

# 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.

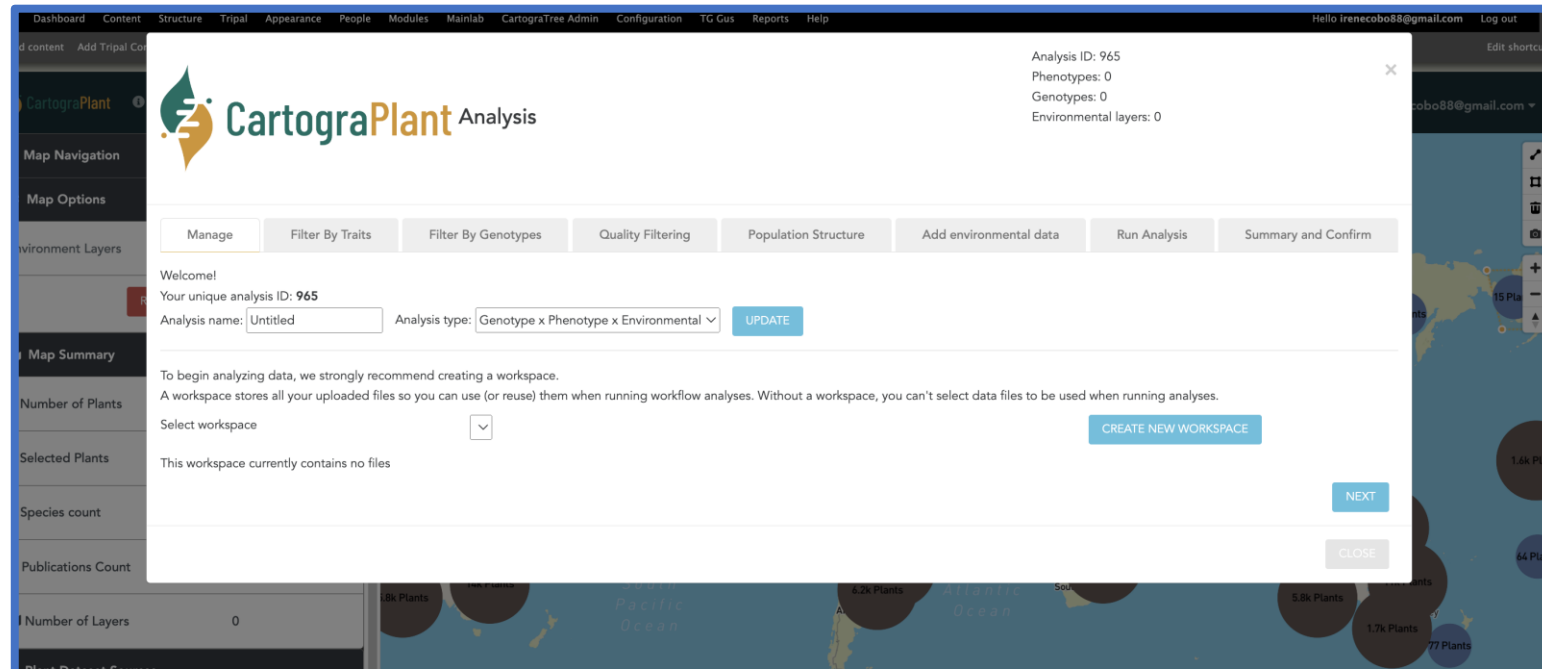
MSc. degree in biology student

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**



The screenshot displays the CartograPlant Analysis web interface. At the top, a navigation menu includes Dashboard, Content, Structure, Tripal, Appearance, People, Modules, Mainlab, CartograTree Admin, Configuration, TG Gus, Reports, and Help. The user is logged in as 'Hello irenecobo88@gmail.com'. The main content area features the CartograPlant logo and a 'Welcome!' message. A unique analysis ID of 965 is shown. Below this, there is a form for 'Analysis name' (currently 'Untitled') and 'Analysis type' (set to 'Genotype x Phenotype x Environmental'). A 'CREATE NEW WORKSPACE' button is visible, along with a 'NEXT' button. A 'CLOSE' button is also present. The interface includes a sidebar with 'Map Navigation', 'Map Options', 'Environment Layers', 'Map Summary', 'Number of Plants', 'Selected Plants', 'Species count', 'Publications Count', and 'Number of Layers'. The background shows a world map with various plant distribution markers.

