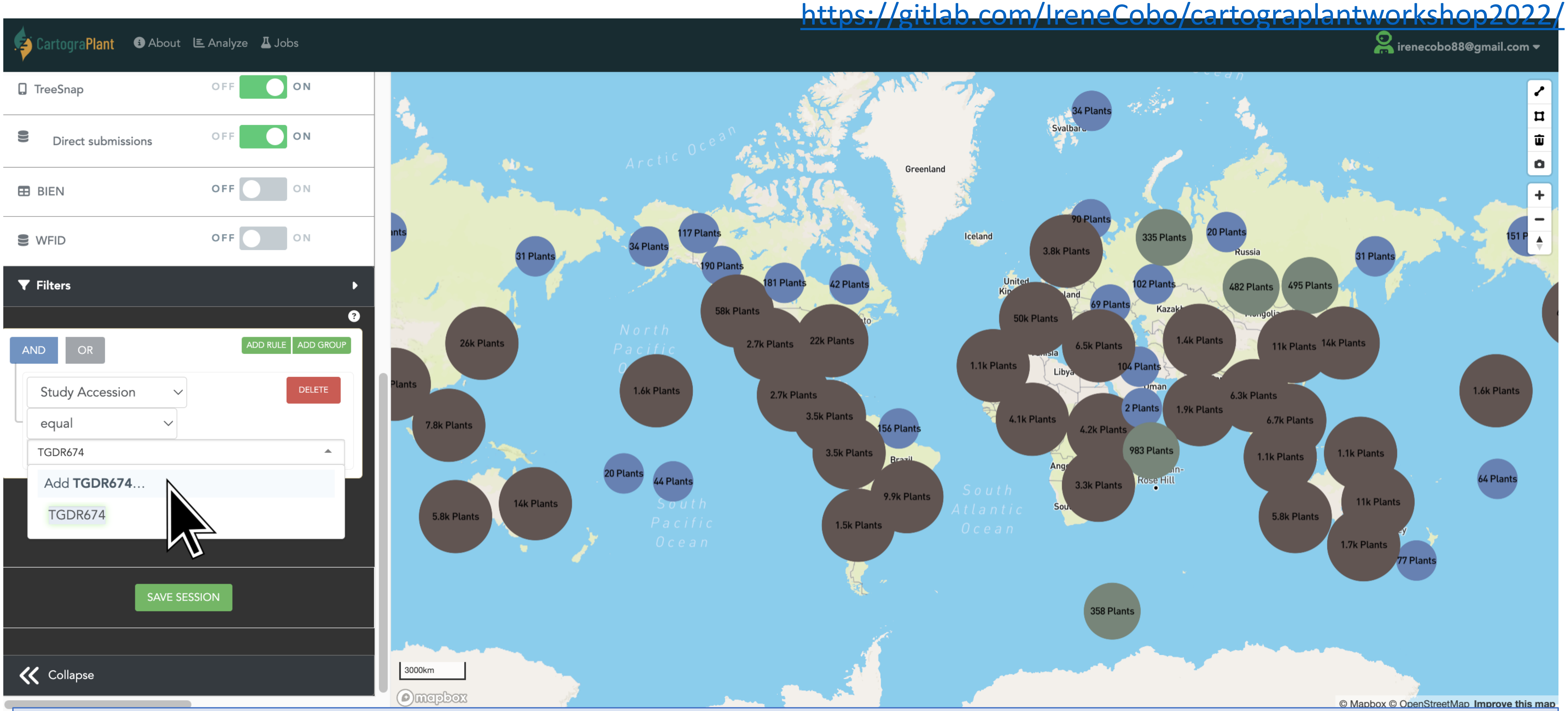
Family
equal
Anacardiaceae

DATA SELECTION FOR ANALYSIS

INTERFACE AND DATA VISUALIZATION

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



TreeSnap OFF ON (ON)

Direct submissions OFF ON (ON)

BIEN OFF ON (OFF)

WFID OFF ON (OFF)

Filters

AND

Study Accession

equal

TGDR674

DATA SELECTION FOR ANALYSIS

INTERFACE AND DATA VISUALIZATION

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/

TreeSnap OFF ON

Direct submissions OFF ON

BIEN OFF ON

WFID OFF ON

Study Accession

equal

TGDR674

1 phenotypes found in study: frost free days

DATA SELECTION FOR ANALYSIS

INTERFACE AND DATA VISUALIZATION

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/

Study Accession
equal
TGDR674

1 phenotypes found in study: frost free days

Study Accession
equal
TGDR682

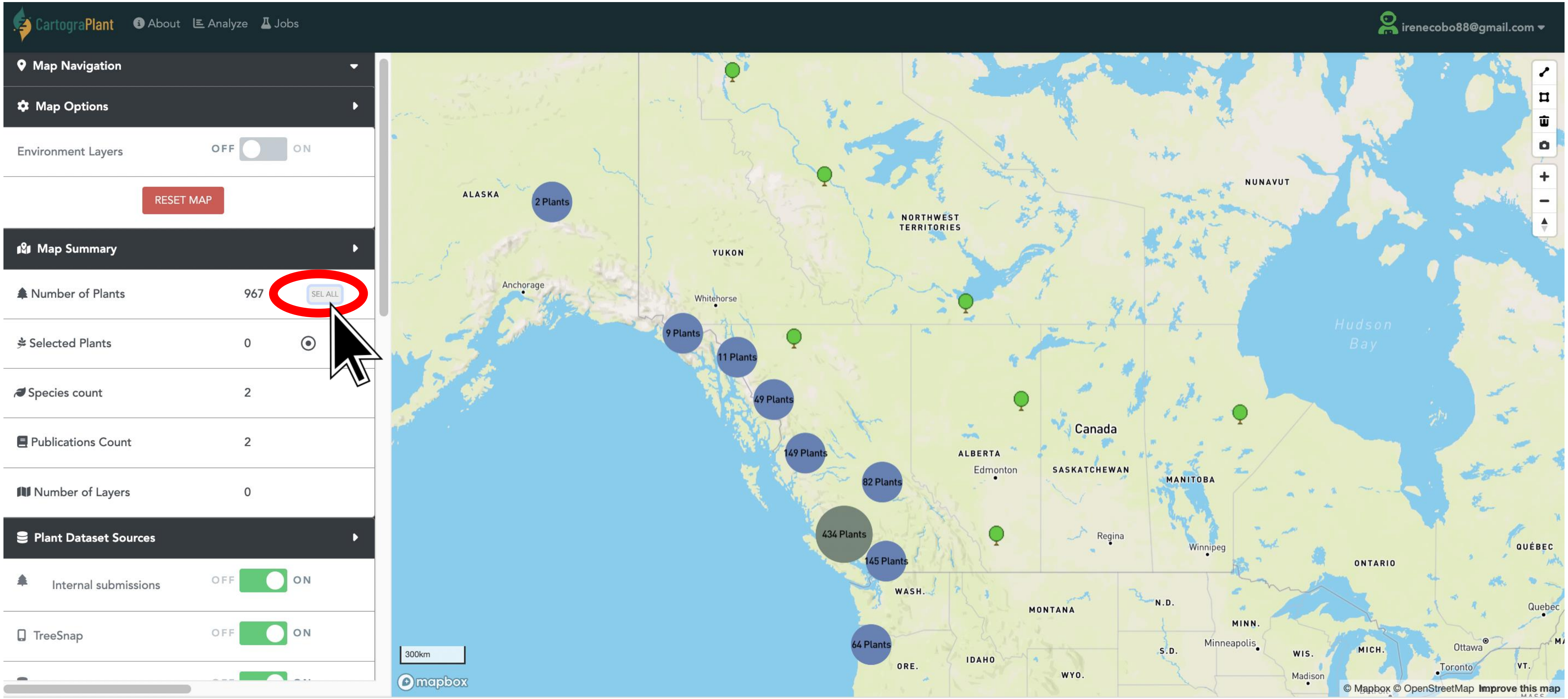
17 phenotypes found in study: bud break, bud set, canopy duration, frost free days, growth period, leaf carbon to nitrogen ratio, leaf chlorophyll content, leaf nitrogen 15 content, leaf shape, mean annual precipitation, mean annual temperature, plant height, ratio of average warmest month temperature to average summer precipitation, ratio of mean annual temperature to mean annual precipitation, volume, whole plant mass, wood carbon 13 content

DATA SELECTION FOR ANALYSIS

INTERFACE AND DATA VISUALIZATION

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



Environment Layers OFF ON

Number of Plants 967

Selected Plants 0

Species count 2

Publications Count 2

Number of Layers 0

Internal submissions OFF ON

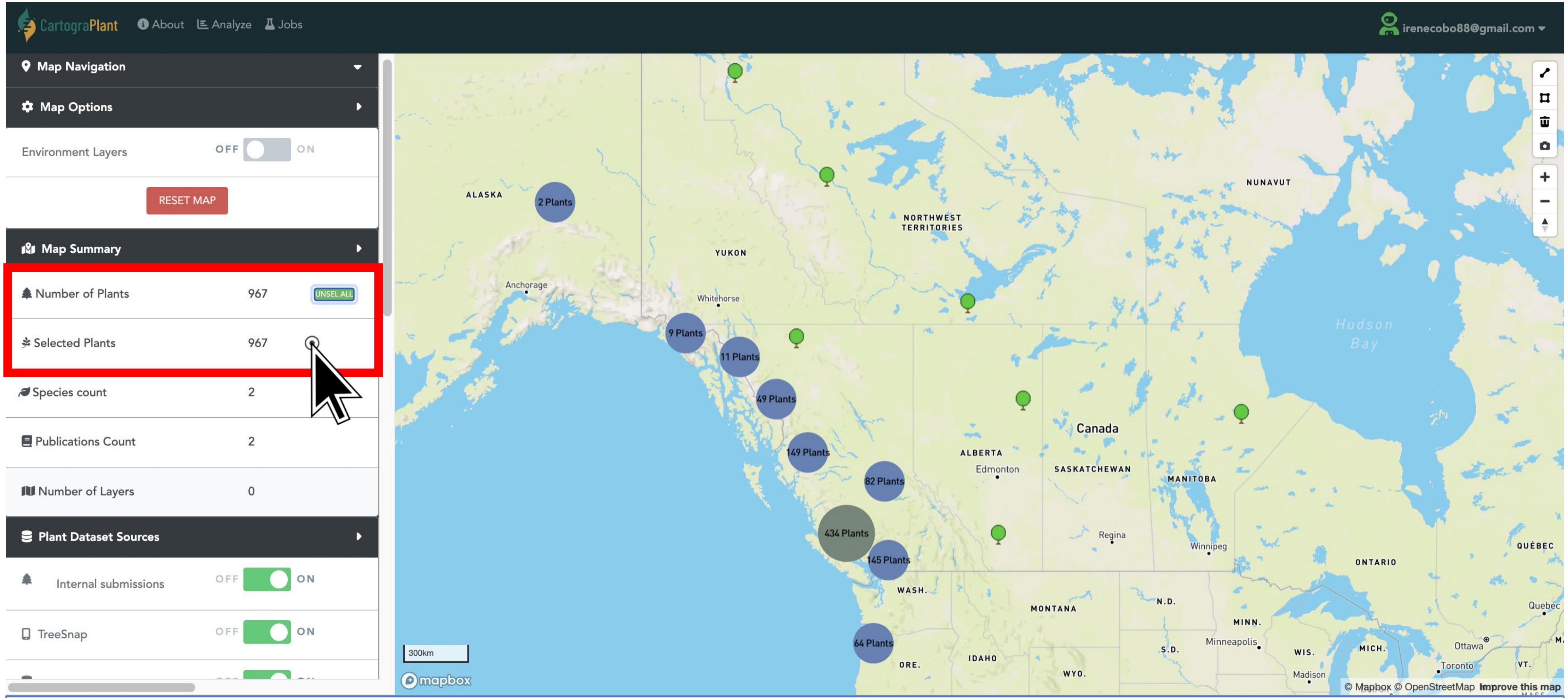
TreeSnap OFF ON

DATA SELECTION FOR ANALYSIS

INTERFACE AND DATA VISUALIZATION

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



Environment Layers OFF ○ ON

Number of Plants	967
Selected Plants	967
Species count	2
Publications Count	2
Number of Layers	0

Internal submissions OFF ● ON

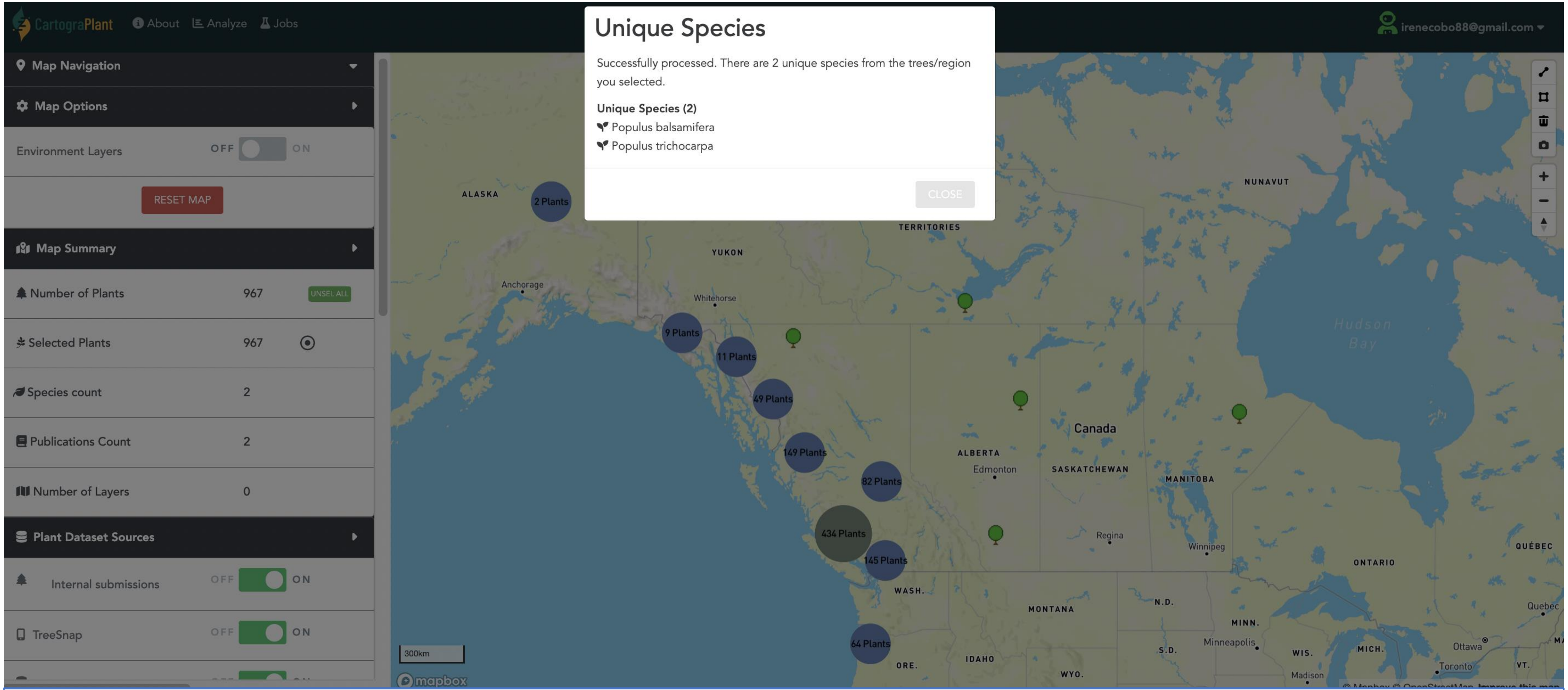
TreeSnap OFF ● ON

DATA SELECTION FOR ANALYSIS

INTERFACE AND DATA VISUALIZATION

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



Unique Species

Successfully processed. There are 2 unique species from the trees/region you selected.

Unique Species (2)

- Populus balsamifera
- Populus trichocarpa

Environment Layers OFF ○ ON

Number of Plants	967
Selected Plants	967
Species count	2
Publications Count	2
Number of Layers	0

Internal submissions OFF ● ON

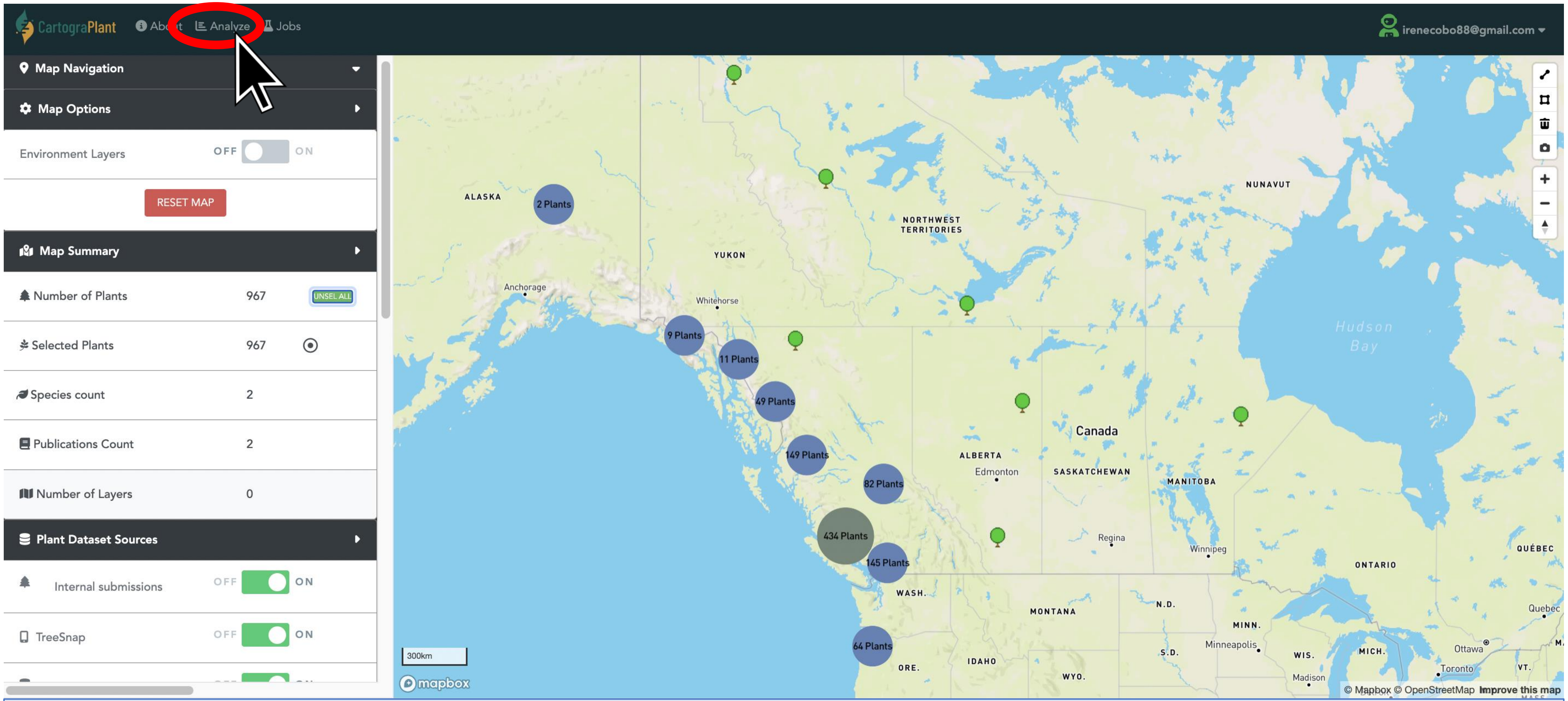
TreeSnap OFF ● ON

DATA SELECTION FOR ANALYSIS

INTERFACE AND DATA VISUALIZATION

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



Environment Layers OFF ON

Map Summary

Number of Plants	967
Selected Plants	967
Species count	2
Publications Count	2
Number of Layers	0

Plant Dataset Sources

Internal submissions OFF ON

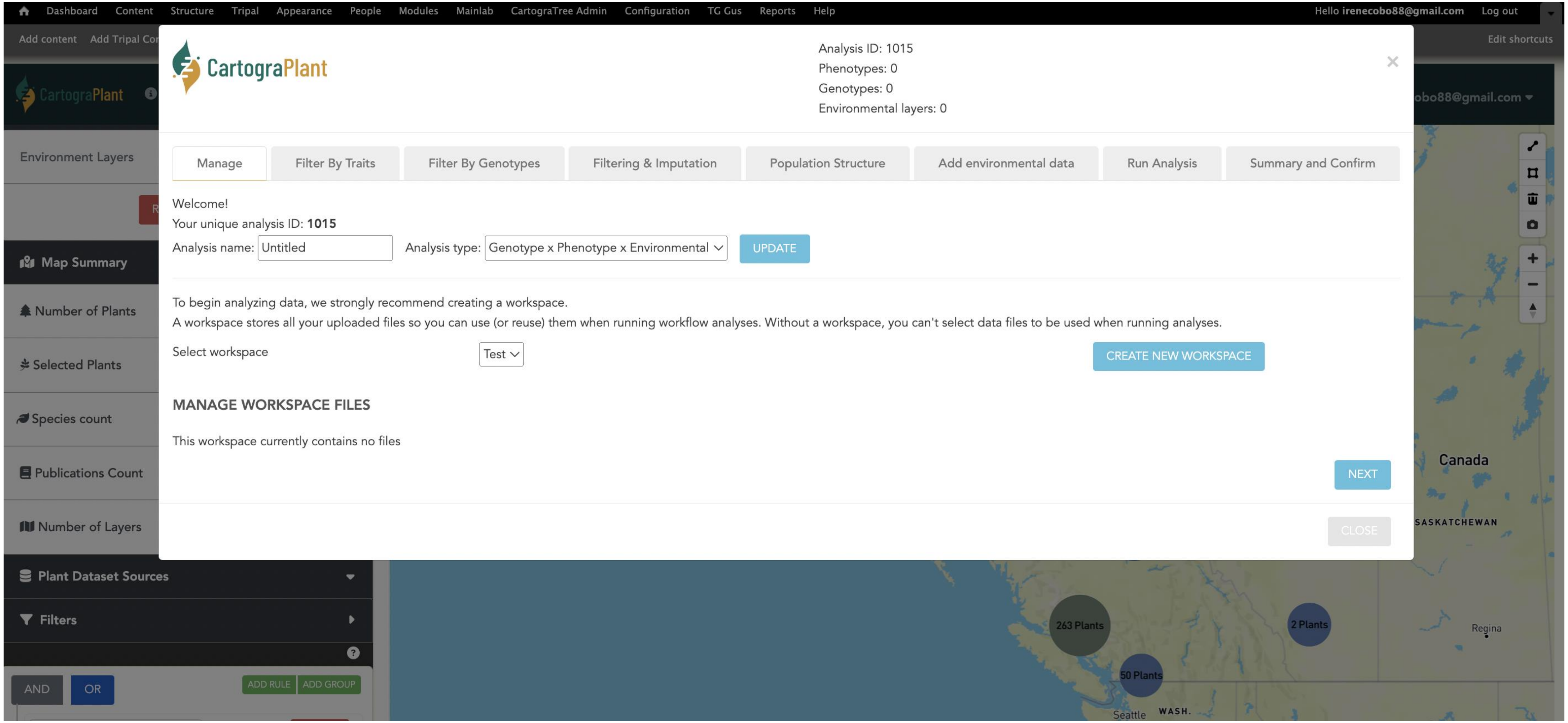
TreeSnap OFF ON

DATA SELECTION FOR ANALYSIS

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

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

Analysis name: Untitled
Analysis type: Genotype x Phenotype x Environmental

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

MANAGE WORKSPACE FILES

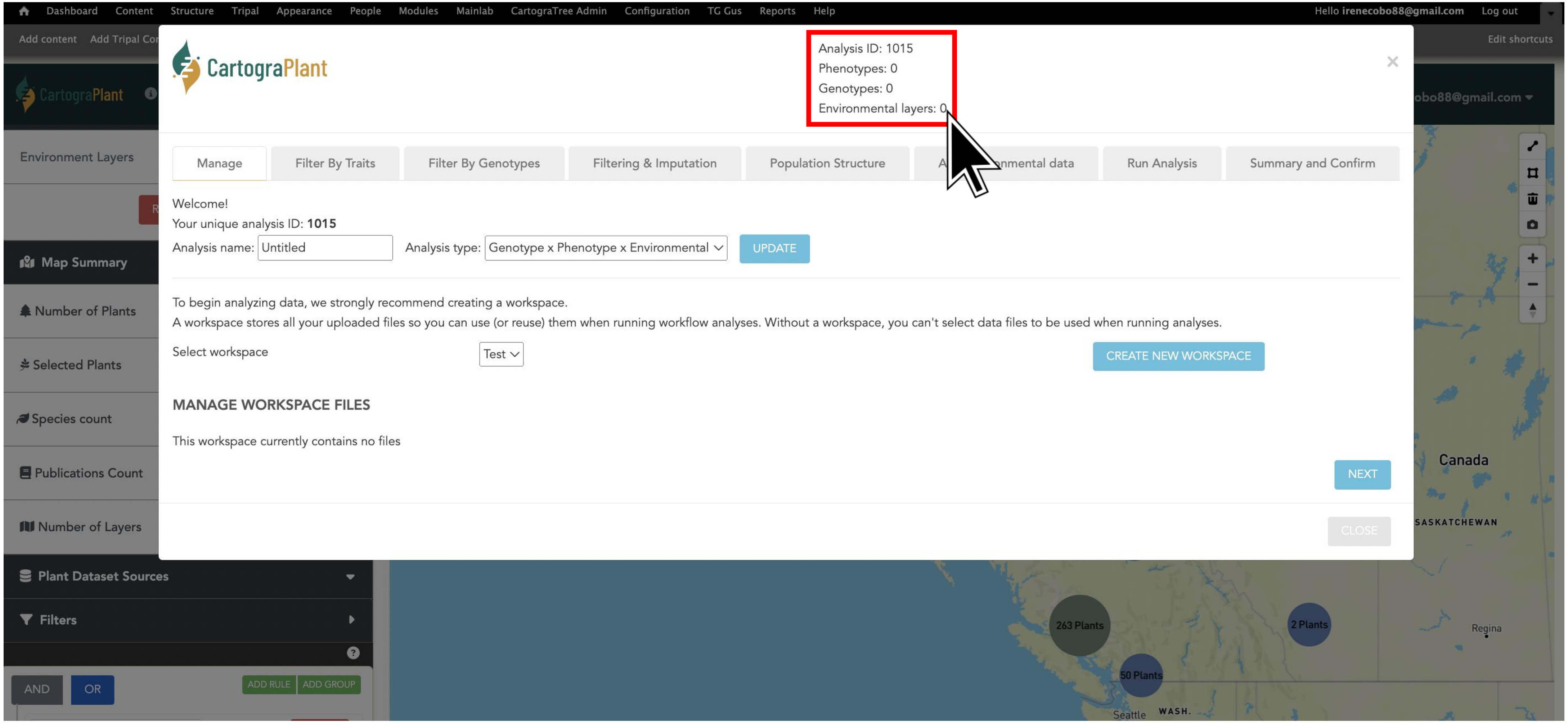
This workspace currently contains no files

DATA ANALYSIS: WORKSPACE CREATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

Welcome!
Your unique analysis ID: **1015**
Analysis name: Untitled
Analysis type: Genotype x Phenotype x Environmental

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

MANAGE WORKSPACE FILES

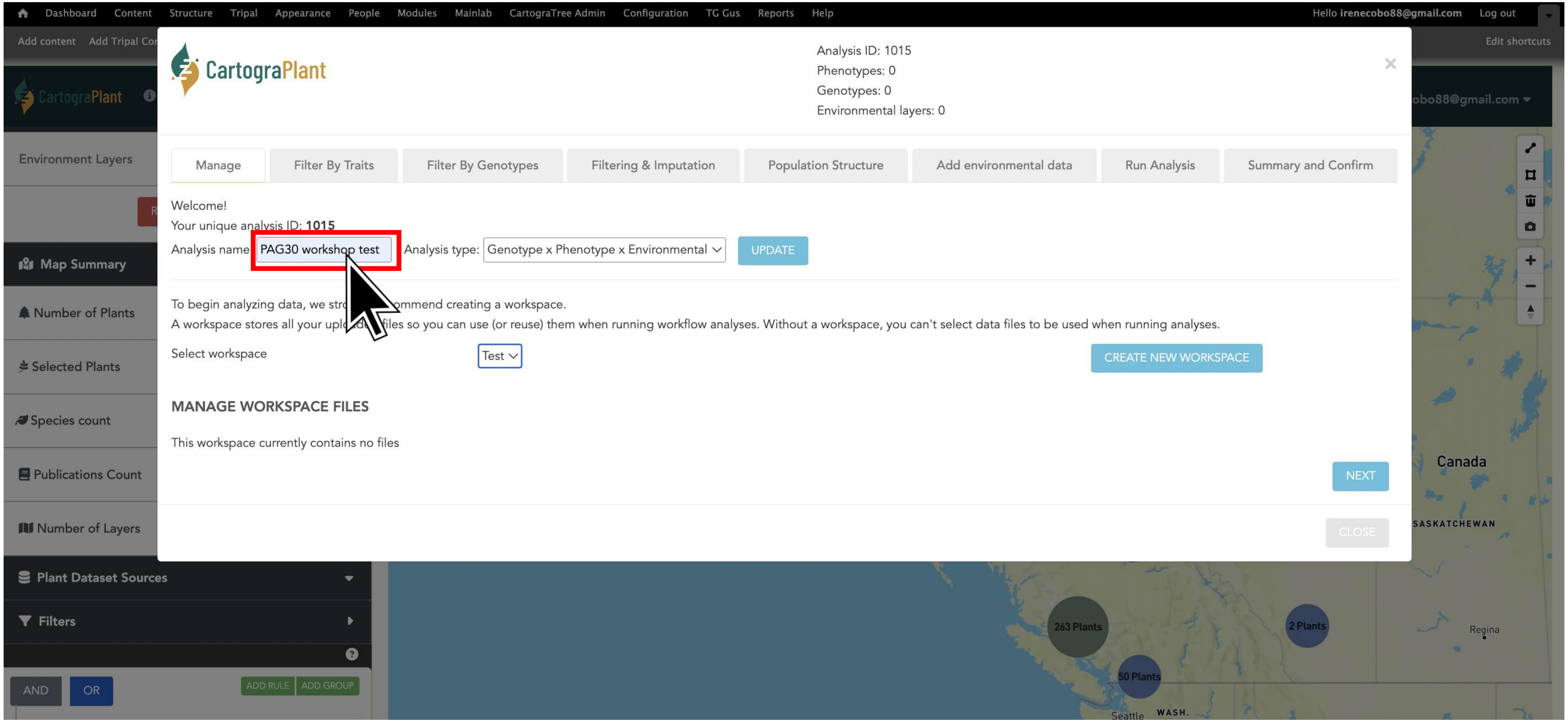
This workspace currently contains no files

DATA ANALYSIS: WORKSPACE CREATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

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

Analysis name: PAG30 workshop test
Analysis type: Genotype x Phenotype x Environmental

To begin analyzing data, we stro mmend creating a workspace.
A workspace stores all your upl 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

MANAGE WORKSPACE FILES

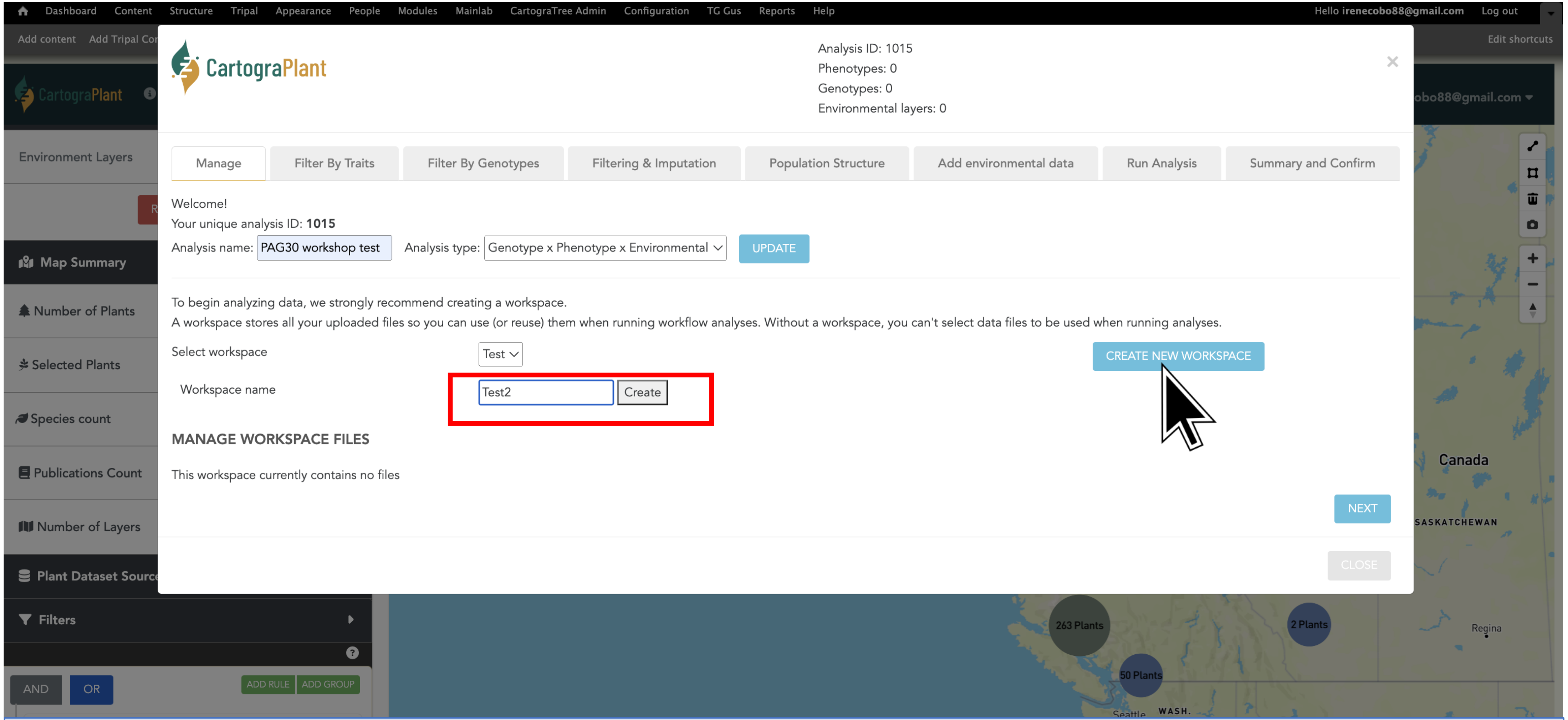
This workspace currently contains no files

DATA ANALYSIS: WORKSPACE CREATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

Welcome!
Your unique analysis ID: **1015**
Analysis name: PAG30 workshop test
Analysis type: Genotype x Phenotype x Environmental

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

Workspace name: Test2

MANAGE WORKSPACE FILES

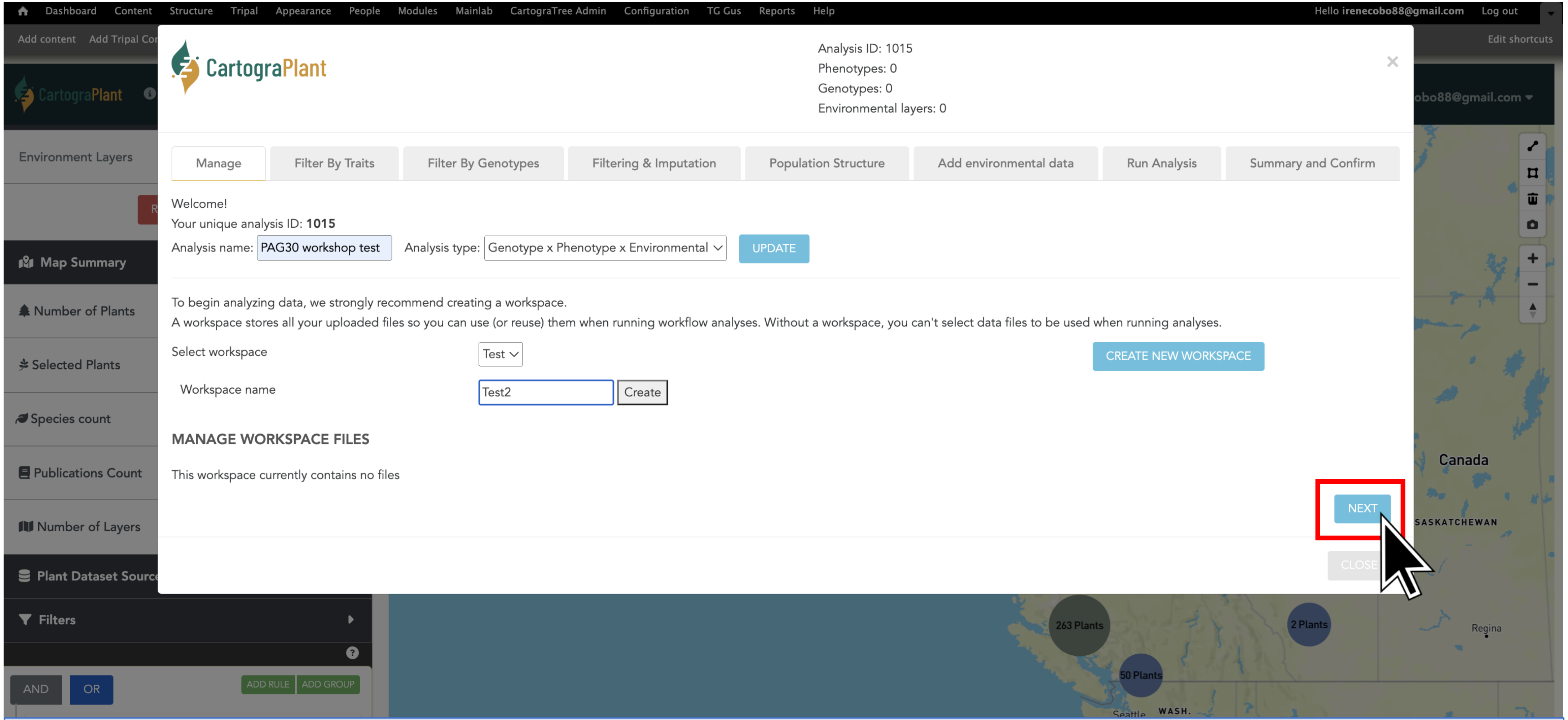
This workspace currently contains no files

DATA ANALYSIS: WORKSPACE CREATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

Welcome!
Your unique analysis ID: **1015**
Analysis name: PAG30 workshop test
Analysis type: Genotype x Phenotype x Environmental

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
Workspace name: Test2

MANAGE WORKSPACE FILES

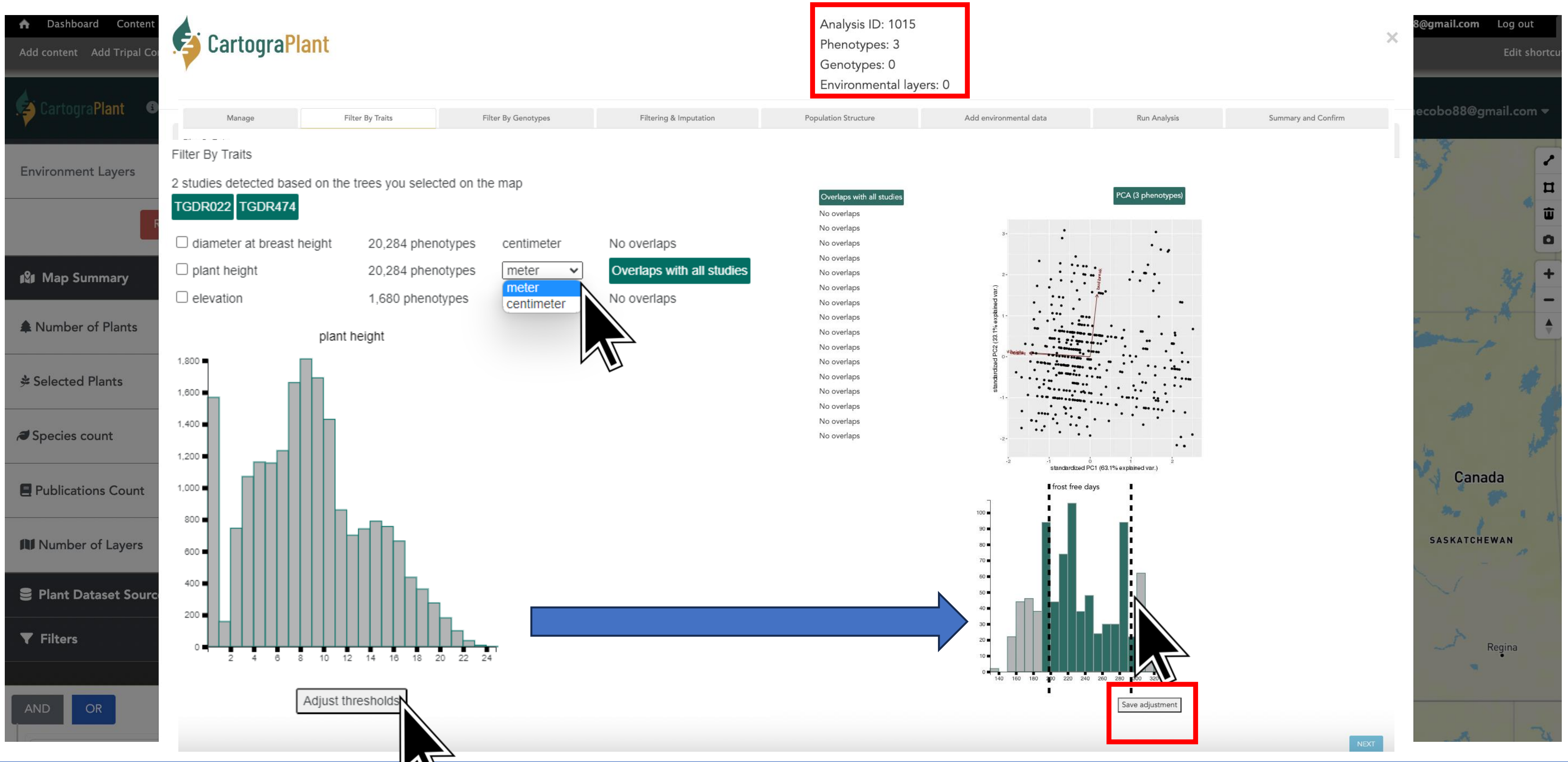
This workspace currently contains no files

DATA ANALYSIS: WORKSPACE CREATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

Filter By Traits

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

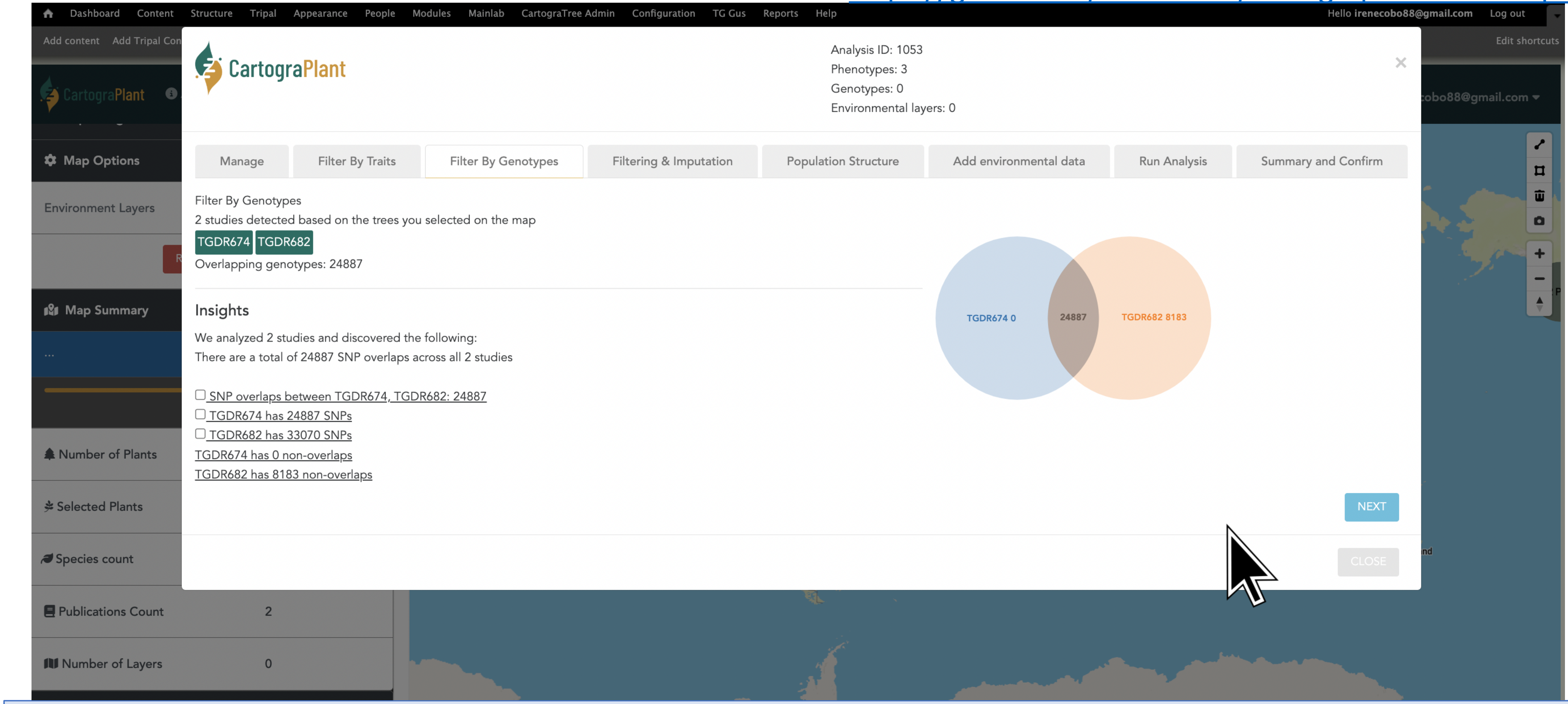
☐ diameter at breast height	20,284 phenotypes	centimeter	No overlaps
☐ plant height	20,284 phenotypes	meter meter centimeter	Overlaps with all studies
☐ elevation	1,680 phenotypes		No overlaps