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

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

Welcome!  
Your unique analysis ID: 965  
Analysis name:  Analysis type:

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.**



The image shows a screenshot of the Galaxy TG/HWG web interface. The main content area displays the CartograPlant logo, which consists of a stylized green and orange leaf-like shape next to the text "CartograPlant". Below this is a large plus sign (+) and the Galaxy logo, which is a stylized "G" followed by the word "Galaxy". The interface includes a left sidebar with a "Tools" section containing various bioinformatics tools like "Get Data", "Send Data", and "Collection Operations". The top navigation bar includes "Analyze Data", "Workflow", "Visualize", "Shared Data", "Help", and "User". The right sidebar shows a "History" section with a search bar and a list of datasets.

# 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	

## RESPONSE VARIABLE

Variation in phenotype

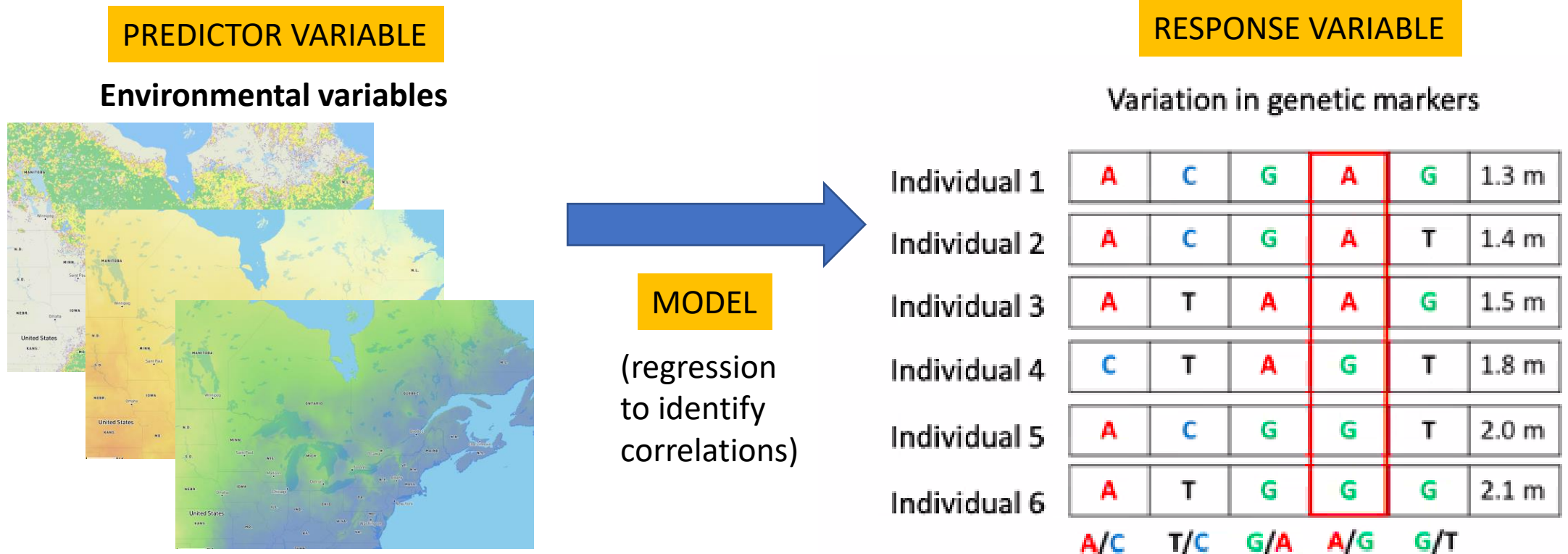


## MODEL

(regression to identify correlations)