DATA ANALYSIS: PHENOTYPIC TRAITS SELECTION/VISUALIZATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

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

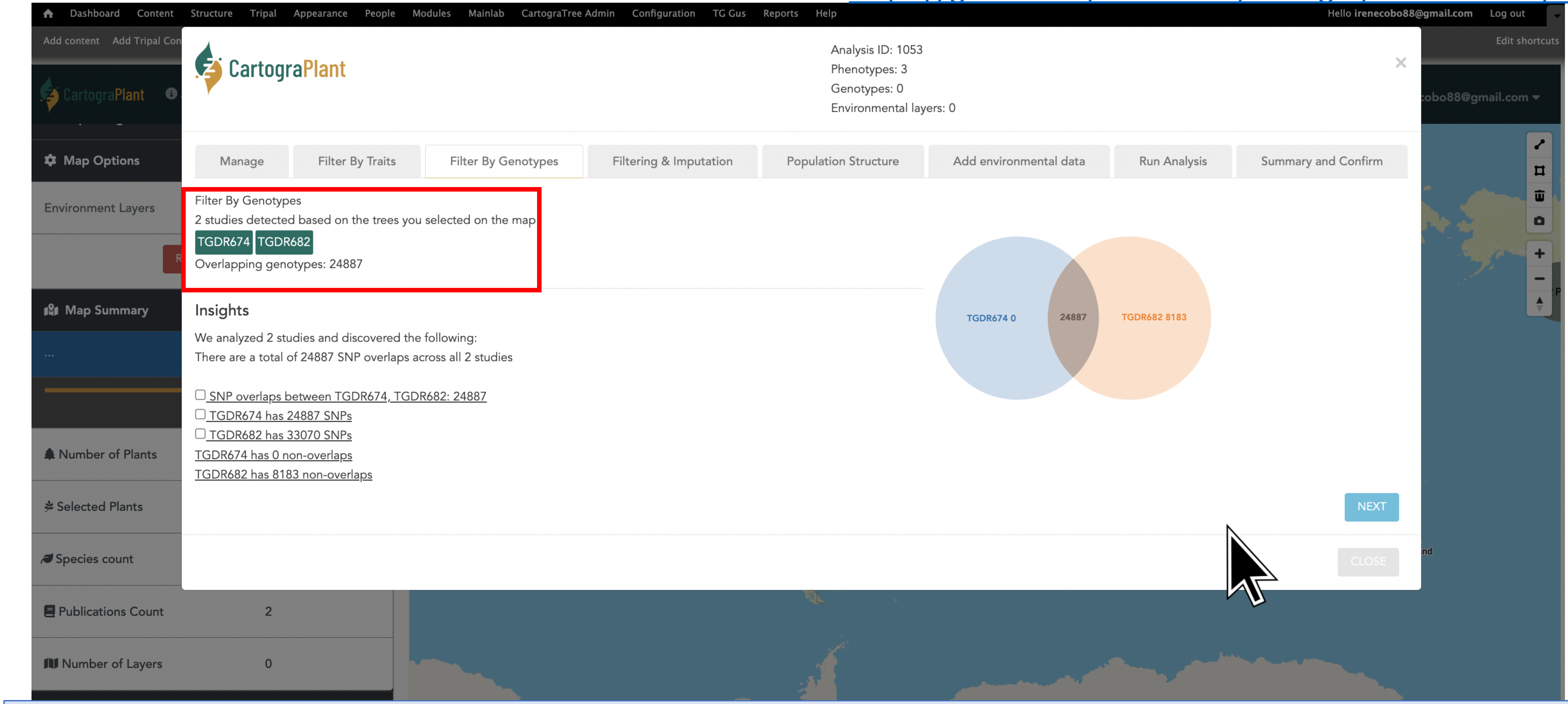
☐ SNP overlaps between TGDR674, TGDR682: 24887
☐ TGDR674 has 24887 SNPs
☐ TGDR682 has 33070 SNPs
TGDR674 has 0 non-overlaps
TGDR682 has 8183 non-overlaps

DATA ANALYSIS: GENOTYPIC DATA SELECTION/VISUALIZATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

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

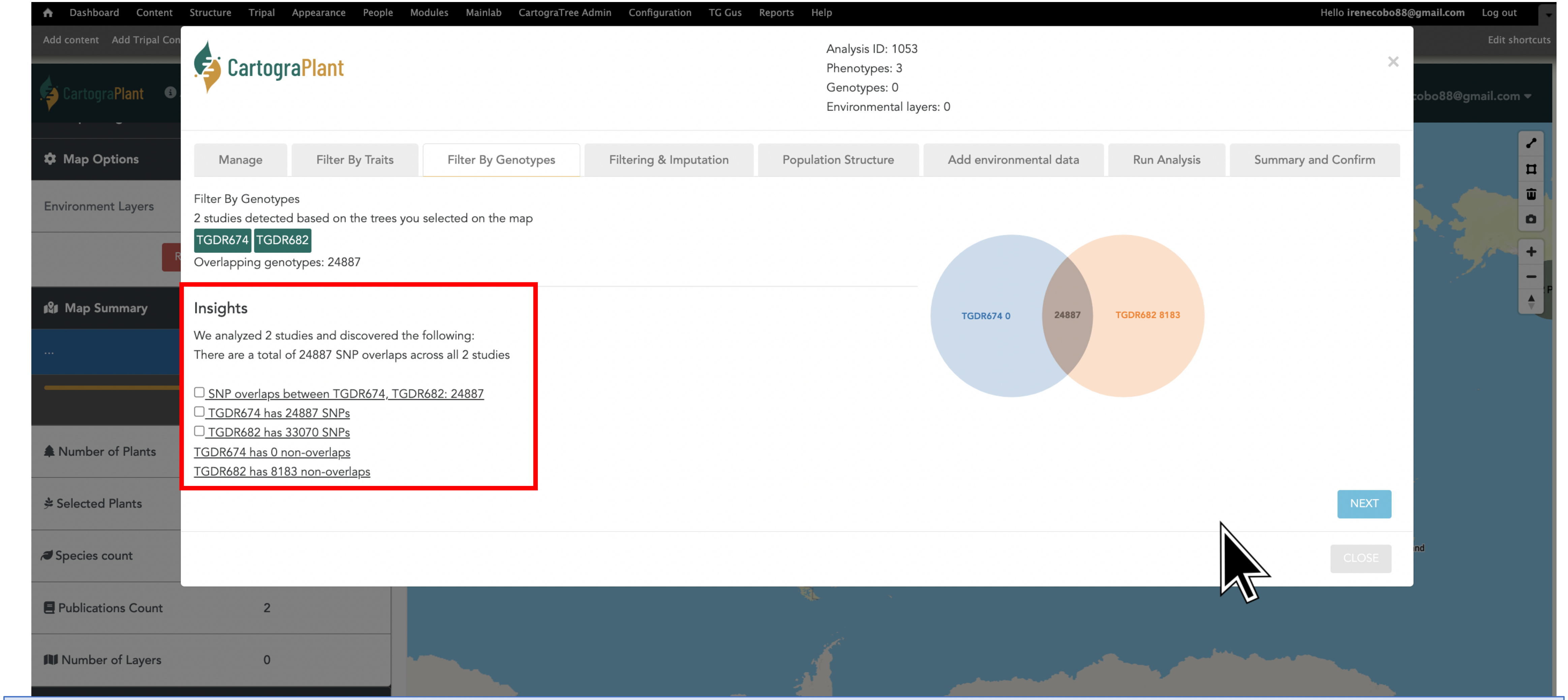
☐ SNP overlaps between TGDR674, TGDR682: 24887
☐ TGDR674 has 24887 SNPs
☐ TGDR682 has 33070 SNPs
TGDR674 has 0 non-overlaps
TGDR682 has 8183 non-overlaps

Publications Count 2
Number of Layers 0

DATA ANALYSIS: GENOTYPIC DATA SELECTION/VISUALIZATION

DATA ANALYSIS

https://cartograplant.org/



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

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

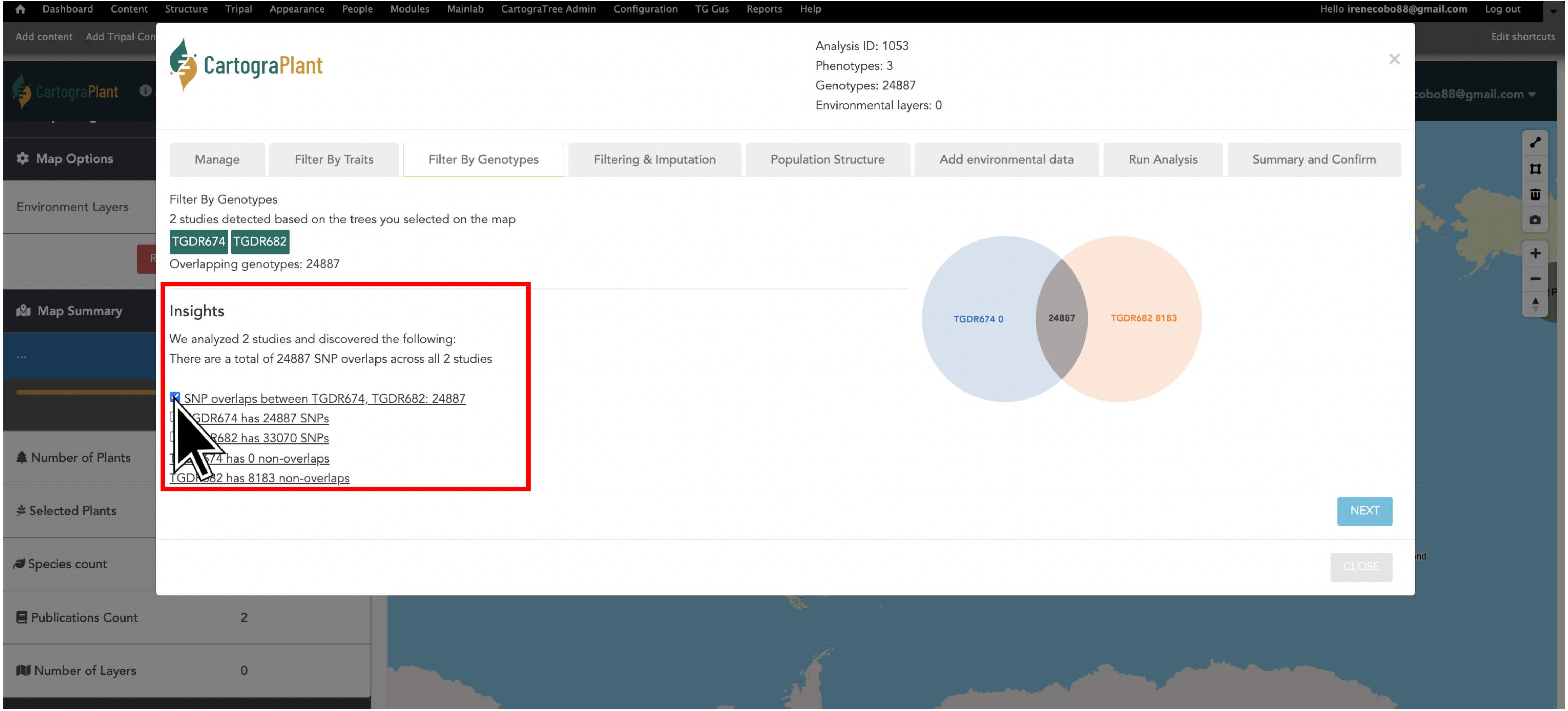
☐ SNP overlaps between TGDR674, TGDR682: 24887
☐ TGDR674 has 24887 SNPs
☐ TGDR682 has 33070 SNPs
TGDR674 has 0 non-overlaps
TGDR682 has 8183 non-overlaps

DATA ANALYSIS: GENOTYPIC DATA SELECTION/VISUALIZATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

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

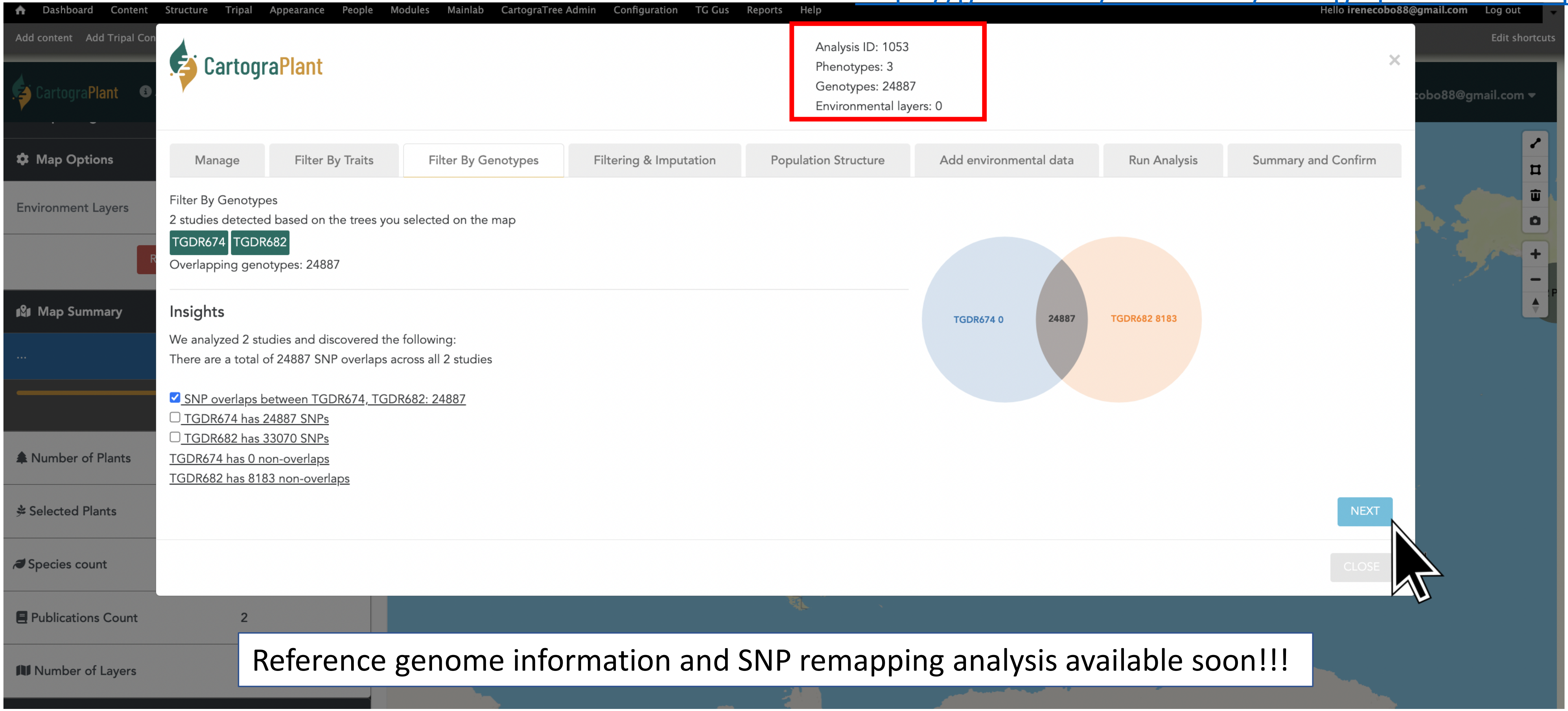
☑ SNP overlaps between TGDR674, TGDR682: 24887
☐ GDR674 has 24887 SNPs
☐ R682 has 33070 SNPs
T 674 has 0 non-overlaps
TGD 82 has 8183 non-overlaps

DATA ANALYSIS: GENOTYPIC DATA SELECTION/VISUALIZATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

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

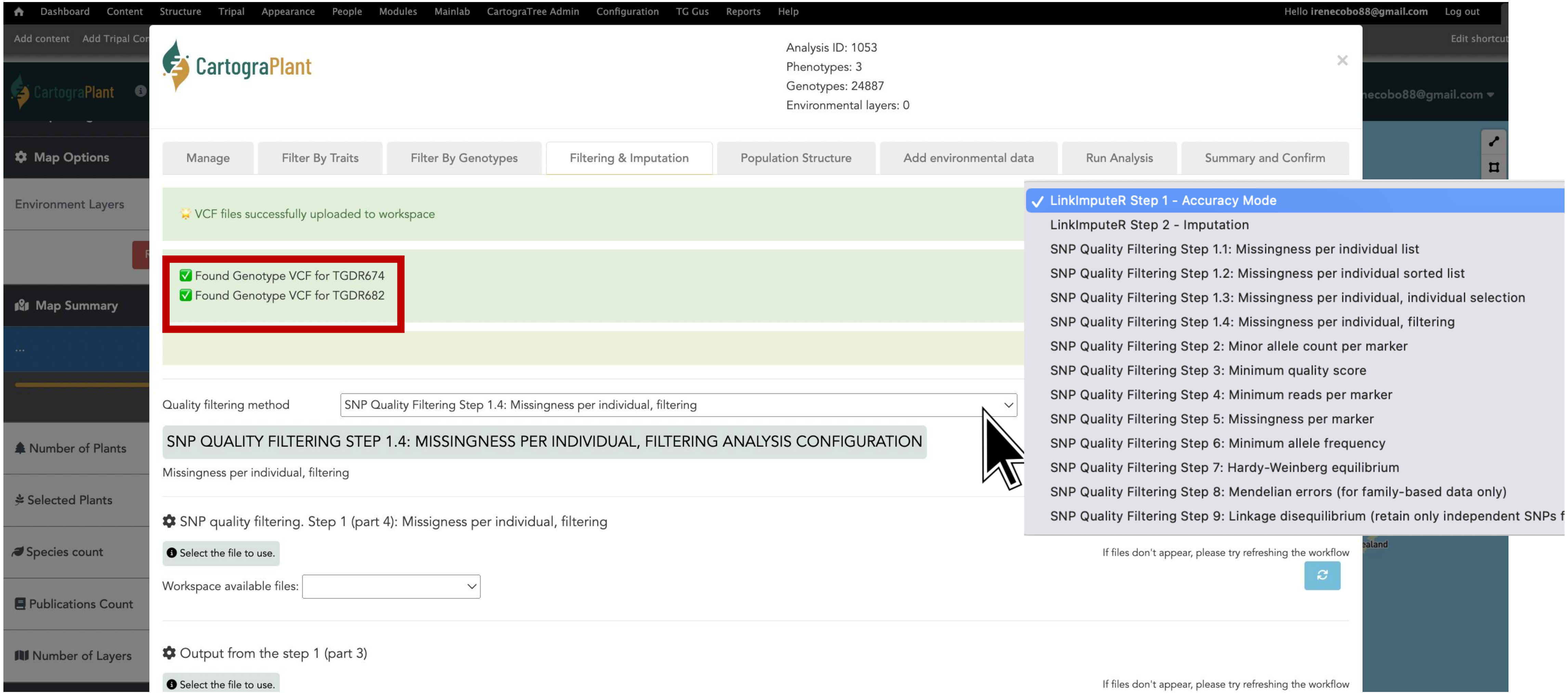
Reference genome information and SNP remapping analysis available soon!!!

DATA ANALYSIS: GENOTYPIC DATA SELECTION/VISUALIZATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

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): Missigness per individual, filtering

If files don't appear, please try refreshing the workflow

Workspace available files:

Output from the step 1 (part 3)

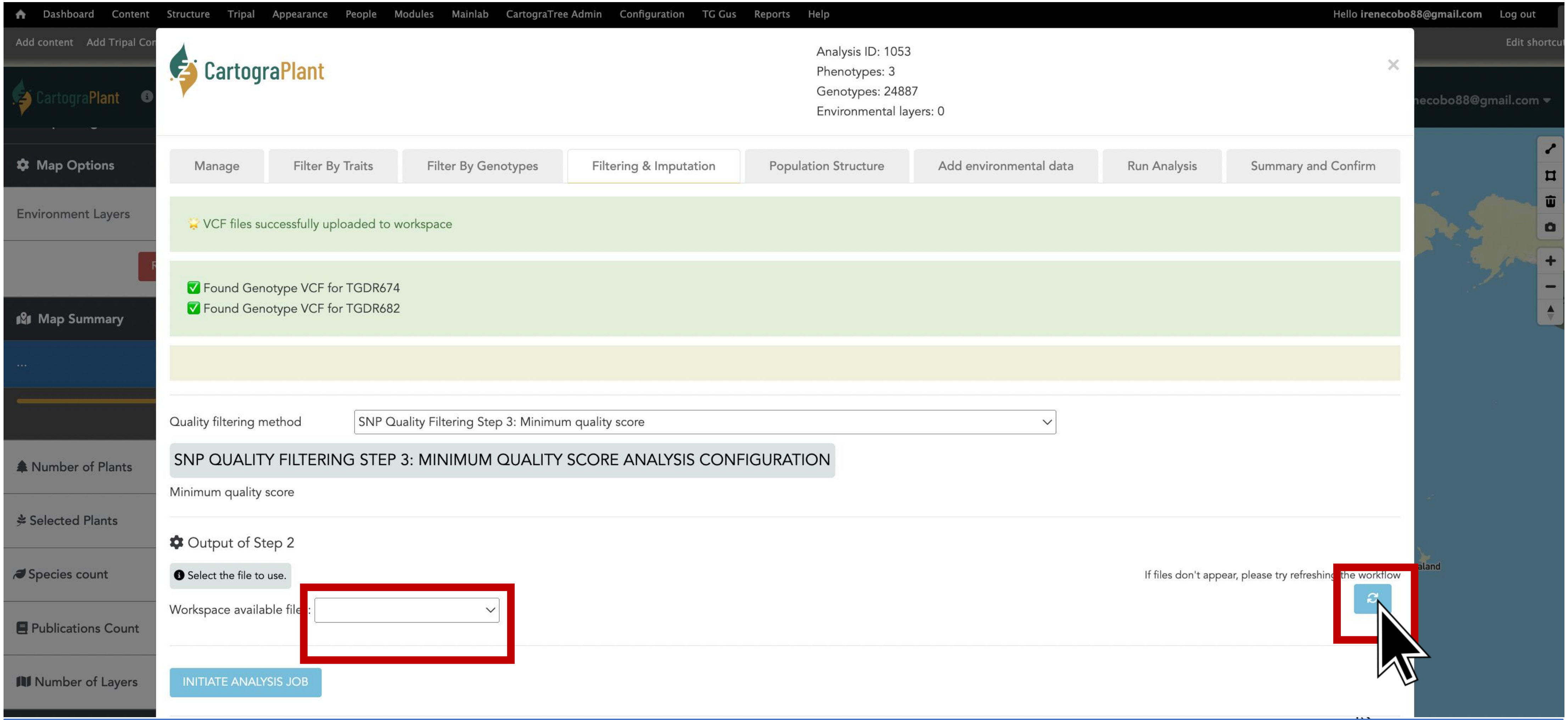
If files don't appear, please try refreshing the workflow

DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

VCF files successfully uploaded to workspace

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

Quality filtering method: SNP Quality Filtering Step 3: Minimum quality score

SNP QUALITY FILTERING STEP 3: MINIMUM QUALITY SCORE ANALYSIS CONFIGURATION

Minimum quality score

Output of Step 2

Select the file to use.

Workspace available file:

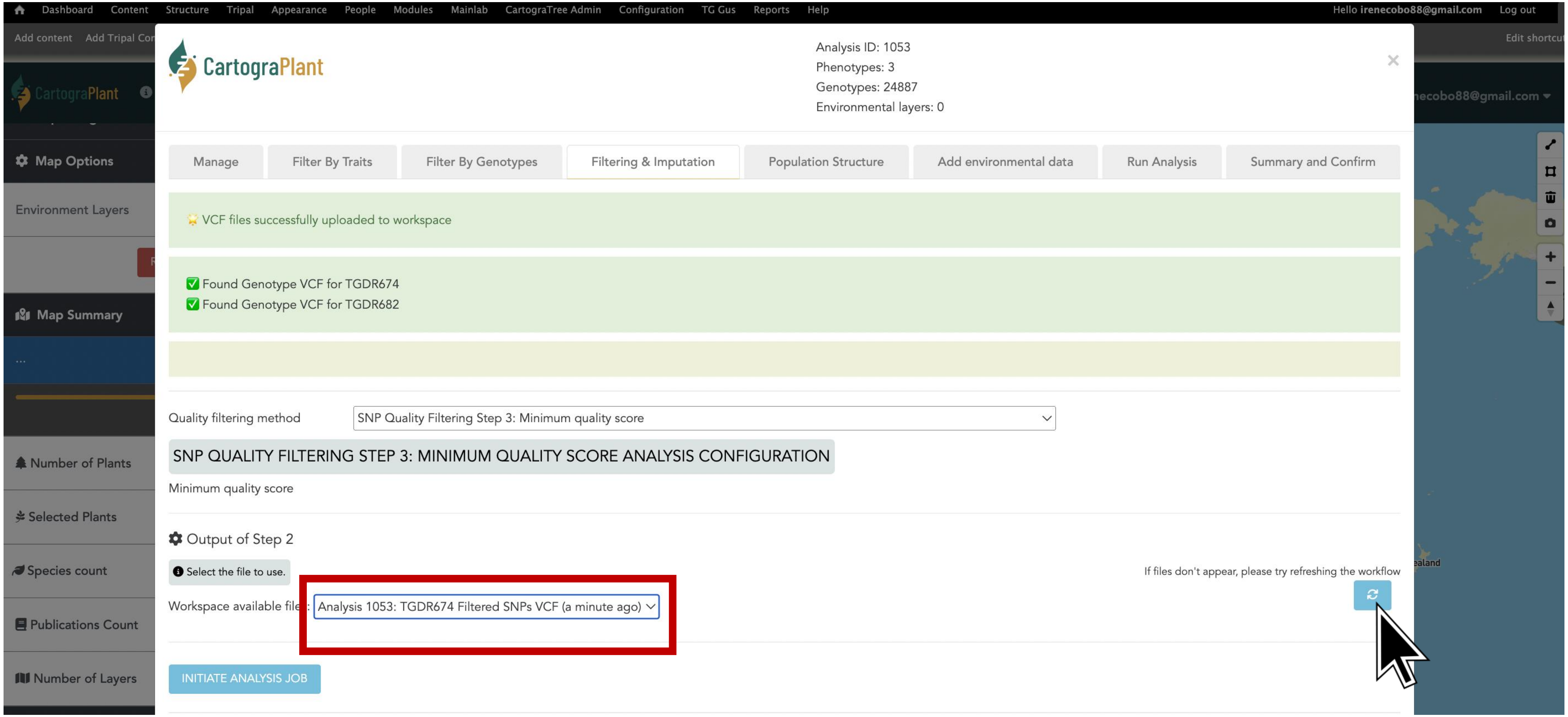
If files don't appear, please try refreshing the workflow

DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

VCF files successfully uploaded to workspace

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

Quality filtering method: SNP Quality Filtering Step 3: Minimum quality score

SNP QUALITY FILTERING STEP 3: MINIMUM QUALITY SCORE ANALYSIS CONFIGURATION

Minimum quality score

Output of Step 2

Select the file to use.

If files don't appear, please try refreshing the workflow

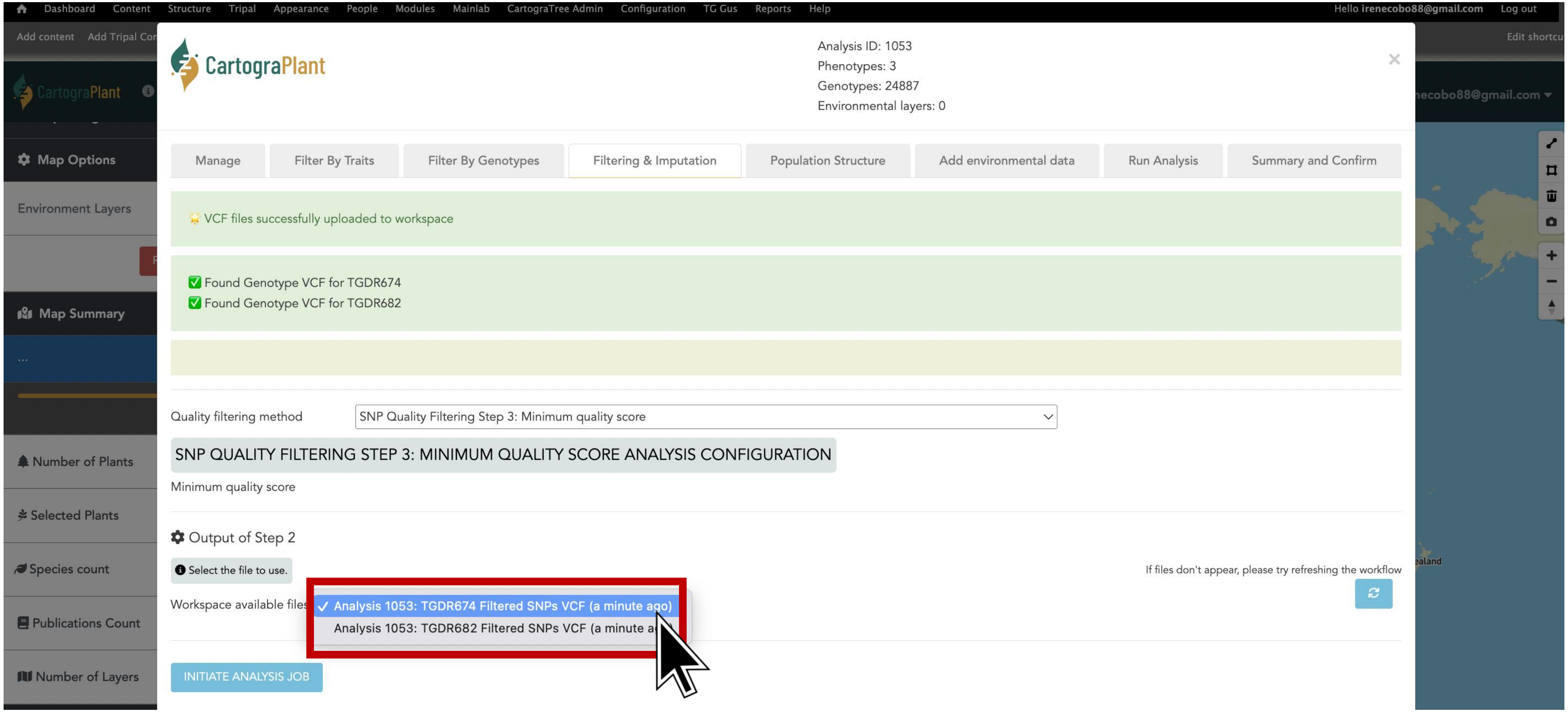
Workspace available file: Analysis 1053: TGDR674 Filtered SNPs VCF (a minute ago)

DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

VCF files successfully uploaded to workspace

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

Quality filtering method: SNP Quality Filtering Step 3: Minimum quality score

SNP QUALITY FILTERING STEP 3: MINIMUM QUALITY SCORE ANALYSIS CONFIGURATION

Minimum quality score

Output of Step 2

Select the file to use.

If files don't appear, please try refreshing the workflow

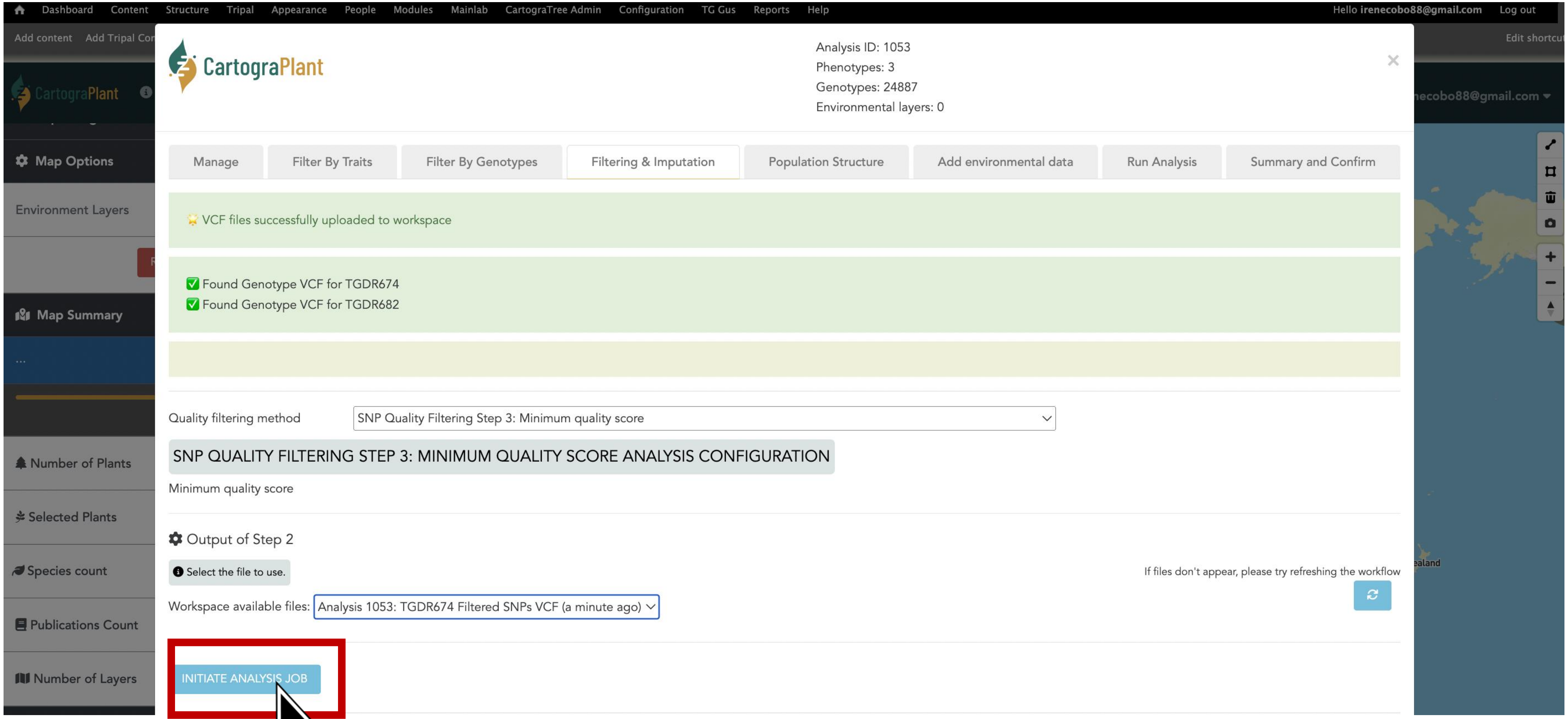
Workspace available files:
✓ Analysis 1053: TGDR674 Filtered SNPs VCF (a minute ago)
Analysis 1053: TGDR682 Filtered SNPs VCF (a minute a

DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

VCF files successfully uploaded to workspace

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

Quality filtering method: SNP Quality Filtering Step 3: Minimum quality score

SNP QUALITY FILTERING STEP 3: MINIMUM QUALITY SCORE ANALYSIS CONFIGURATION

Minimum quality score

Output of Step 2

Select the file to use.

If files don't appear, please try refreshing the workflow

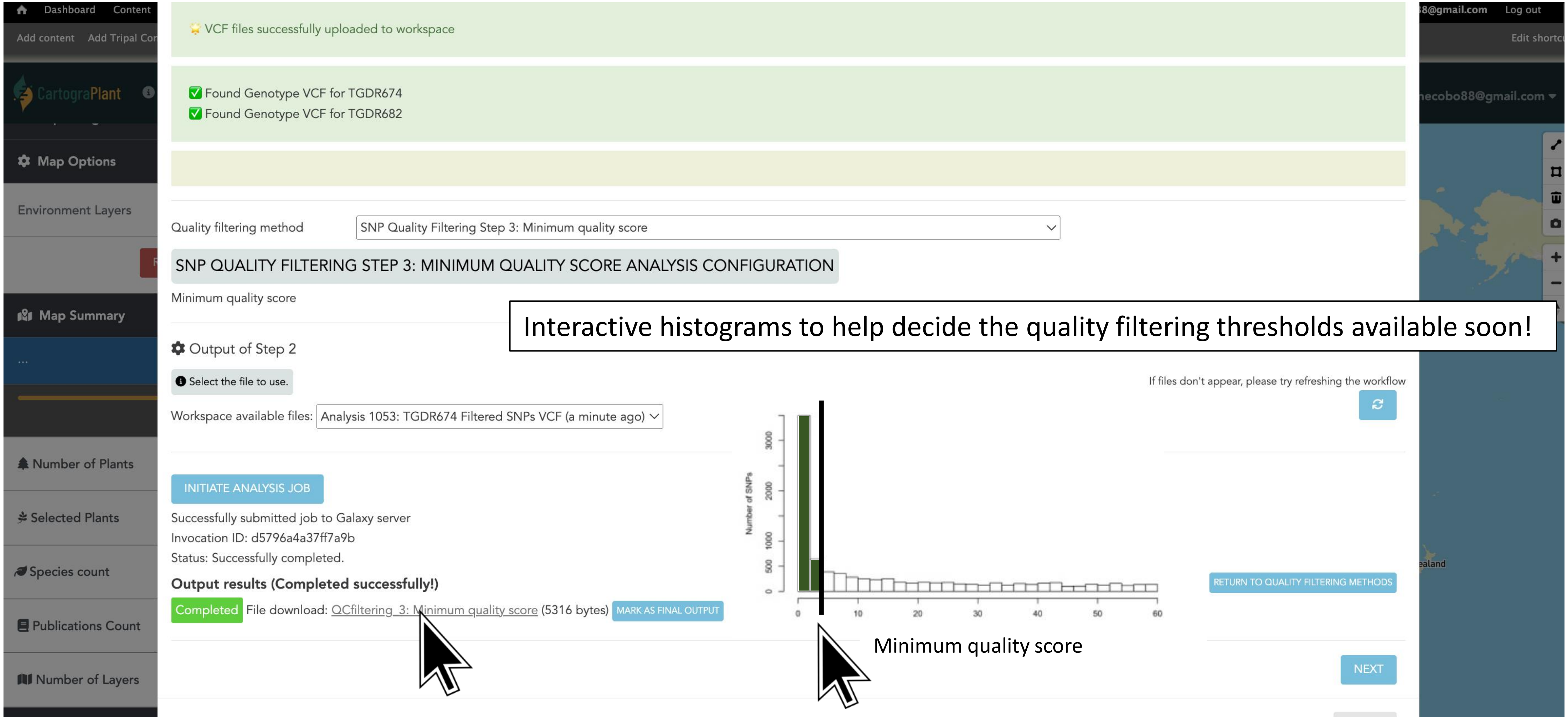
Workspace available files: Analysis 1053: TGDR674 Filtered SNPs VCF (a minute ago)

DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



VCF files successfully uploaded to workspace

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

Quality filtering method: SNP Quality Filtering Step 3: Minimum quality score

SNP QUALITY FILTERING STEP 3: MINIMUM QUALITY SCORE ANALYSIS CONFIGURATION

Minimum quality score

Output of Step 2

Select the file to use.

If files don't appear, please try refreshing the workflow

Workspace available files: Analysis 1053: TGDR674 Filtered SNPs VCF (a minute ago)

INITIATE ANALYSIS JOB

Successfully submitted job to Galaxy server
Invocation ID: d5796a4a37ff7a9b
Status: Successfully completed.

Output results (Completed successfully!)

Completed File download: QCfiltering_3: Minimum quality score (5316 bytes) MARK AS FINAL OUTPUT

RETURN TO QUALITY FILTERING METHODS

NEXT

Interactive histograms to help decide the quality filtering thresholds available soon!

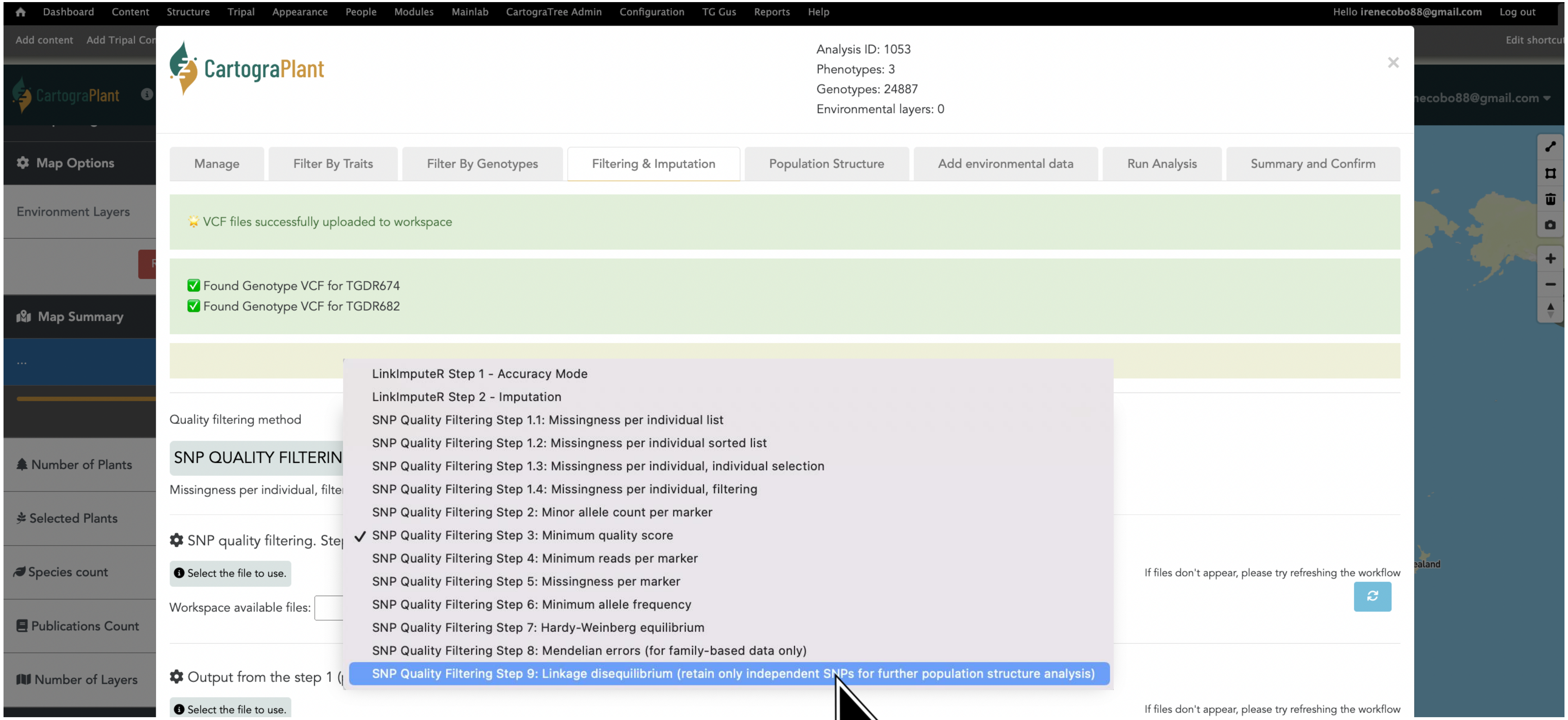
Minimum quality score

DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

VCF files successfully uploaded to workspace

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

Quality filtering method

SNP QUALITY FILTERIN

Missingness per individual, filte

SNP quality filtering. Ste

Select the file to use.

Workspace available files:

If files don't appear, please try refreshing the workflow

Output from the step 1 (

Select the file to use.

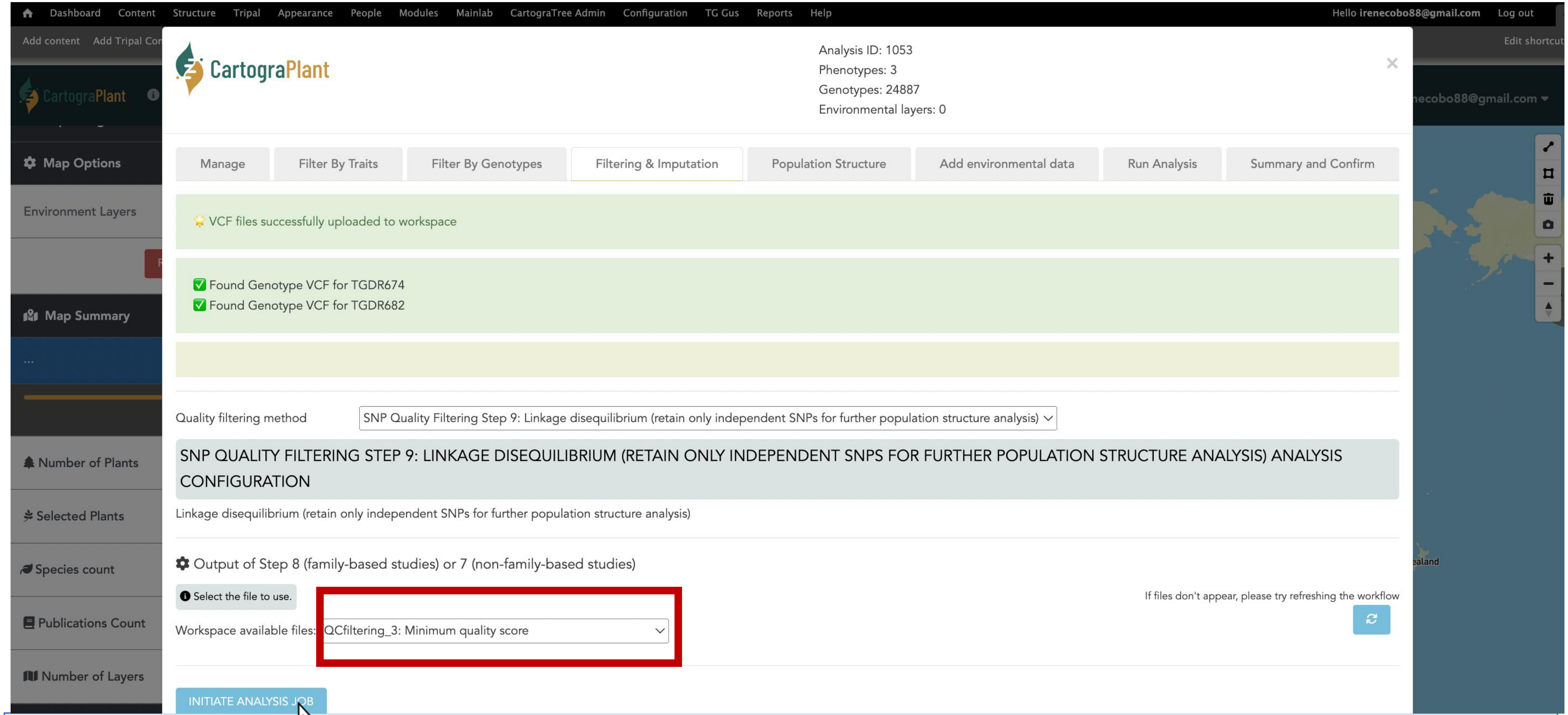
If files don't appear, please try refreshing the workflow

DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

VCF files successfully uploaded to workspace

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

Quality filtering method: SNP Quality Filtering Step 9: Linkage disequilibrium (retain only independent SNPs for further population structure analysis)

SNP QUALITY FILTERING STEP 9: LINKAGE DISEQUILIBRIUM (RETAIN ONLY INDEPENDENT SNPS FOR FURTHER POPULATION STRUCTURE ANALYSIS) ANALYSIS CONFIGURATION

Linkage disequilibrium (retain only independent SNPs for further population structure analysis)

Output of Step 8 (family-based studies) or 7 (non-family-based studies)

Select the file to use.