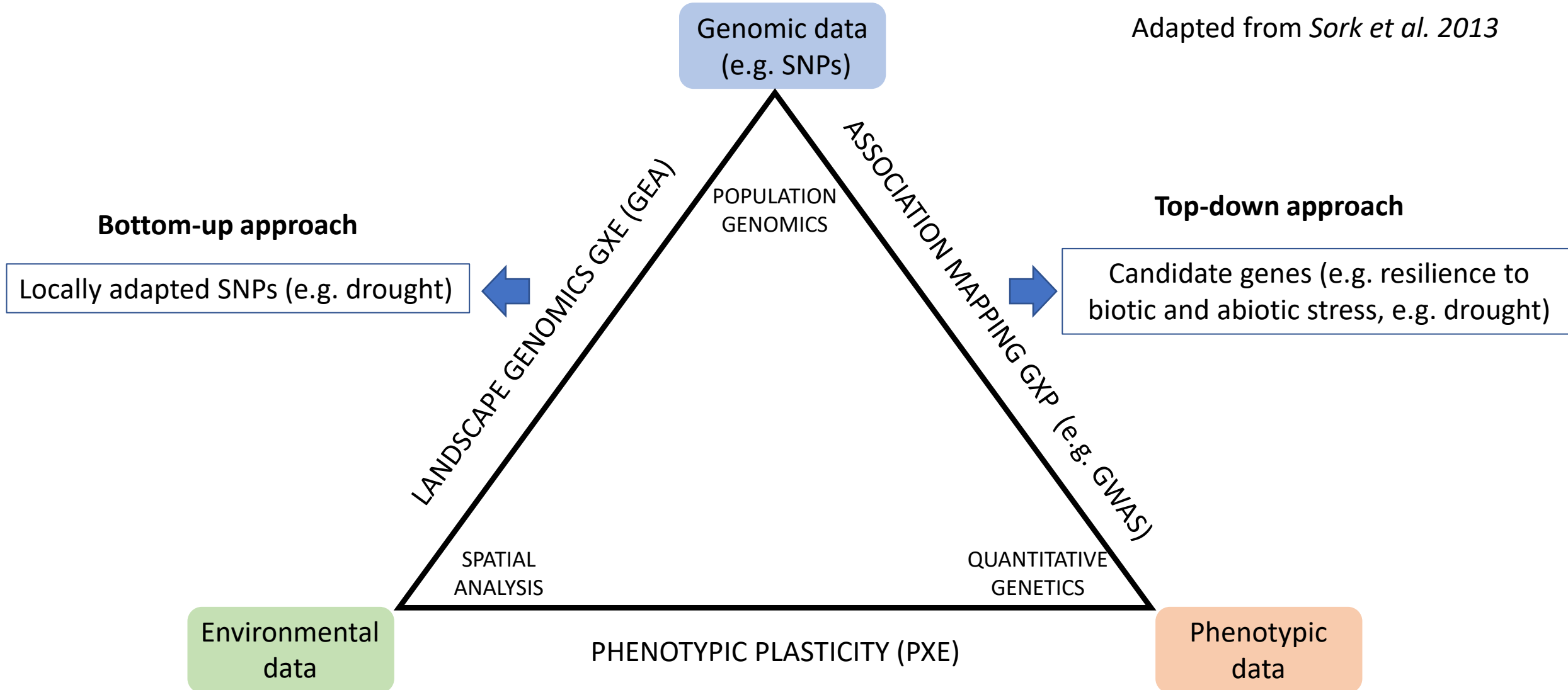
# WHAT IS LANDSCAPE GENOMICS?