If files don't appear, please try refreshing the workflow

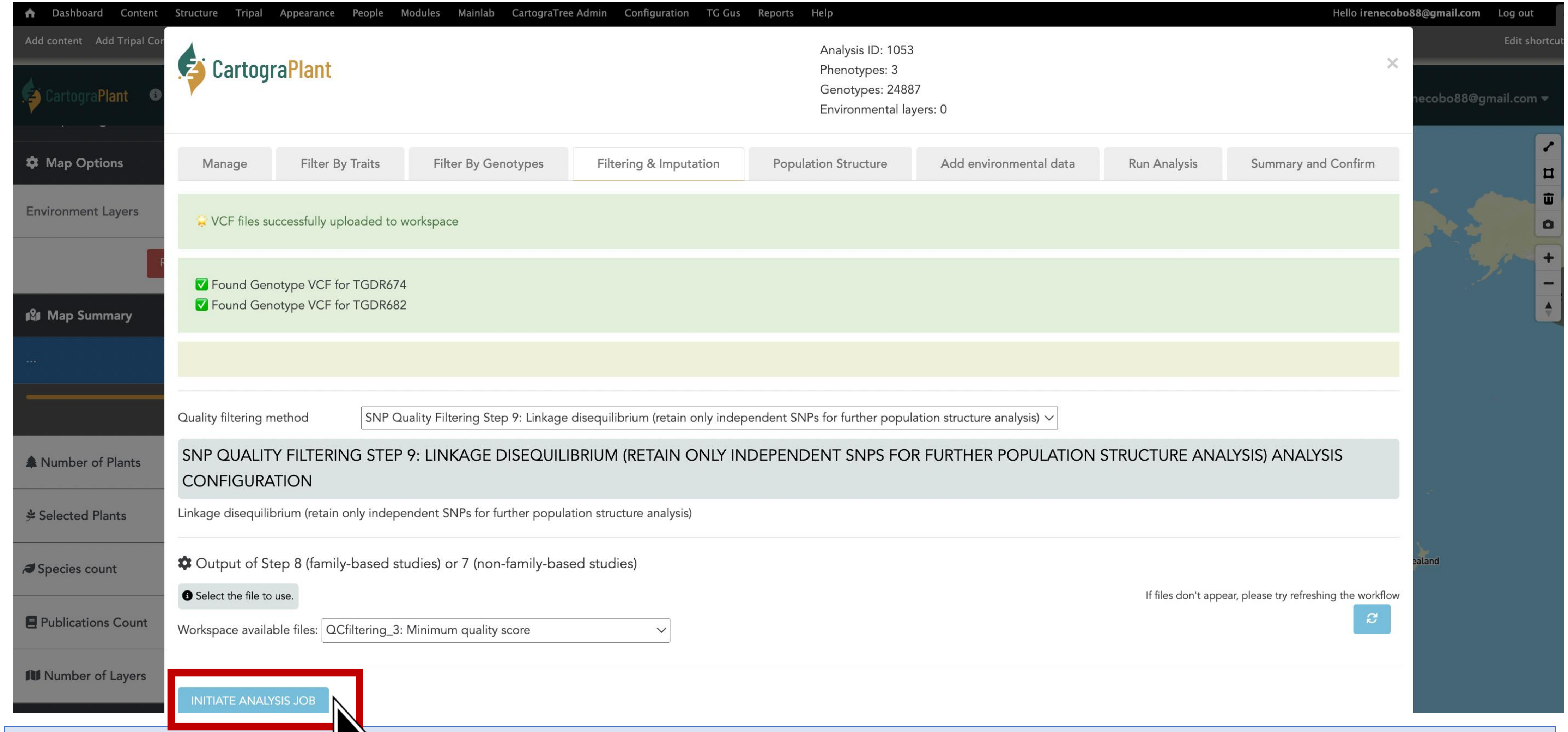
Workspace available files: QCfiltering_3: Minimum quality score

DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

VCF files successfully uploaded to workspace

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

Quality filtering method: SNP Quality Filtering Step 9: Linkage disequilibrium (retain only independent SNPs for further population structure analysis)

SNP QUALITY FILTERING STEP 9: LINKAGE DISEQUILIBRIUM (RETAIN ONLY INDEPENDENT SNPS FOR FURTHER POPULATION STRUCTURE ANALYSIS) ANALYSIS CONFIGURATION

Linkage disequilibrium (retain only independent SNPs for further population structure analysis)

Output of Step 8 (family-based studies) or 7 (non-family-based studies)

Select the file to use.

If files don't appear, please try refreshing the workflow

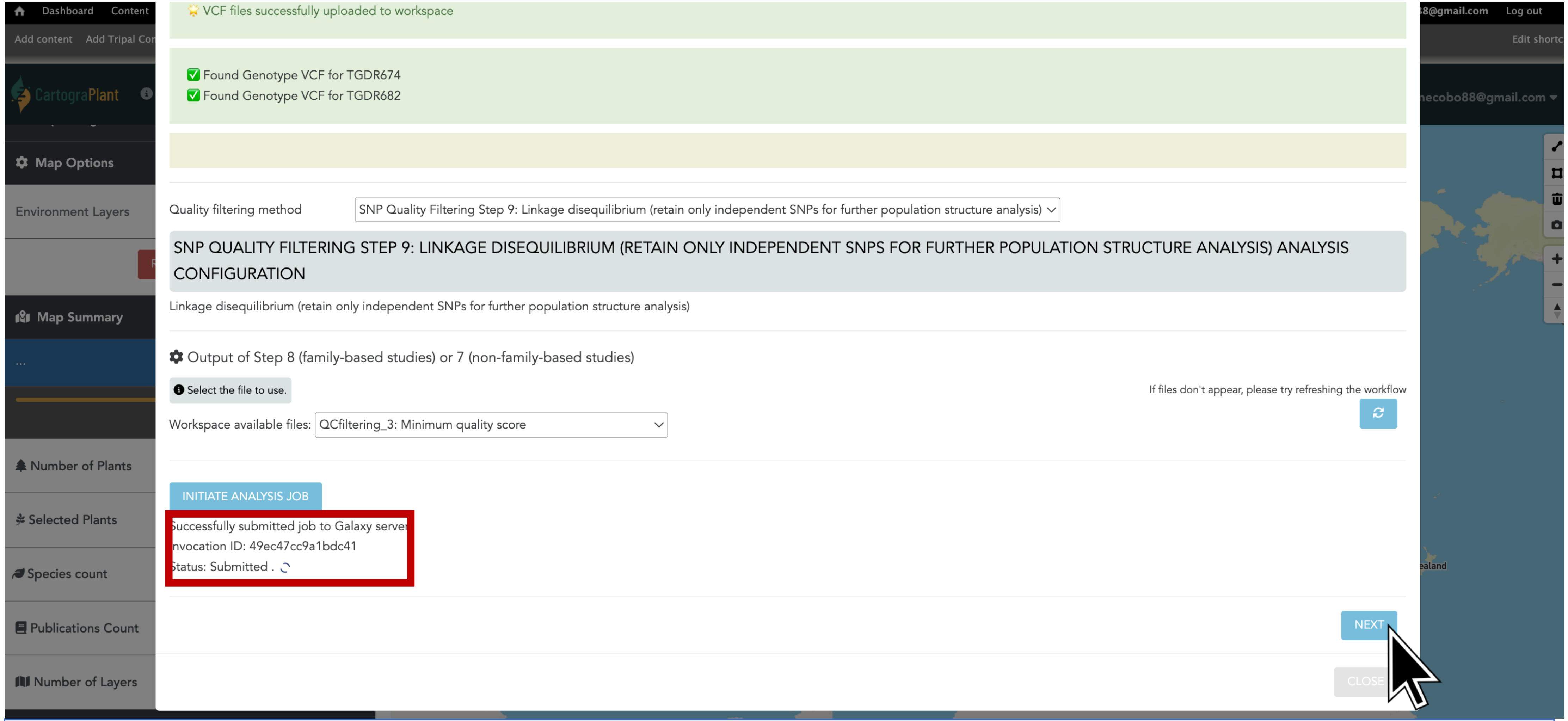
Workspace available files: QCfiltering_3: Minimum quality score

DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



VCF files successfully uploaded to workspace

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

Quality filtering method: SNP Quality Filtering Step 9: Linkage disequilibrium (retain only independent SNPs for further population structure analysis)

SNP QUALITY FILTERING STEP 9: LINKAGE DISEQUILIBRIUM (RETAIN ONLY INDEPENDENT SNPS FOR FURTHER POPULATION STRUCTURE ANALYSIS) ANALYSIS CONFIGURATION

Linkage disequilibrium (retain only independent SNPs for further population structure analysis)

Output of Step 8 (family-based studies) or 7 (non-family-based studies)

Select the file to use.

If files don't appear, please try refreshing the workflow

Workspace available files: QCfiltering_3: Minimum quality score

INITIATE ANALYSIS JOB

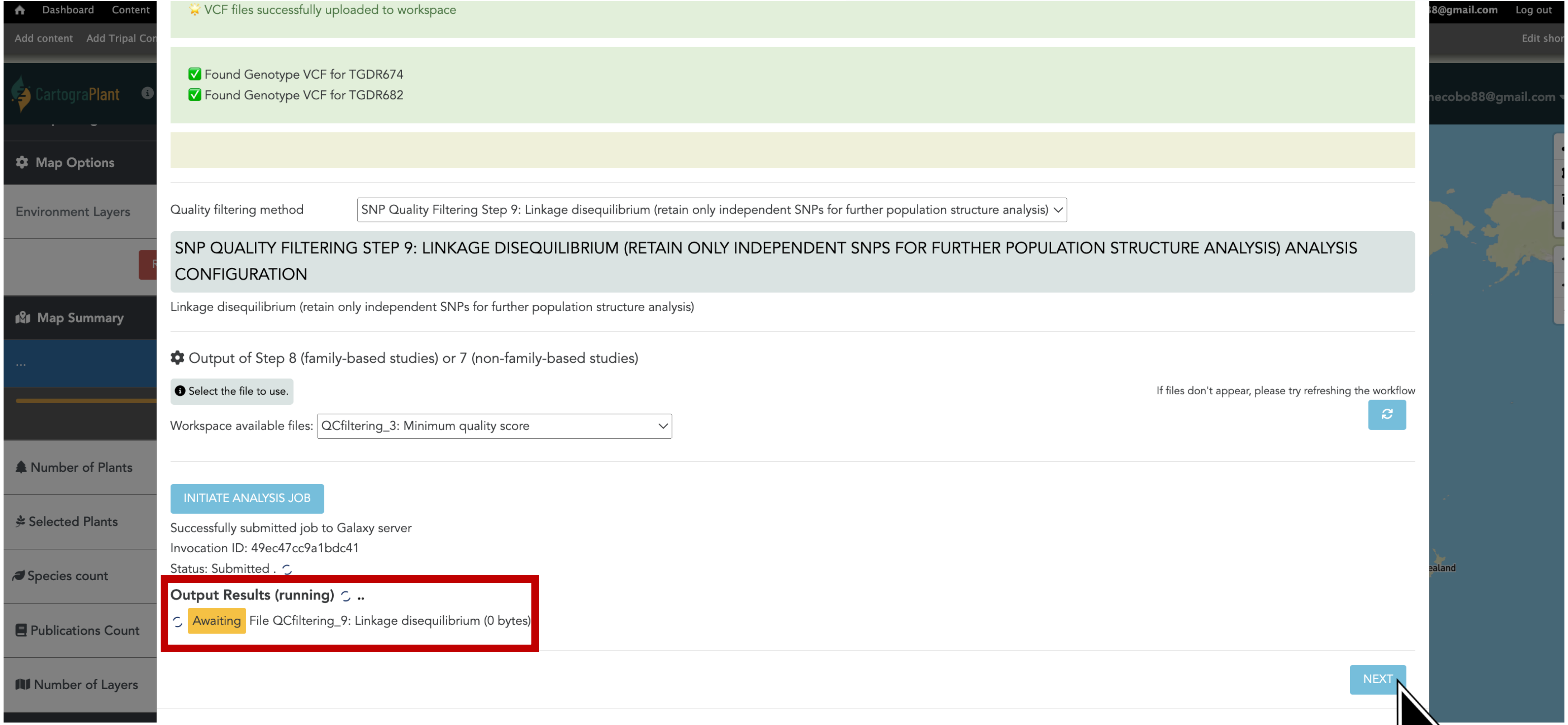
Successfully submitted job to Galaxy server
nvocation ID: 49ec47cc9a1bdc41
Status: Submitted .

DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



VCF files successfully uploaded to workspace

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

Quality filtering method: SNP Quality Filtering Step 9: Linkage disequilibrium (retain only independent SNPs for further population structure analysis)

SNP QUALITY FILTERING STEP 9: LINKAGE DISEQUILIBRIUM (RETAIN ONLY INDEPENDENT SNPS FOR FURTHER POPULATION STRUCTURE ANALYSIS) ANALYSIS CONFIGURATION

Linkage disequilibrium (retain only independent SNPs for further population structure analysis)

Output of Step 8 (family-based studies) or 7 (non-family-based studies)

Select the file to use.

If files don't appear, please try refreshing the workflow

Workspace available files: QCfiltering_3: Minimum quality score

Successfully submitted job to Galaxy server
Invocation ID: 49ec47cc9a1bdc41
Status: Submitted .

Output Results (running) ..

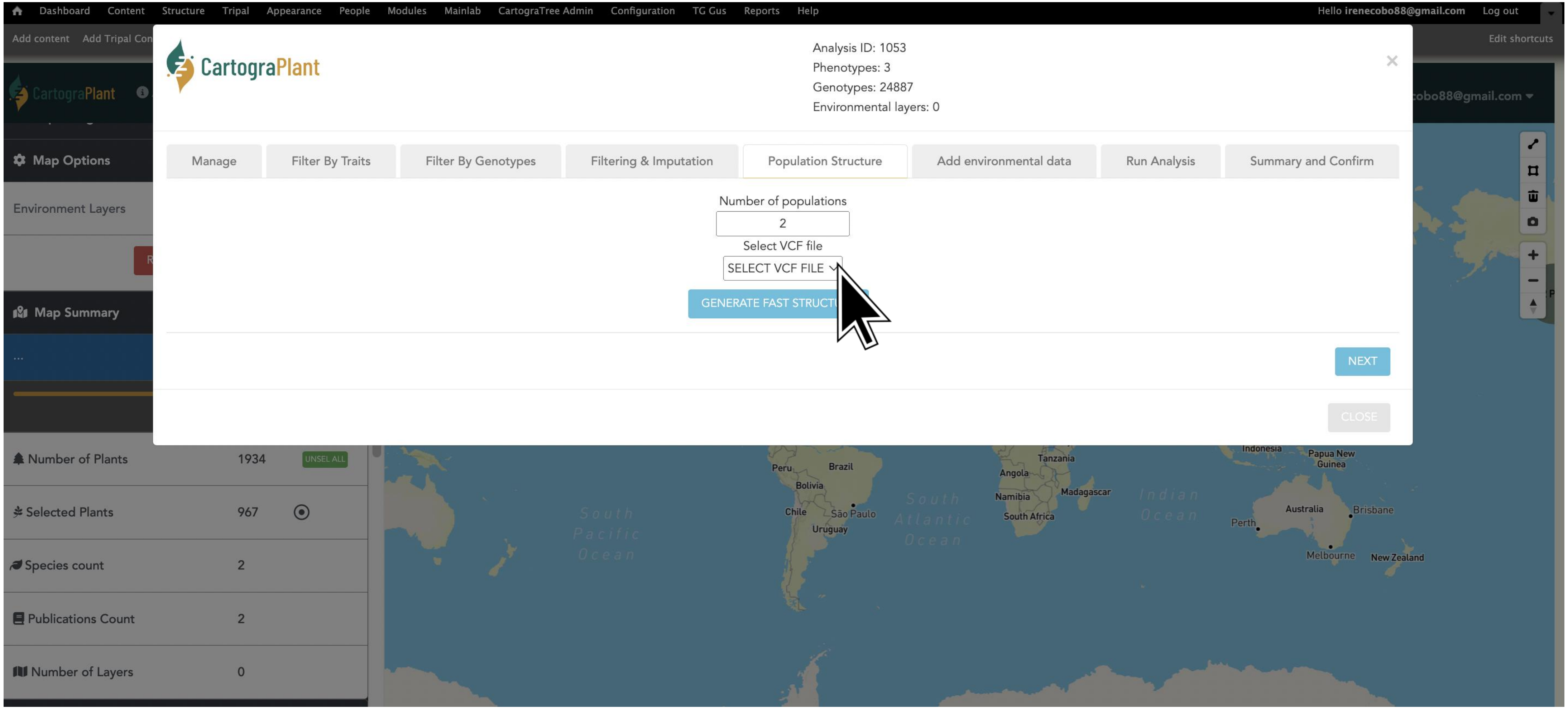
Awaiting File QCfiltering_9: Linkage disequilibrium (0 bytes)

DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

Number of populations: 2

Select VCF file: SELECT VCF FILE

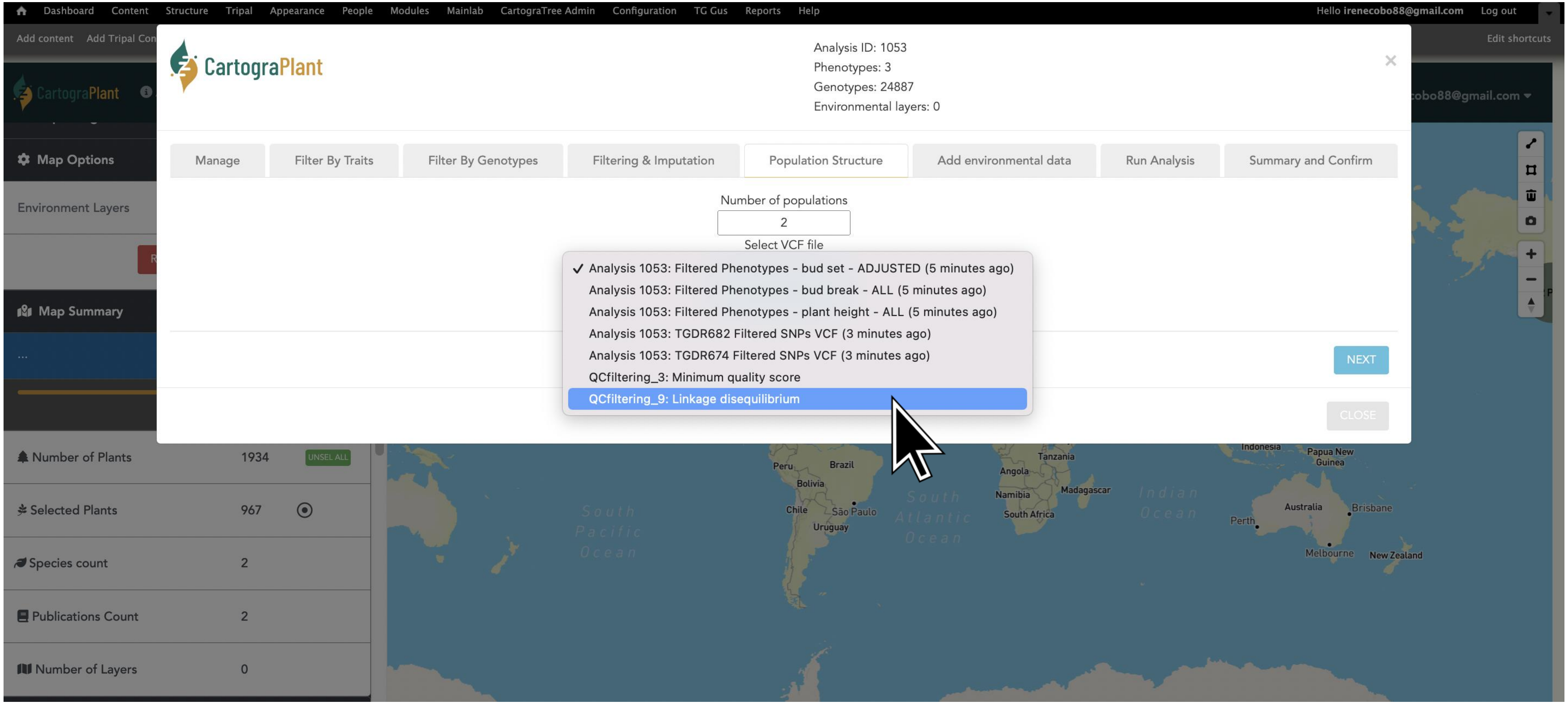
Number of Plants 1934
Selected Plants 967
Species count 2
Publications Count 2
Number of Layers 0

DATA ANALYSIS: POPULATION STRUCTURE CALCULATION/VISUALIZATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

Number of populations: 2

Select VCF file:
- ✓ Analysis 1053: Filtered Phenotypes - bud set - ADJUSTED (5 minutes ago)
- Analysis 1053: Filtered Phenotypes - bud break - ALL (5 minutes ago)
- Analysis 1053: Filtered Phenotypes - plant height - ALL (5 minutes ago)
- Analysis 1053: TGDR682 Filtered SNPs VCF (3 minutes ago)
- Analysis 1053: TGDR674 Filtered SNPs VCF (3 minutes ago)
- QCfiltering_3: Minimum quality score
- QCfiltering_9: Linkage disequilibrium

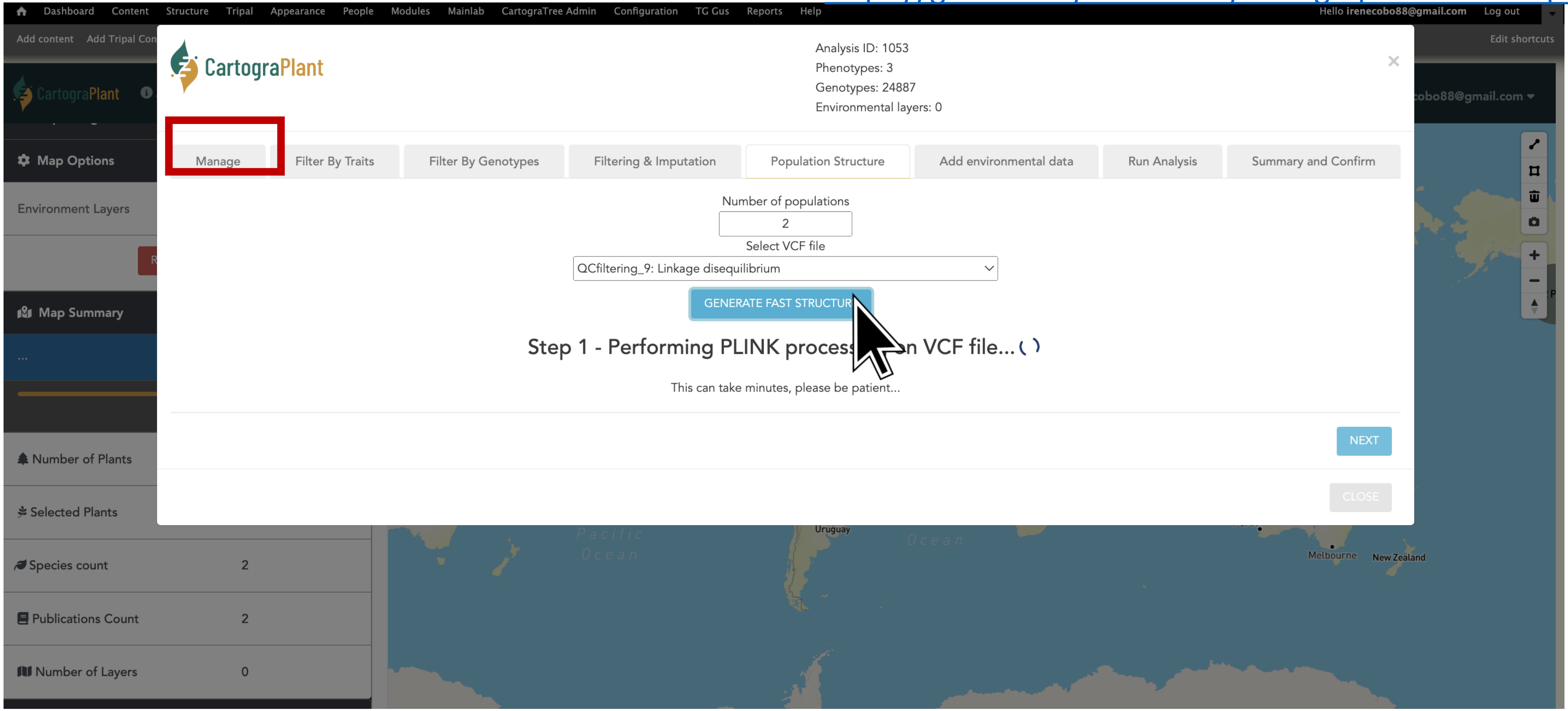
Number of Plants: 1934
Selected Plants: 967
Species count: 2
Publications Count: 2
Number of Layers: 0

DATA ANALYSIS: POPULATION STRUCTURE CALCULATION/VISUALIZATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

Number of populations: 2

Select VCF file: QCfiltering_9: Linkage disequilibrium

Step 1 - Performing PLINK proces n VCF file...

This can take minutes, please be patient...

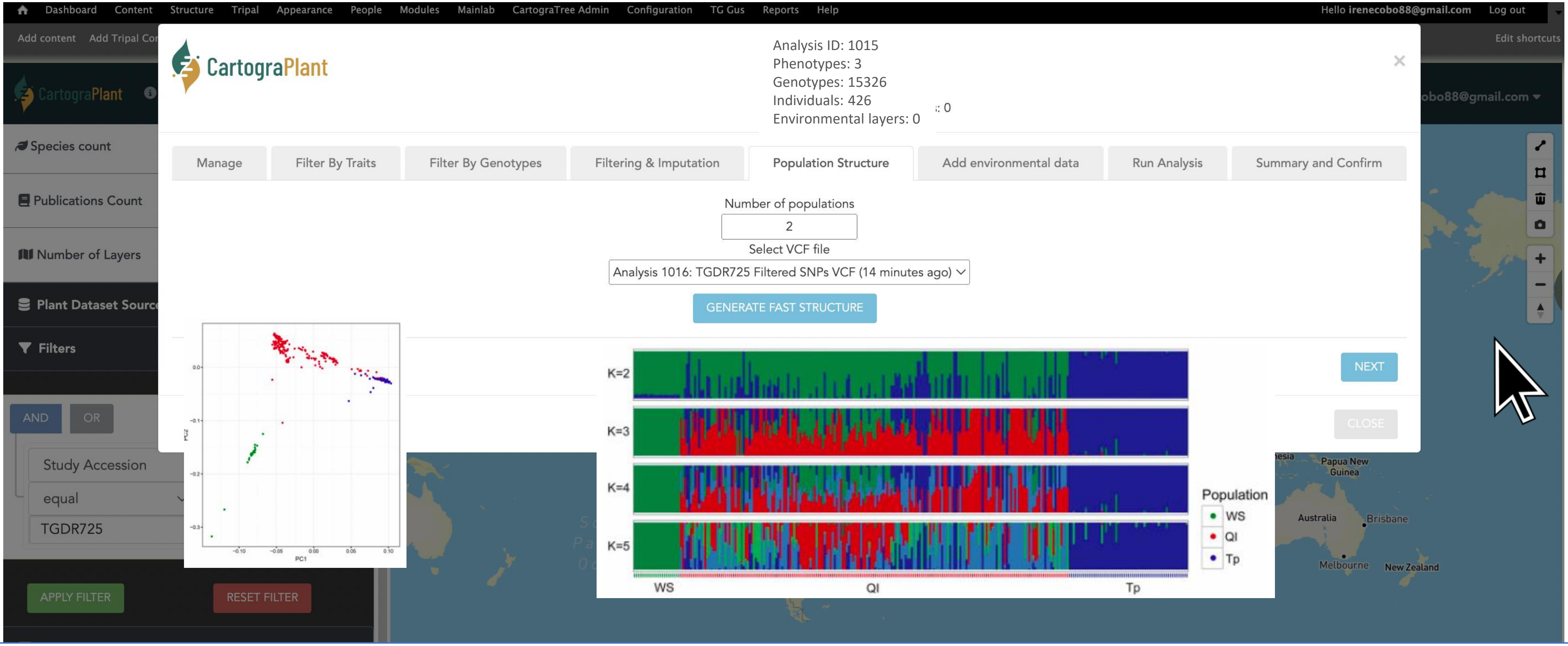
Number of Plants	
Selected Plants	
Species count	2
Publications Count	2
Number of Layers	0

DATA ANALYSIS: POPULATION STRUCTURE CALCULATION/VISUALIZATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/


Analysis ID: 1015
Phenotypes: 3
Genotypes: 15326
Individuals: 426
Environmental layers: 0

Number of populations: 2
Select VCF file: Analysis 1016: TGDR725 Filtered SNPs VCF (14 minutes ago)

Study Accession
equal
TGDR725

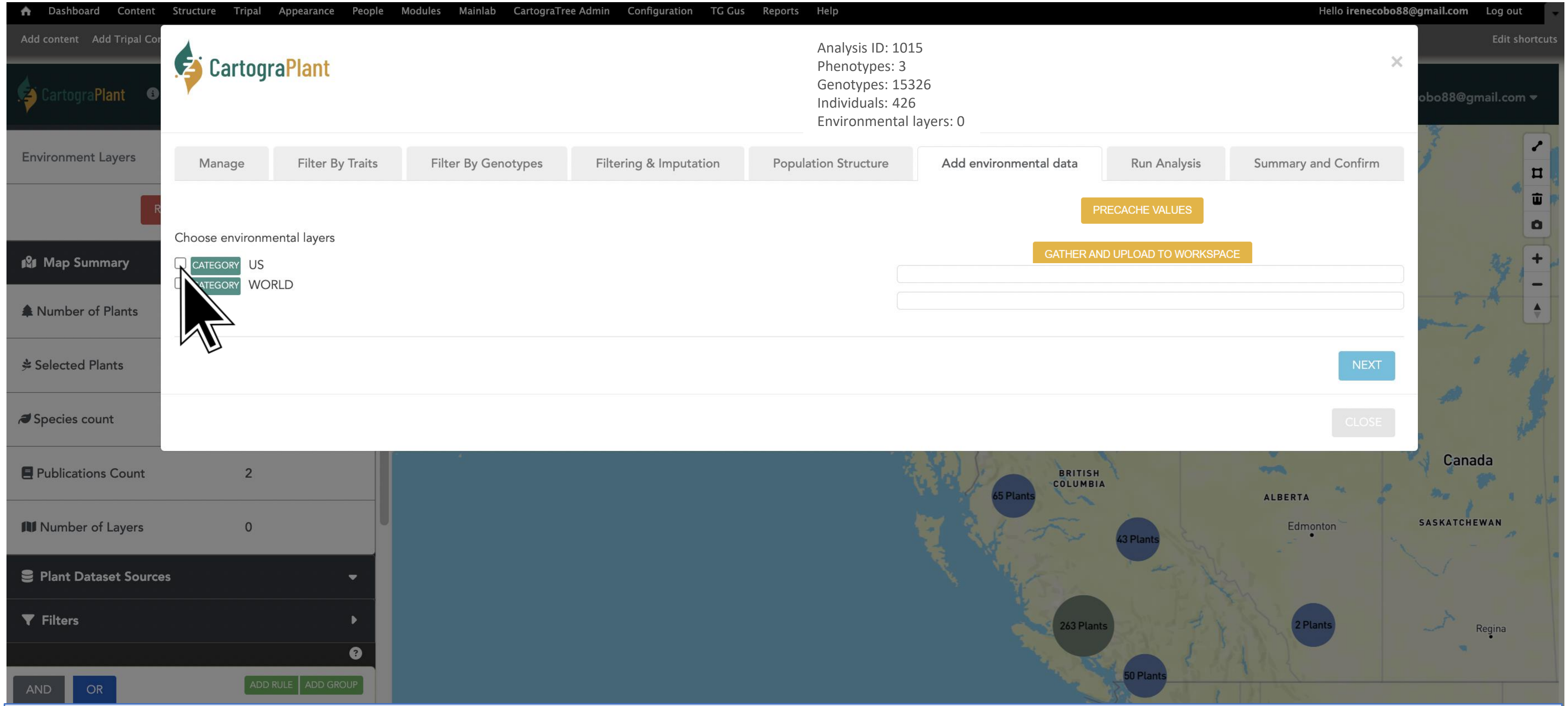
Population structure calculation (PCA and DAPC) and visualization (fastSTRUCTURE, PCA and DAPC) available soon!!!

DATA ANALYSIS: POPULATION STRUCTURE CALCULATION/VISUALIZATION

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



Analysis ID: 1015
Phenotypes: 3
Genotypes: 15326
Individuals: 426
Environmental layers: 0

Choose environmental layers

☐ CATEGORY US
☐ CATEGORY WORLD

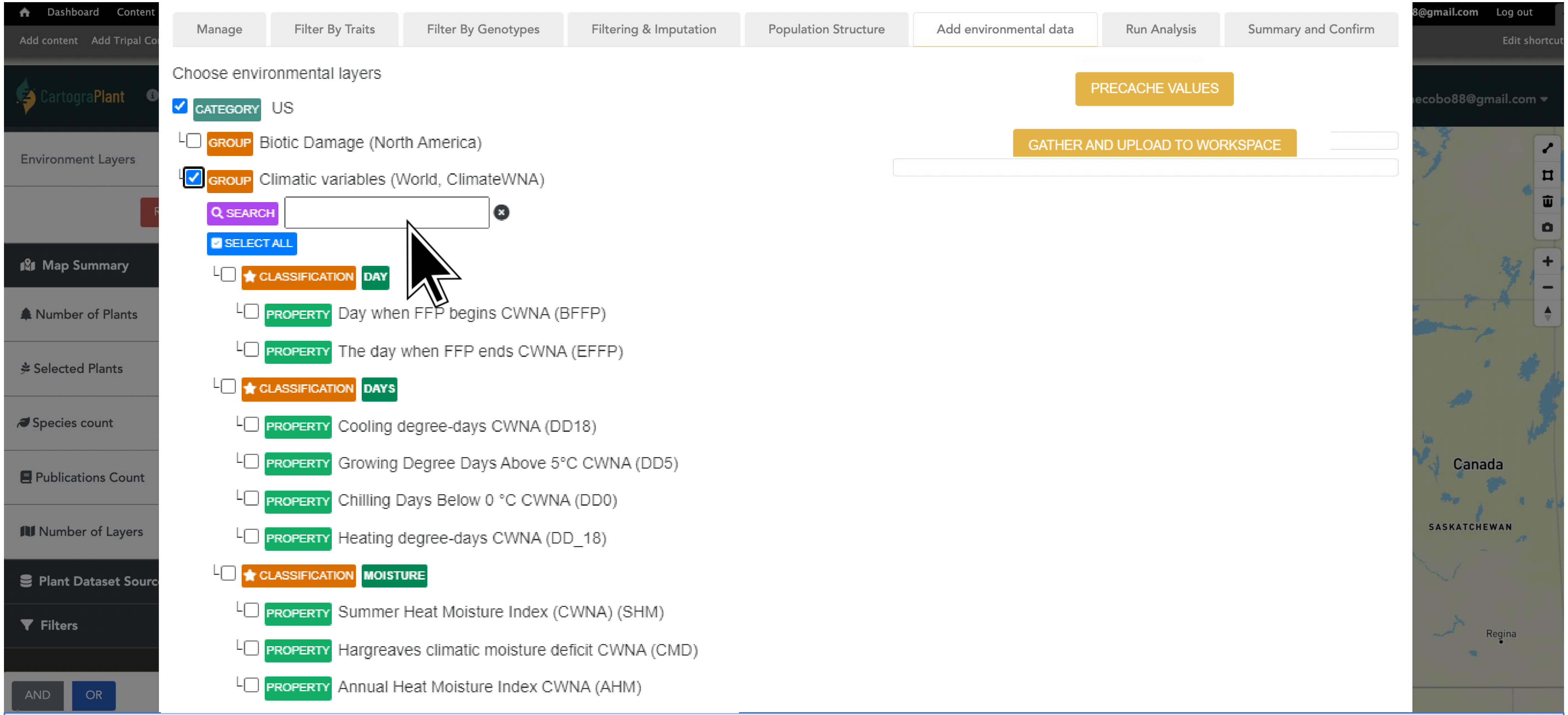
Publications Count 2
Number of Layers 0

DATA ANALYSIS: ENVIRONMENTAL DATA SELECTION/ MULTICOLLINEARITY

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



Choose environmental layers

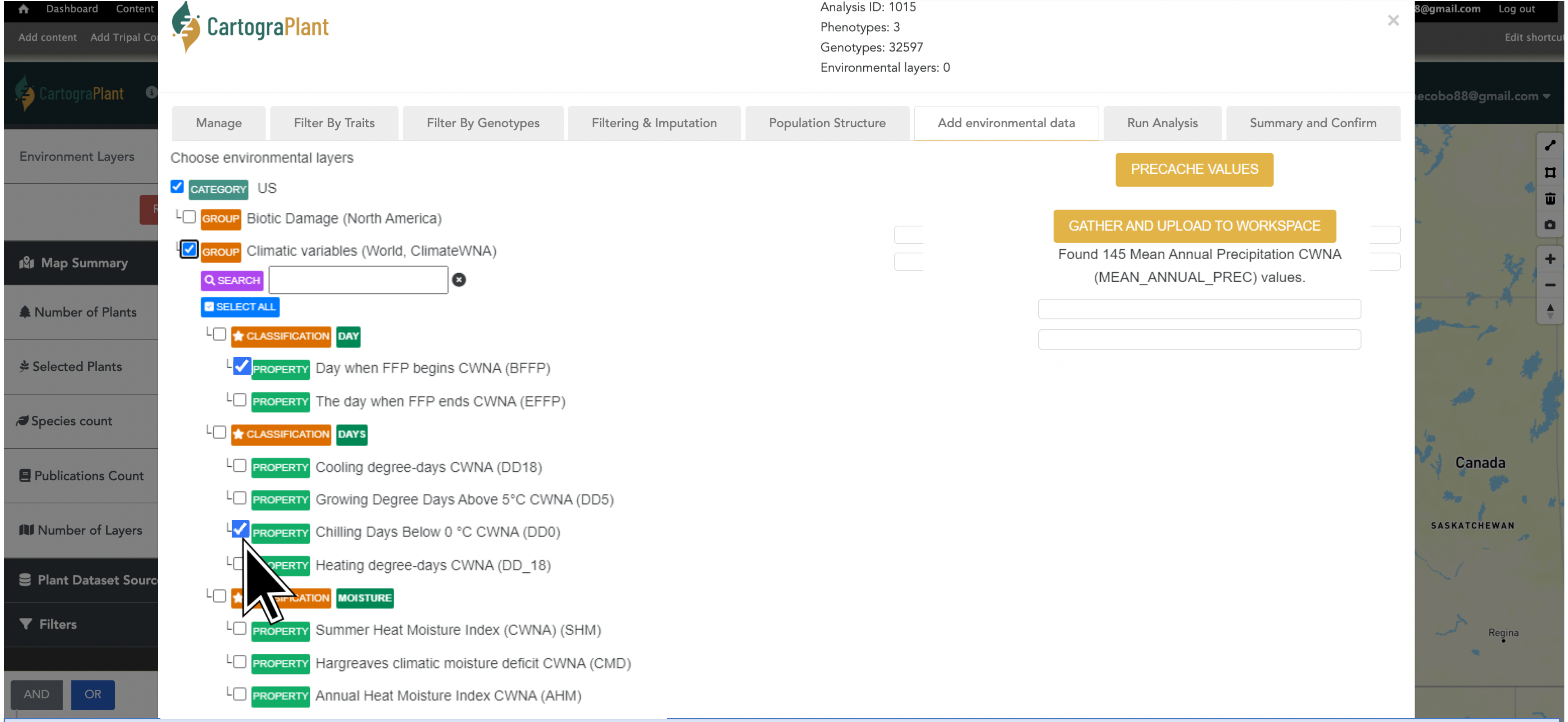
☑ CATEGORY US
- ☐ GROUP Biotic Damage (North America)
- ☑ GROUP Climatic variables (World, ClimateWNA)
  - SEARCH
  - SELECT ALL
  - ☐ CLASSIFICATION DAY
    - ☐ PROPERTY Day when FFP begins CWNA (BFFP)
    - ☐ PROPERTY The day when FFP ends CWNA (EFFP)
  - ☐ CLASSIFICATION DAYS
    - ☐ PROPERTY Cooling degree-days CWNA (DD18)
    - ☐ PROPERTY Growing Degree Days Above 5°C CWNA (DD5)
    - ☐ PROPERTY Chilling Days Below 0 °C CWNA (DD0)
    - ☐ PROPERTY Heating degree-days CWNA (DD_18)
  - ☐ CLASSIFICATION MOISTURE
    - ☐ PROPERTY Summer Heat Moisture Index (CWNA) (SHM)
    - ☐ PROPERTY Hargreaves climatic moisture deficit CWNA (CMD)
    - ☐ PROPERTY Annual Heat Moisture Index CWNA (AHM)

DATA ANALYSIS: ENVIRONMENTAL DATA SELECTION/ MULTICOLLINEARITY

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

Choose environmental layers

- ☑ CATEGORY US
  - ☐ GROUP Biotic Damage (North America)
  - ☑ GROUP Climatic variables (World, ClimateWNA)
    - ☐ CLASSIFICATION DAY
      - ☑ PROPERTY Day when FFP begins CWNA (BFFP)
      - ☐ PROPERTY The day when FFP ends CWNA (EFFP)
    - ☐ CLASSIFICATION DAYS
      - ☐ PROPERTY Cooling degree-days CWNA (DD18)
      - ☐ PROPERTY Growing Degree Days Above 5°C CWNA (DD5)
      - ☑ PROPERTY Chilling Days Below 0 °C CWNA (DD0)
      - ☐ PROPERTY Heating degree-days CWNA (DD_18)
    - ☐ CLASSIFICATION MOISTURE
      - ☐ PROPERTY Summer Heat Moisture Index (CWNA) (SHM)
      - ☐ PROPERTY Hargreaves climatic moisture deficit CWNA (CMD)
      - ☐ PROPERTY Annual Heat Moisture Index CWNA (AHM)

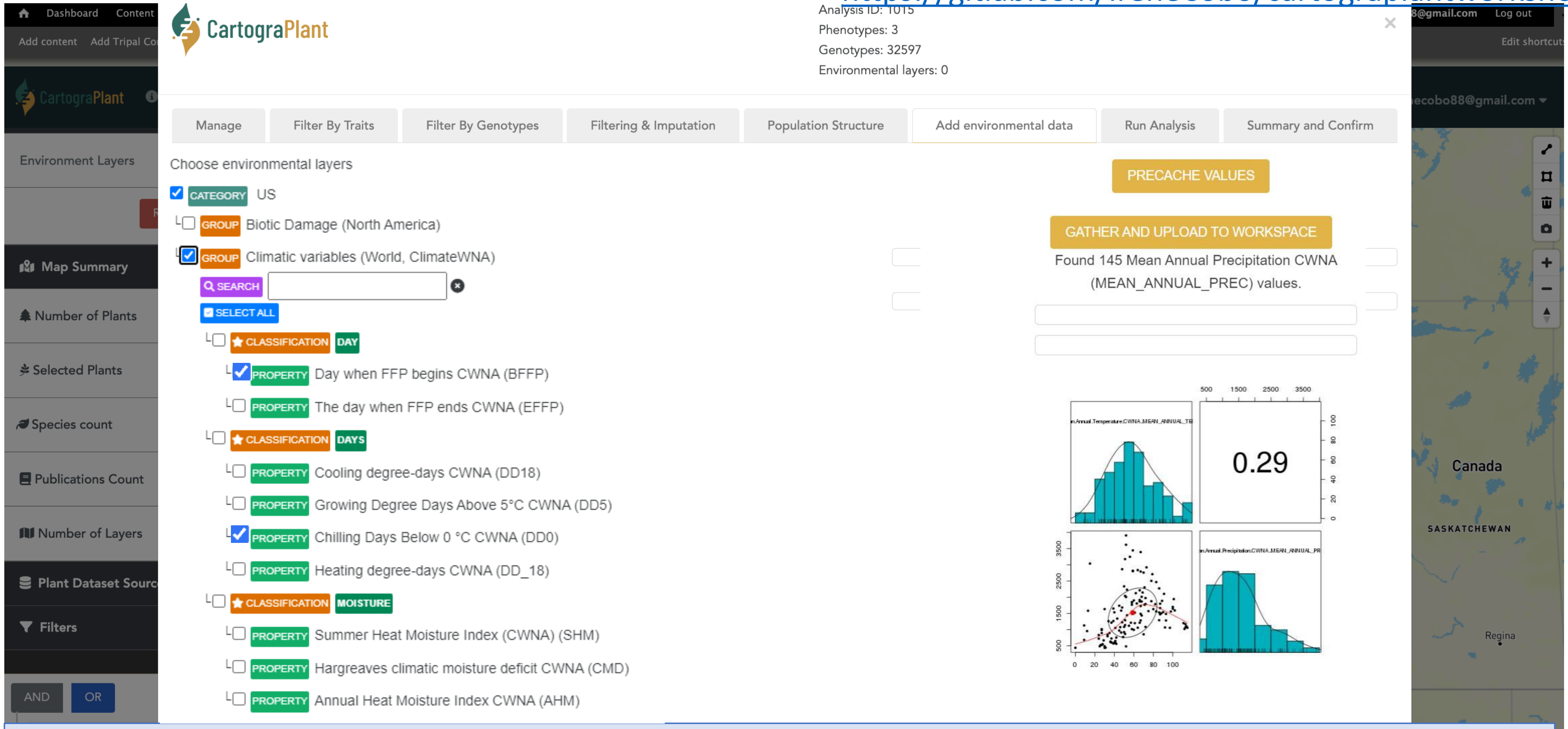
Found 145 Mean Annual Precipitation CWNA (MEAN_ANNUAL_PREC) values.