- 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.



# USEFULNESS OF PERFORMING GXP AND GXE TOGETHER

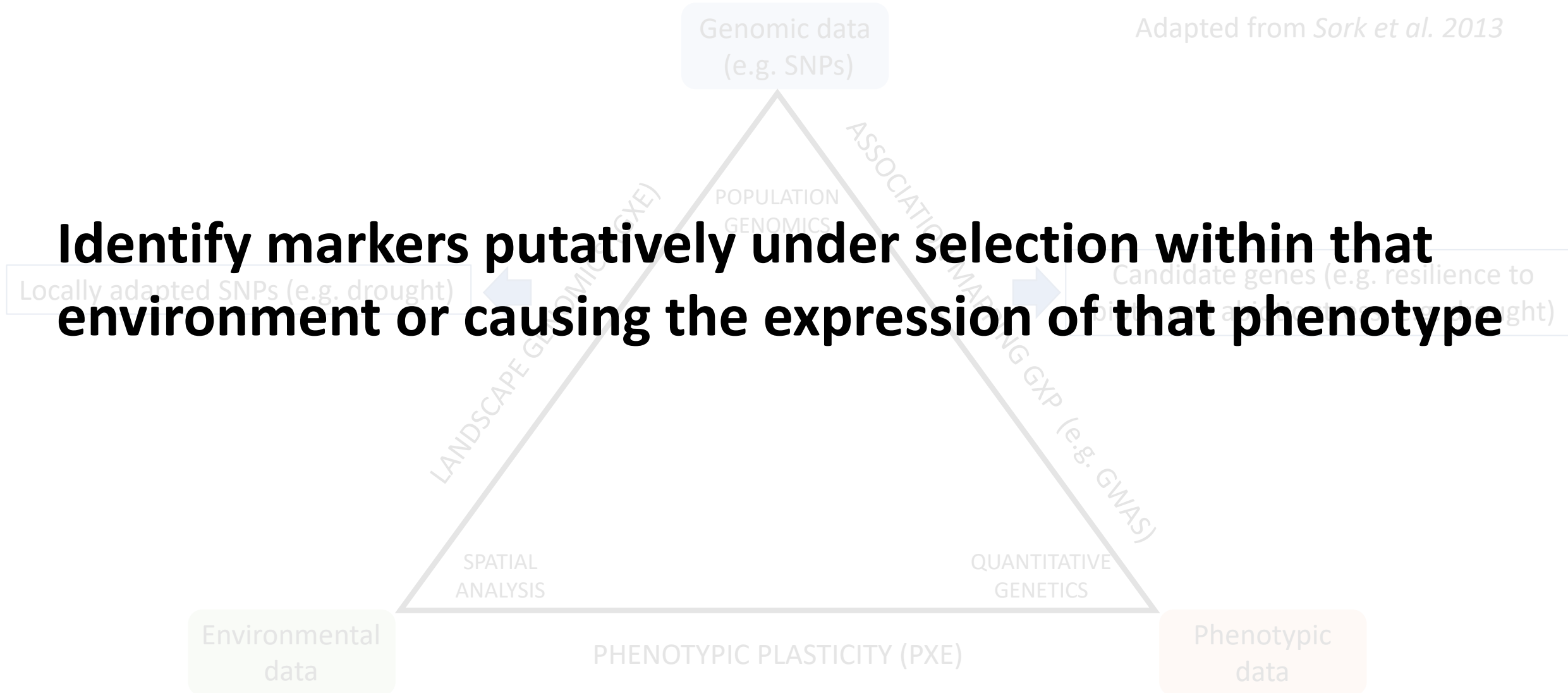
Adapted from *Sork et al. 2013*



# USEFULNESS OF PERFORMING GXP AND GXE TOGETHER

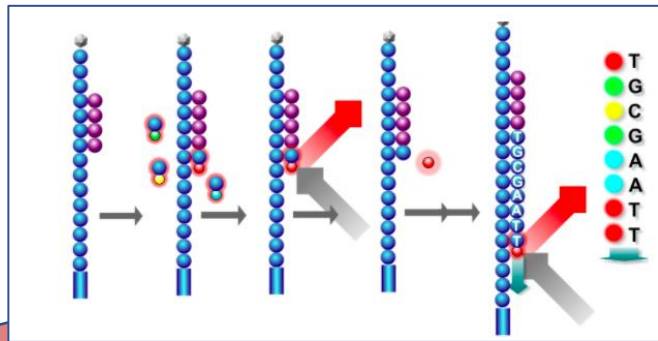
Adapted from *Sork et al. 2013*

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





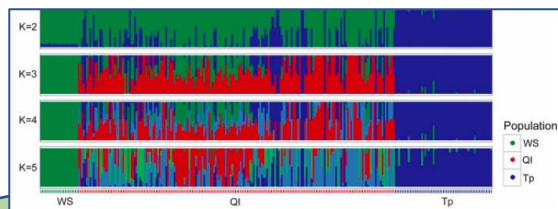
# FACTORS THAT CAN AFFECT THE ACCURACY OF THE GWAS/GEA RESULTS



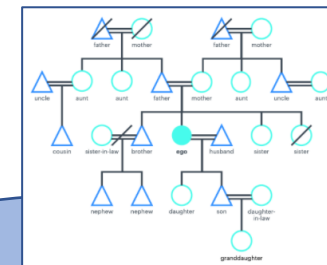