DATA ANALYSIS: ENVIRONMENTAL DATA SELECTION/ MULTICOLLINEARITY

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

Choose environmental layers

- ☑ CATEGORY US
  - ☐ GROUP Biotic Damage (North America)
  - ☑ GROUP Climatic variables (World, ClimateWNA)
    - SEARCH
    - SELECT ALL
    - ☐ CLASSIFICATION DAY
      - ☑ PROPERTY Day when FFP begins CWNA (BFFP)
      - ☐ PROPERTY The day when FFP ends CWNA (EFFP)
    - ☐ CLASSIFICATION DAYS
      - ☐ PROPERTY Cooling degree-days CWNA (DD18)
      - ☐ PROPERTY Growing Degree Days Above 5°C CWNA (DD5)
      - ☑ PROPERTY Chilling Days Below 0 °C CWNA (DD0)
      - ☐ PROPERTY Heating degree-days CWNA (DD_18)
    - ☐ CLASSIFICATION MOISTURE
      - ☐ PROPERTY Summer Heat Moisture Index (CWNA) (SHM)
      - ☐ PROPERTY Hargreaves climatic moisture deficit CWNA (CMD)
      - ☐ PROPERTY Annual Heat Moisture Index CWNA (AHM)

PRECACHE VALUES

GATHER AND UPLOAD TO WORKSPACE

Found 145 Mean Annual Precipitation CWNA (MEAN_ANNUAL_PREC) values.

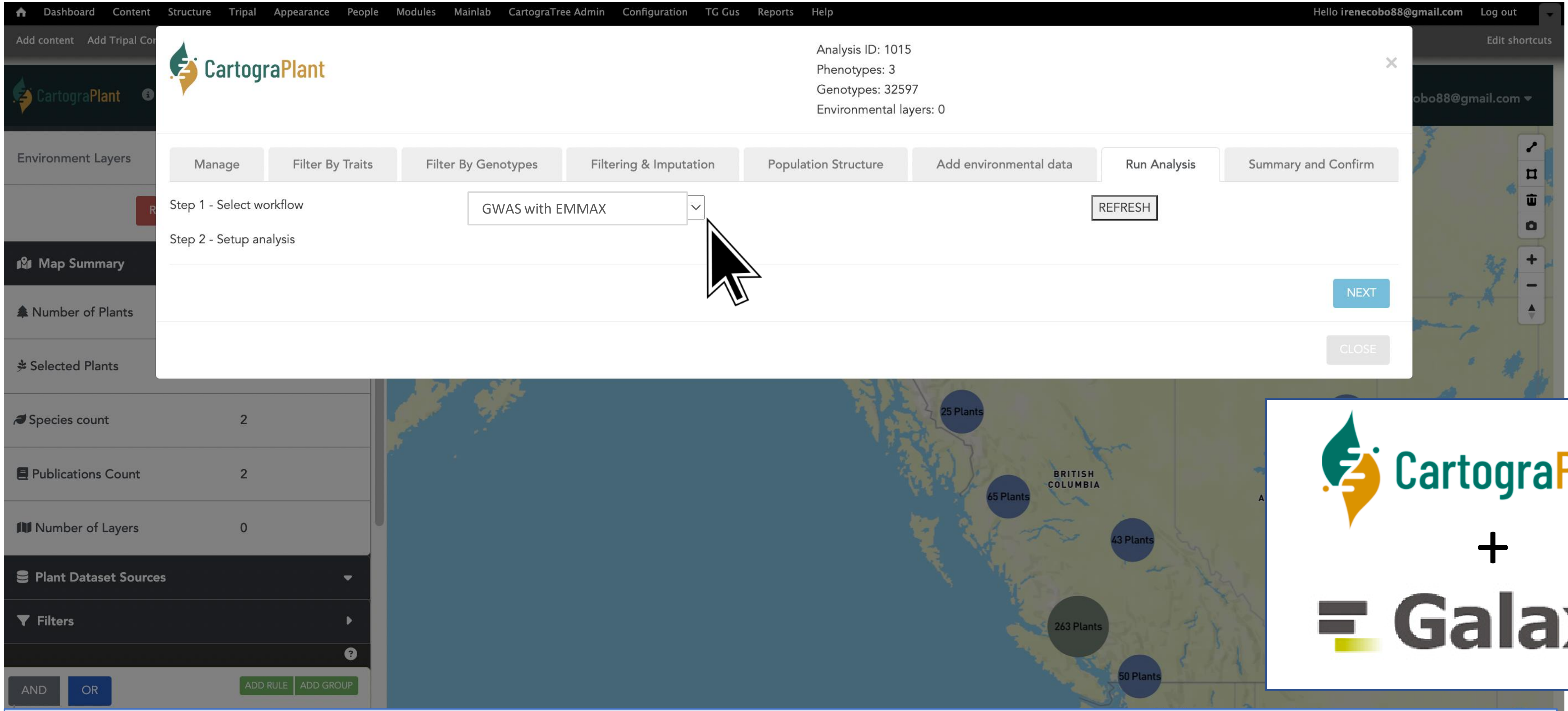
0.29

DATA ANALYSIS: ENVIRONMENTAL DATA SELECTION/ MULTICOLLINEARITY

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

Step 1 - Select workflow: GWAS with EMMAX
Step 2 - Setup analysis

Species count 2
Publications Count 2
Number of Layers 0

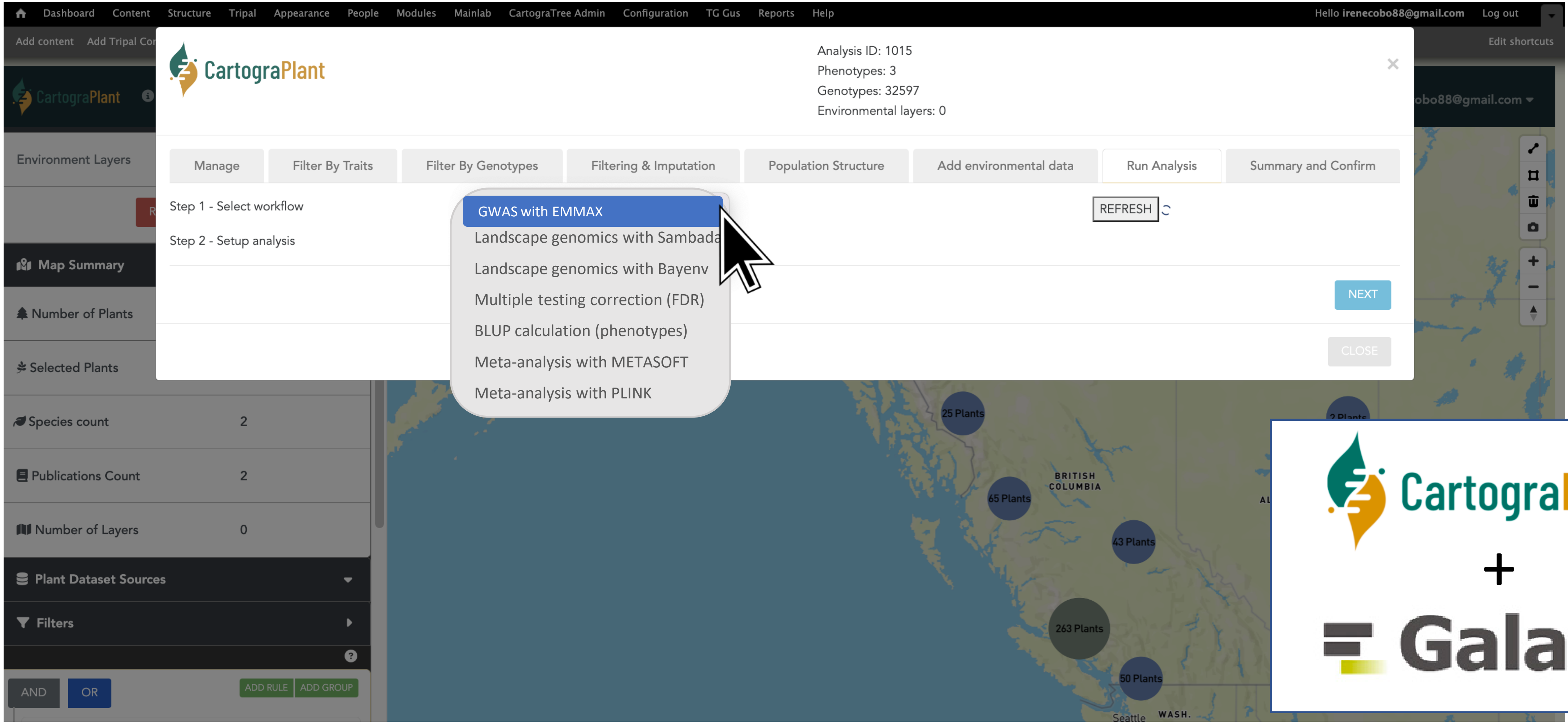
CartograPlant
+
Galaxy

DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

Step 1 - Select workflow

Step 2 - Setup analysis

Species count	2
Publications Count	2
Number of Layers	0

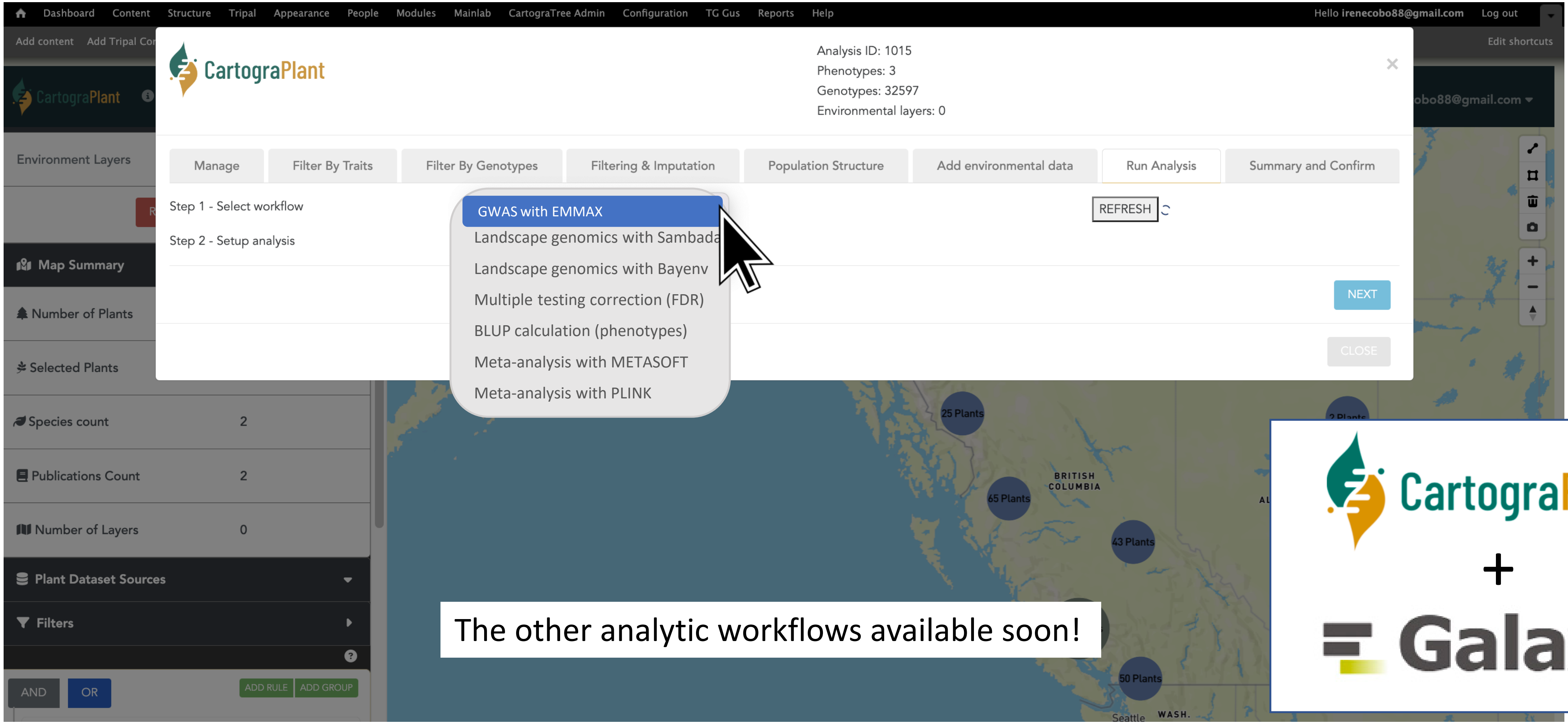
CartograPlant
+
Galaxy

DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



Analysis ID: 1015

Phenotypes: 3

Genotypes: 32597

Environmental layers: 0

Step 1 - Select workflow

Step 2 - Setup analysis

The other analytic workflows available soon!

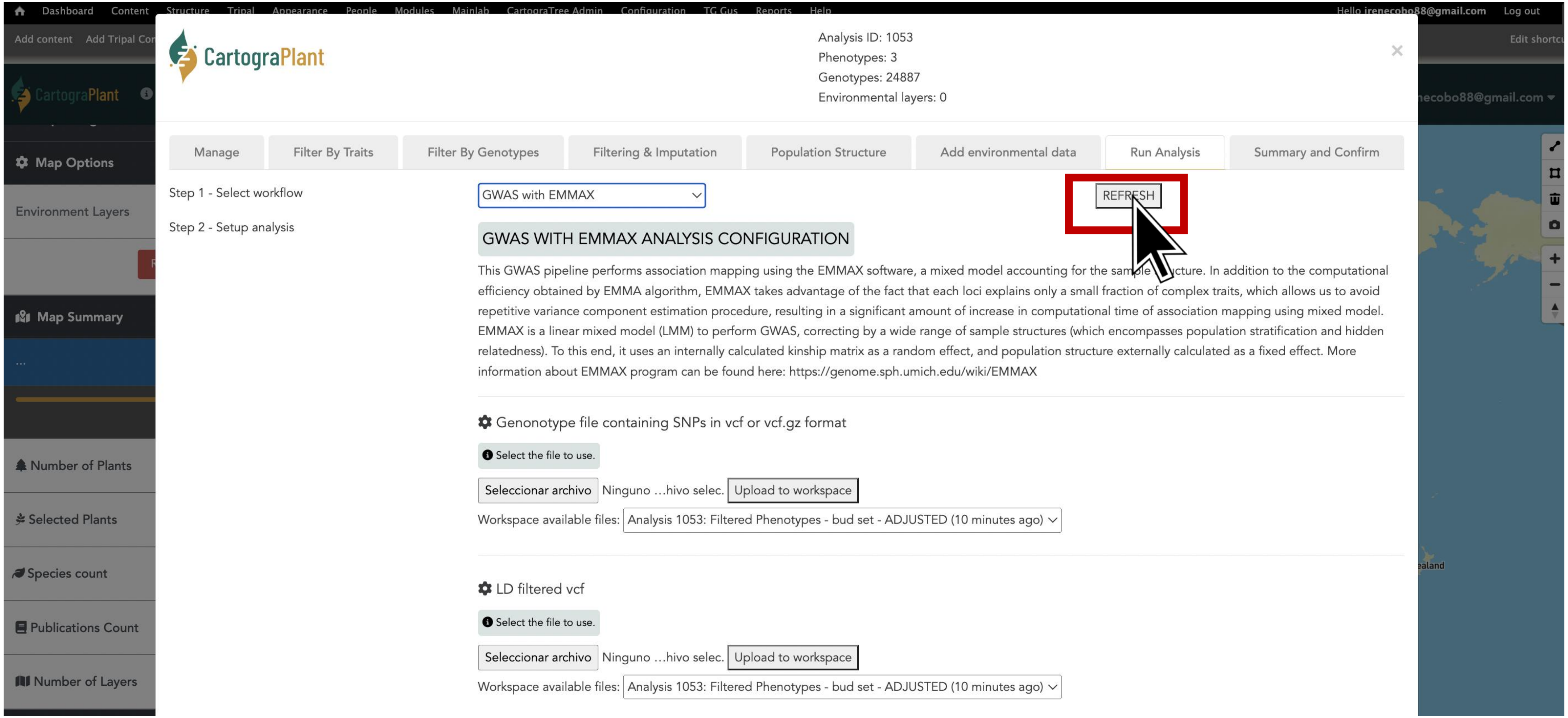
CartograPlant + Galaxy

DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

Step 1 - Select workflow: GWAS with EMMAX

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

Genonotype file containing SNPs in vcf or vcf.gz format

Select the file to use.

Ninguno ...hivo selec.

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

LD filtered vcf

Select the file to use.

Ninguno ...hivo selec.

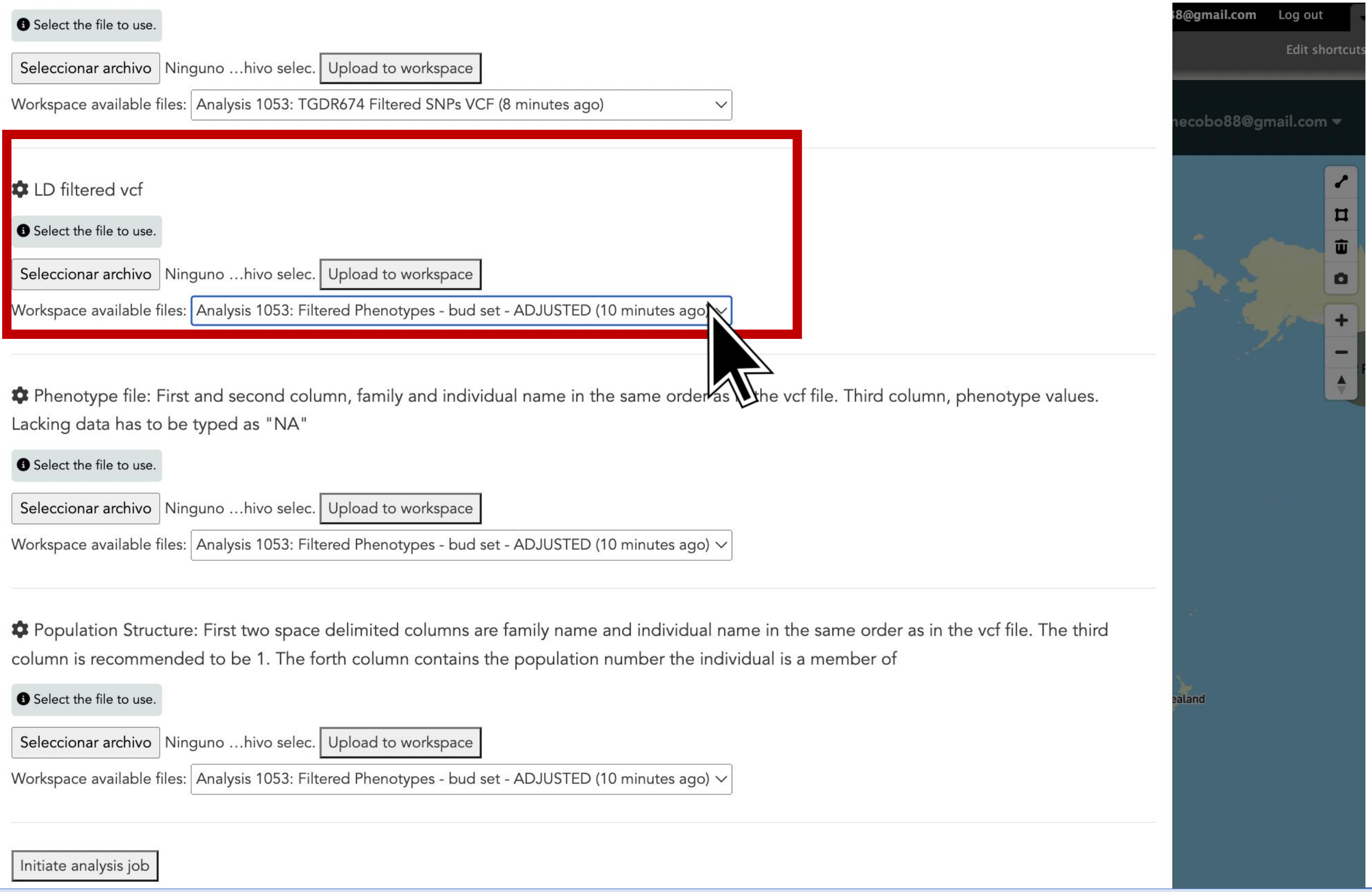
Workspace available files: Analysis 1053: Filtered Phenotypes - bud set - ADJUSTED (10 minutes ago)

DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



Select the file to use.

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

Workspace available files: Analysis 1053: TGDR674 Filtered SNPs VCF (8 minutes ago)

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)

Phenotype file: First and second column, family and individual name in the same order as in the vcf file. Third column, phenotype values. Lacking data has to be typed as "NA"

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)

Population Structure: First two space delimited columns are family name and individual name in the same order as in the vcf file. The third column is recommended to be 1. The forth column contains the population number the individual is a member of

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)

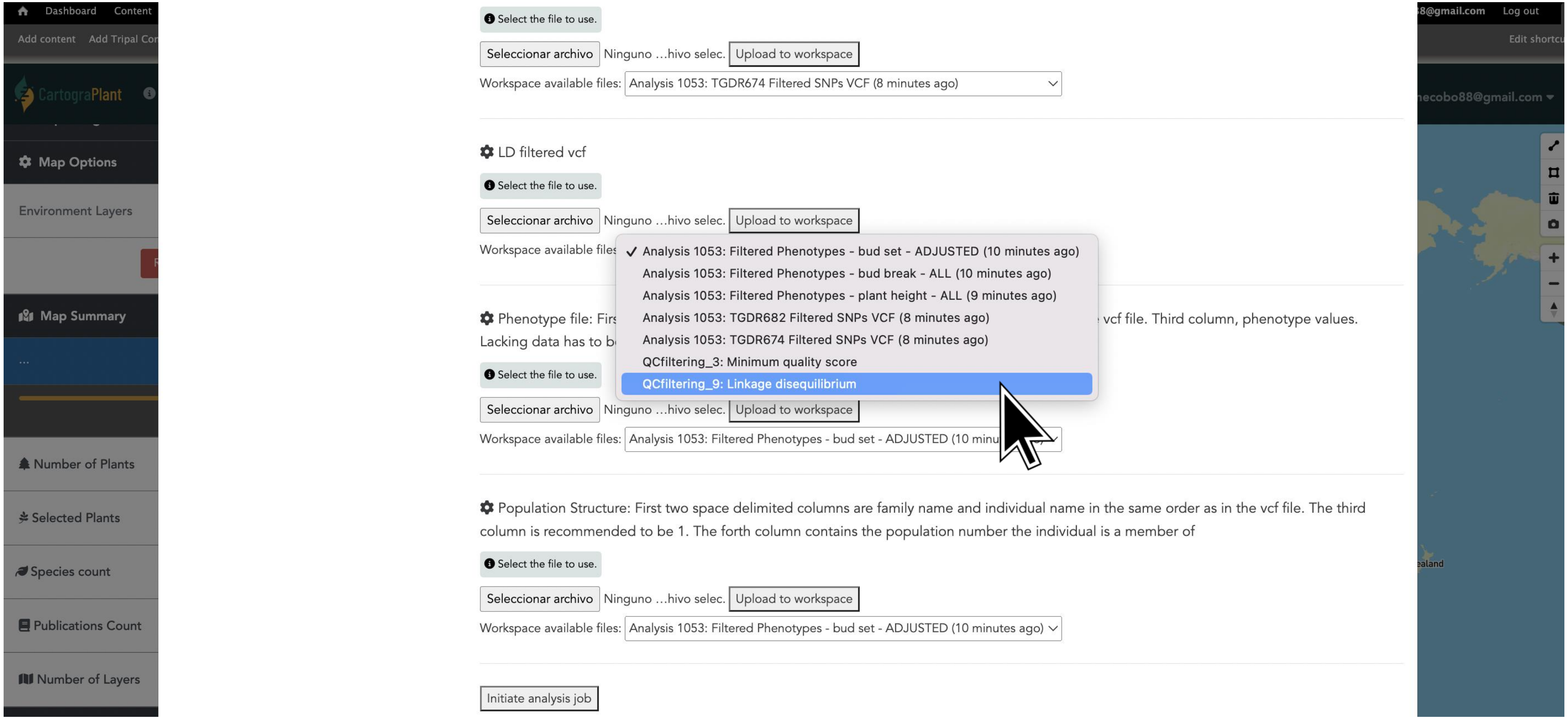
Initiate analysis job

DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



Select the file to use.

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

Workspace available files: Analysis 1053: TGDR674 Filtered SNPs VCF (8 minutes ago)

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)
- Analysis 1053: Filtered Phenotypes - bud break - ALL (10 minutes ago)
- Analysis 1053: Filtered Phenotypes - plant height - ALL (9 minutes ago)
- Analysis 1053: TGDR682 Filtered SNPs VCF (8 minutes ago)
- Analysis 1053: TGDR674 Filtered SNPs VCF (8 minutes ago)
- QCfiltering_3: Minimum quality score
- QCfiltering_9: Linkage disequilibrium

Phenotype file: Firs ... vcf file. Third column, phenotype values. Lacking data has to b

Select the file to use.

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

Workspace available files: Analysis 1053: Filtered Phenotypes - bud set - ADJUSTED (10 minu

Population Structure: First two space delimited columns are family name and individual name in the same order as in the vcf file. The third column is recommended to be 1. The forth column contains the population number the individual is a member of

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)

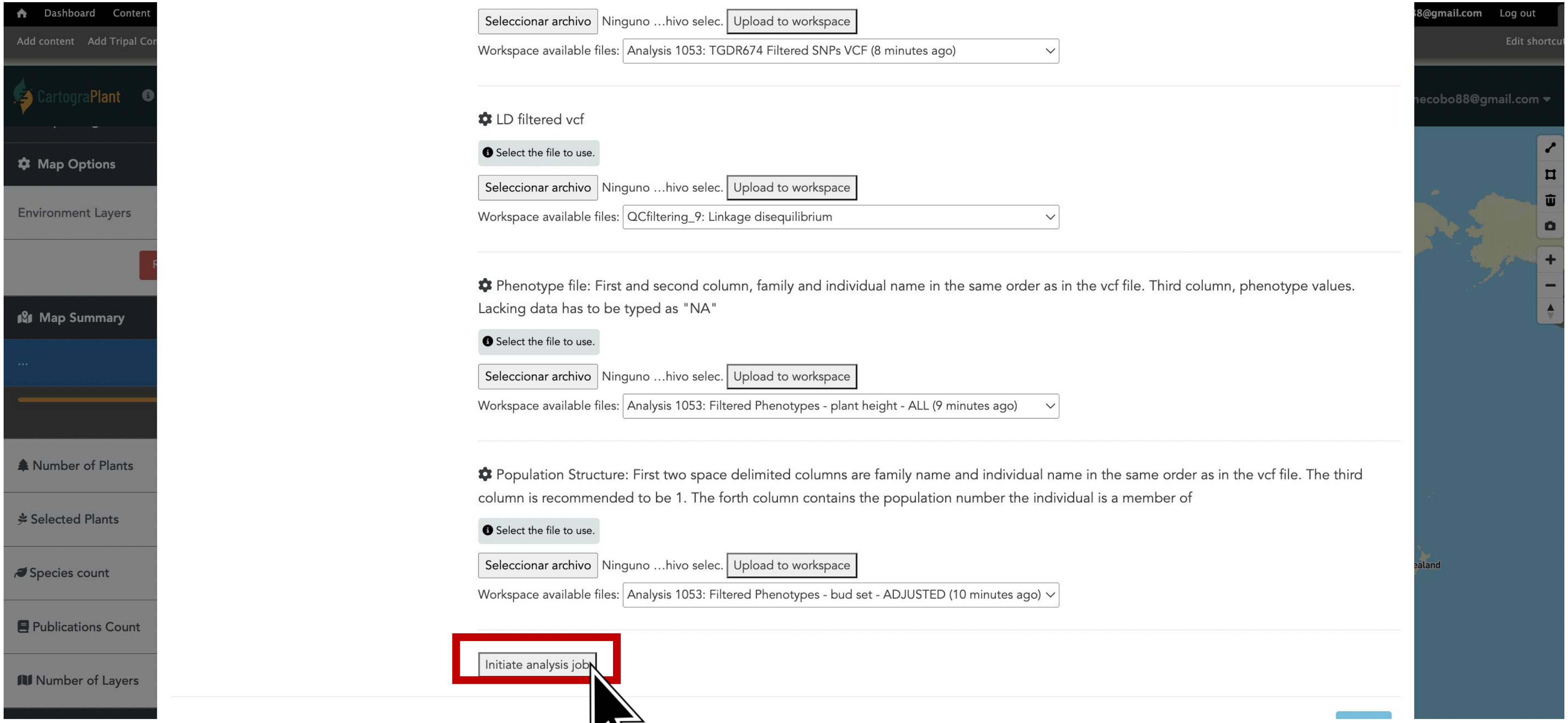
Initiate analysis job

DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



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

Workspace available files: Analysis 1053: TGDR674 Filtered SNPs VCF (8 minutes ago)

LD filtered vcf

Select the file to use.

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

Workspace available files: QCfiltering_9: Linkage disequilibrium

Phenotype file: First and second column, family and individual name in the same order as in the vcf file. Third column, phenotype values. Lacking data has to be typed as "NA"

Select the file to use.

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

Workspace available files: Analysis 1053: Filtered Phenotypes - plant height - ALL (9 minutes ago)

Population Structure: First two space delimited columns are family name and individual name in the same order as in the vcf file. The third column is recommended to be 1. The forth column contains the population number the individual is a member of

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)

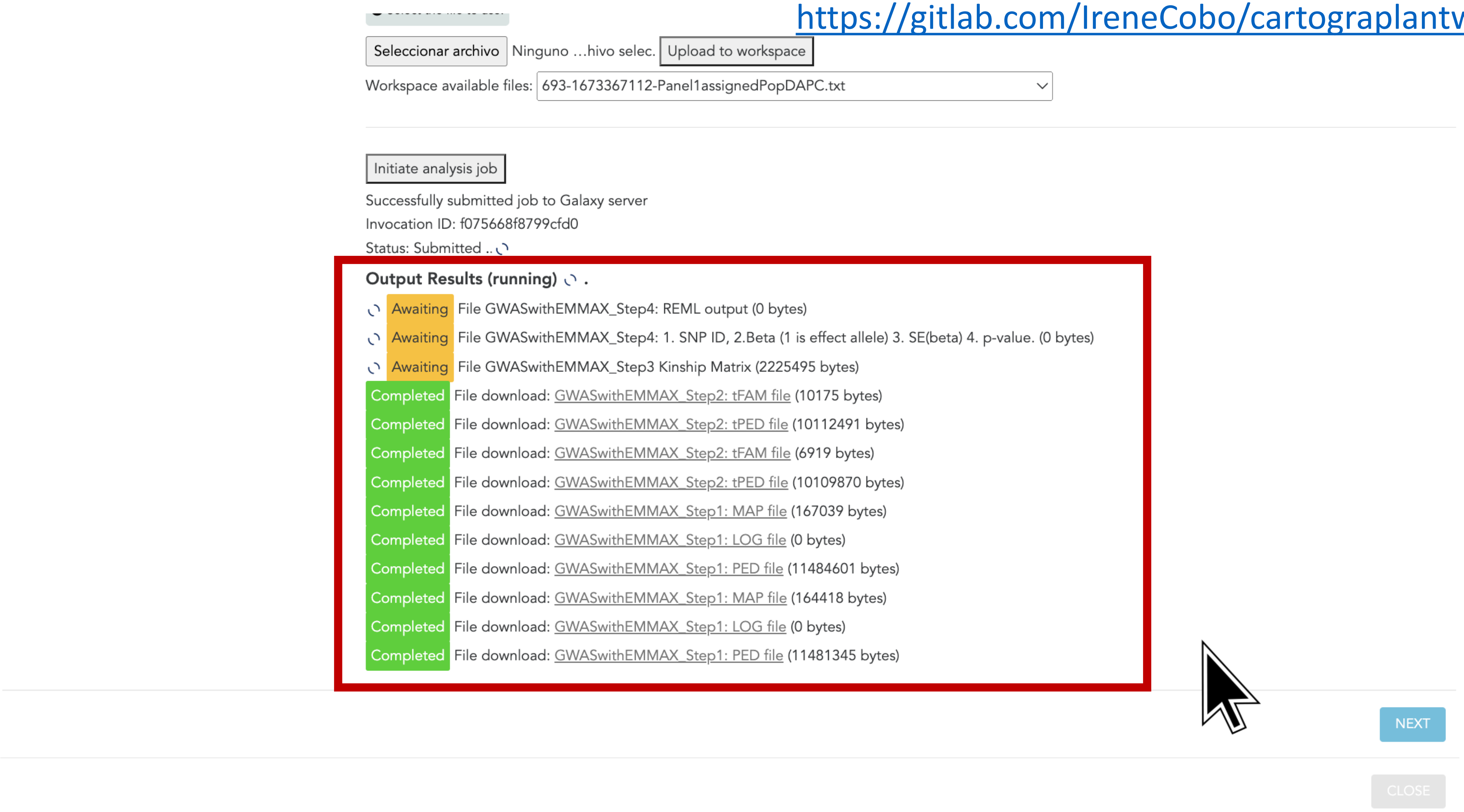
Initiate analysis job

DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



Workspace available files: 693-1673367112-Panel1assignedPopDAPC.txt

Successfully submitted job to Galaxy server
Invocation ID: f075668f8799cfd0
Status: Submitted ..

Output Results (running) .

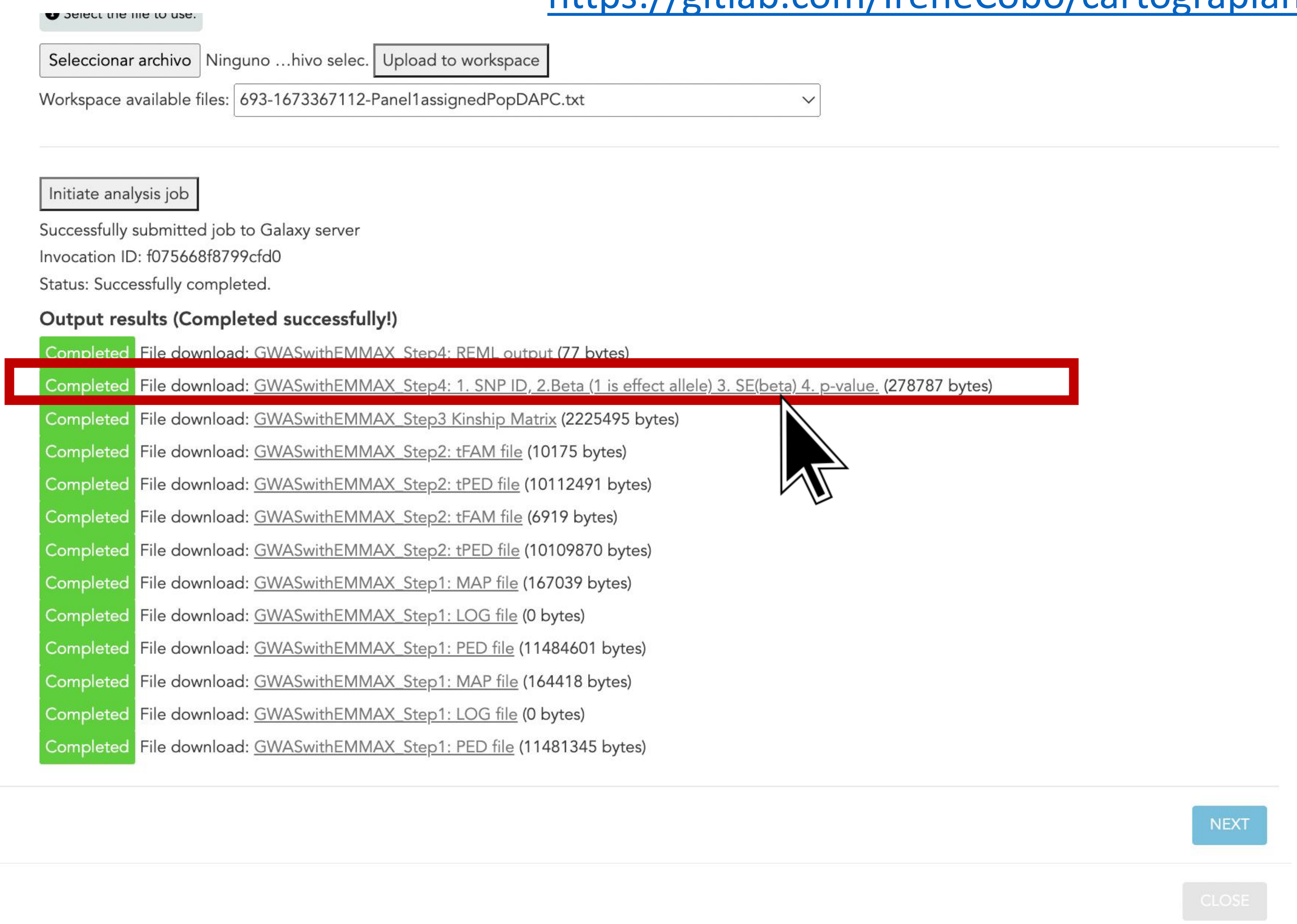
- Awaiting File GWASwithEMMAX_Step4: REML output (0 bytes)
- Awaiting File GWASwithEMMAX_Step4: 1. SNP ID, 2.Beta (1 is effect allele) 3. SE(beta) 4. p-value. (0 bytes)
- Awaiting File GWASwithEMMAX_Step3 Kinship Matrix (2225495 bytes)
- Completed File download: GWASwithEMMAX_Step2: tFAM file (10175 bytes)
- Completed File download: GWASwithEMMAX_Step2: tPED file (10112491 bytes)
- Completed File download: GWASwithEMMAX_Step2: tFAM file (6919 bytes)
- Completed File download: GWASwithEMMAX_Step2: tPED file (10109870 bytes)
- Completed File download: GWASwithEMMAX_Step1: MAP file (167039 bytes)
- Completed File download: GWASwithEMMAX_Step1: LOG file (0 bytes)
- Completed File download: GWASwithEMMAX_Step1: PED file (11484601 bytes)
- Completed File download: GWASwithEMMAX_Step1: MAP file (164418 bytes)
- Completed File download: GWASwithEMMAX_Step1: LOG file (0 bytes)
- Completed File download: GWASwithEMMAX_Step1: PED file (11481345 bytes)

DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



Ninguno …hivo selec.

Workspace available files: 693-1673367112-Panel1assignedPopDAPC.txt

Successfully submitted job to Galaxy server

Invocation ID: f075668f8799cfd0

Status: Successfully completed.

Output results (Completed successfully!)

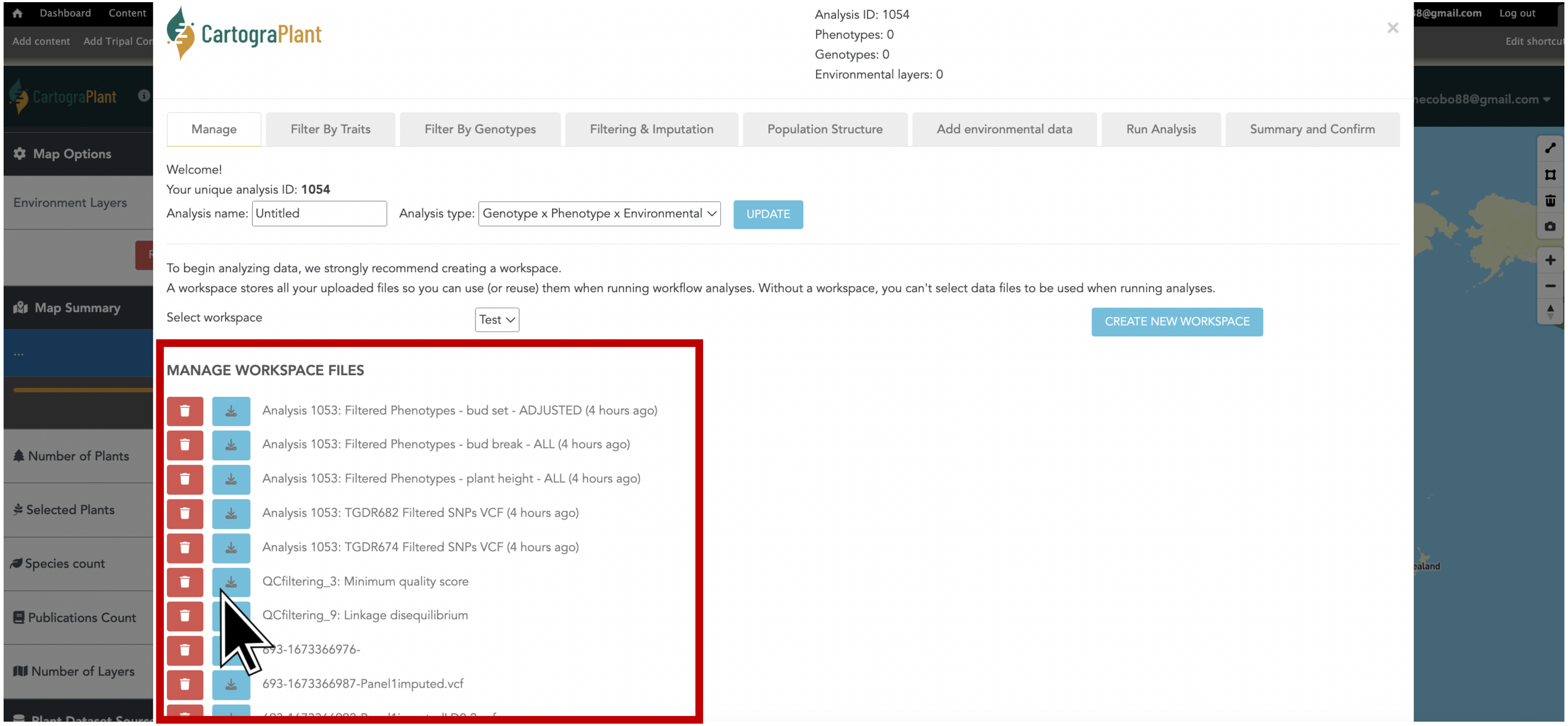
Completed File download: GWASwithEMMAX_Step4: REML output (77 bytes)
Completed File download: GWASwithEMMAX_Step4: 1. SNP ID, 2.Beta (1 is effect allele) 3. SE(beta) 4. p-value. (278787 bytes)
Completed File download: GWASwithEMMAX_Step3 Kinship Matrix (2225495 bytes)
Completed File download: GWASwithEMMAX_Step2: tFAM file (10175 bytes)
Completed File download: GWASwithEMMAX_Step2: tPED file (10112491 bytes)
Completed File download: GWASwithEMMAX_Step2: tFAM file (6919 bytes)
Completed File download: GWASwithEMMAX_Step2: tPED file (10109870 bytes)
Completed File download: GWASwithEMMAX_Step1: MAP file (167039 bytes)
Completed File download: GWASwithEMMAX_Step1: LOG file (0 bytes)
Completed File download: GWASwithEMMAX_Step1: PED file (11484601 bytes)
Completed File download: GWASwithEMMAX_Step1: MAP file (164418 bytes)
Completed File download: GWASwithEMMAX_Step1: LOG file (0 bytes)
Completed File download: GWASwithEMMAX_Step1: PED file (11481345 bytes)

DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS

DATA ANALYSIS

https://cartograplant.org/

https://gitlab.com/IreneCobo/cartograplantworkshop2022/



Analysis ID: 1054
Phenotypes: 0
Genotypes: 0
Environmental layers: 0

Welcome!
Your unique analysis ID: **1054**
Analysis name: Untitled
Analysis type: Genotype x Phenotype x Environmental

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

MANAGE WORKSPACE FILES

- Analysis 1053: Filtered Phenotypes - bud set - ADJUSTED (4 hours ago)
- Analysis 1053: Filtered Phenotypes - bud break - ALL (4 hours ago)
- Analysis 1053: Filtered Phenotypes - plant height - ALL (4 hours ago)
- Analysis 1053: TGDR682 Filtered SNPs VCF (4 hours ago)
- Analysis 1053: TGDR674 Filtered SNPs VCF (4 hours ago)
- QCfiltering_3: Minimum quality score
- QCfiltering_9: Linkage disequilibrium
- 693-1673366976-
- 693-1673366987-Panel1imputed.vcf

DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS

TEAM MEMBERS! QUESTIONS?

Members of the project

WASHINGTON STATE UNIVERSITY

- Stephen P. Ficklin

ECU

- Nic Herndon

UCONN
UNIVERSITY OF CONNECTICUT

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

THE UNIVERSITY OF TENNESSEE
KNOXVILLE

- Margaret Staton
- Abdullah Almsaeed

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

@Irene_Cobo
@TreeGenes

TreeGenes Tripal GMOD AgBioData DBTREES GTTN Galaxy USDA CartograPlant