**GENOTYPING ERRORS**

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**

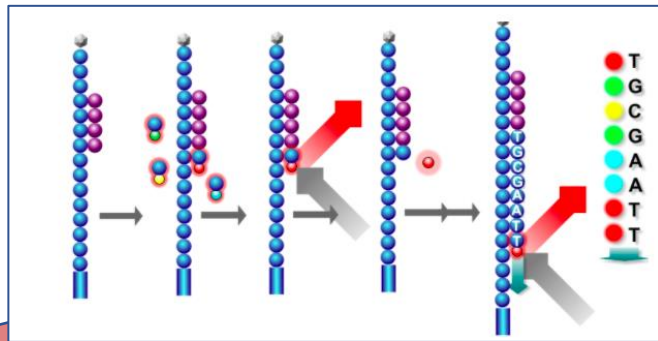


**POPULATION  
STRUCTURE/KINSHIP**



**KINSHIP**

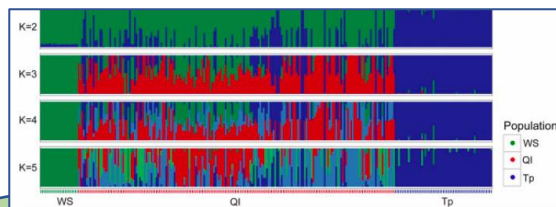
# FACTORS THAT CAN AFFECT THE ACCURACY OF THE GWAS/GEA RESULTS



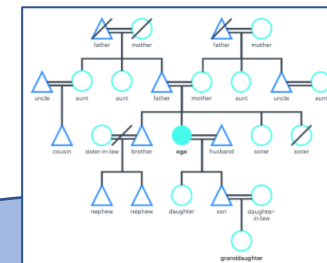
**GENOTYPING ERRORS**

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**



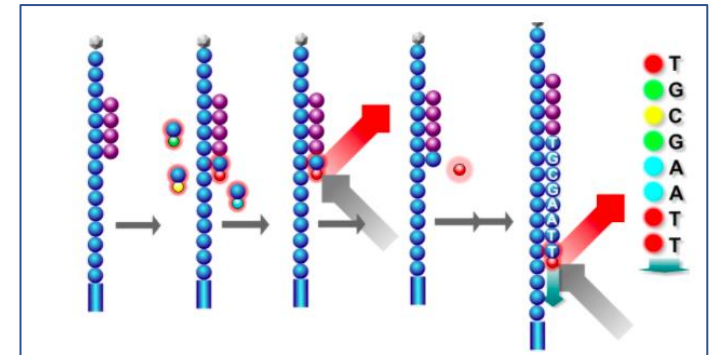
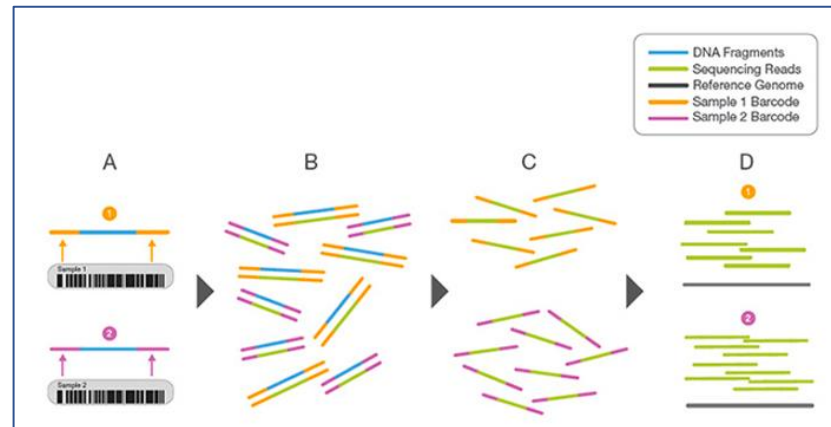
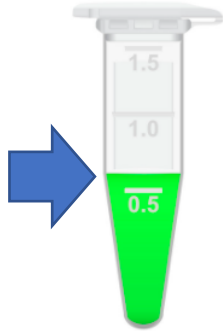
**POPULATION  
STRUCTURE/KINSHIP**



**KINSHIP**

# GENOTYPING ERRORS

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



SAMPLE  
PREPARATION

LIBRARY  
PREPARATION

SEQUENCING

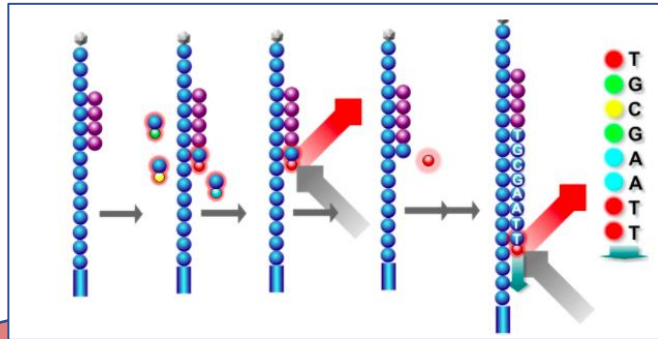
# GENOTYPING ERRORS

## 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						Filtering per-individual				
Individual 1	A	C	G	A	G		A	C	G	A	G
Individual 2	A	C	G	A	T		A	C	G	A	T
Individual 3	A	T	A	A	G		A	T	A	A	G
Individual 4	C	T	A	A	T		C	T	A	G	T
Individual 5	A	C	G	G	T		A	C	G	G	T
Individual 6	A	T	G	G	G		A	T	G	G	G
	A/C	T/C	G/A	A/G	G/T		A/C	T/C	G/A	A/G	G/T

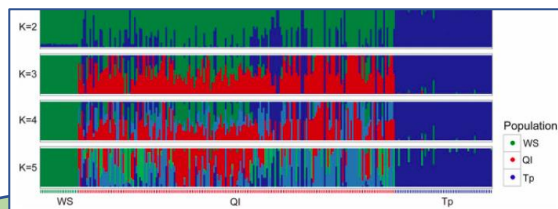
# FACTORS THAT CAN AFFECT THE ACCURACY OF THE GWAS/GEA RESULTS



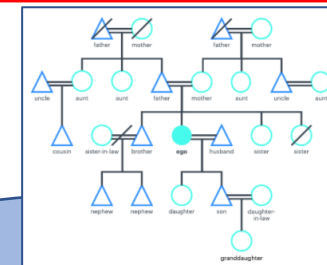
**GENOTYPING ERRORS**

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

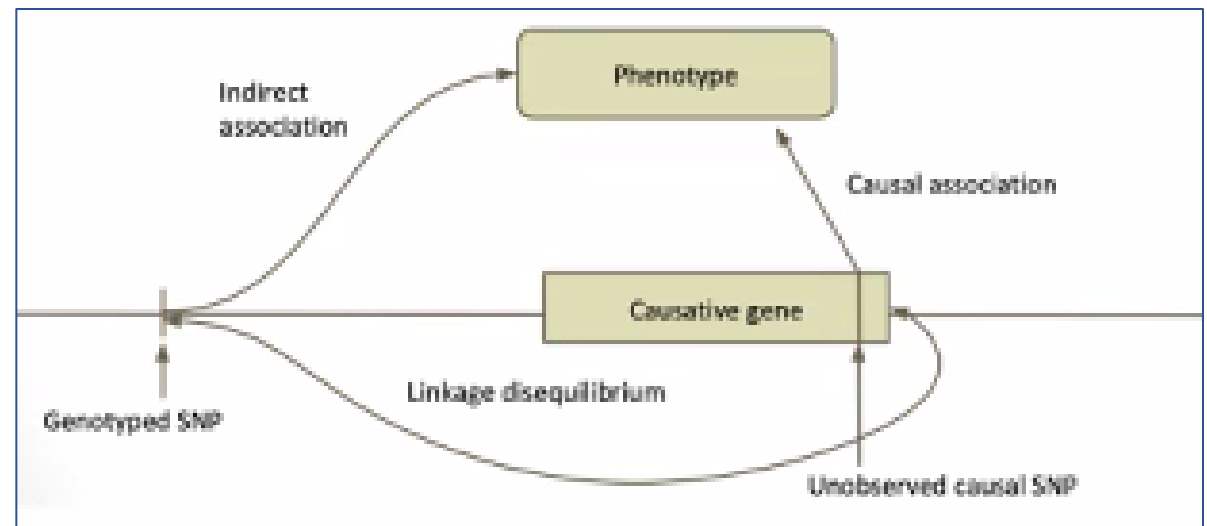
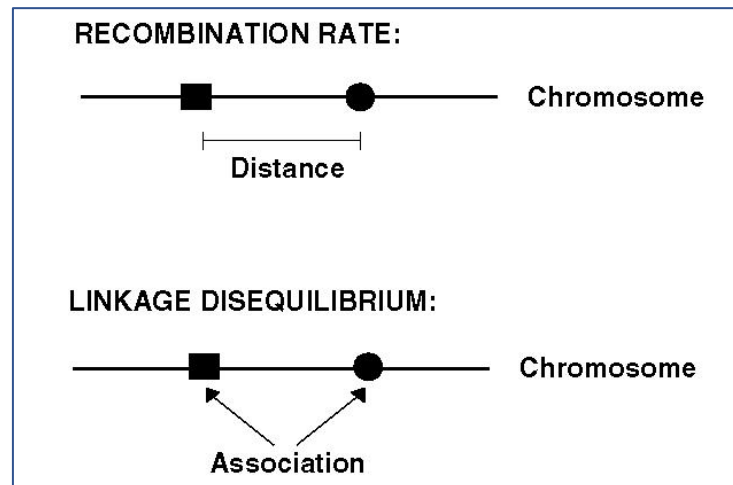
### Imputation of lacking SNPs

- 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**

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	
----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----															
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
----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----															
Reference haplotypes															
Study genotypes															

# 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

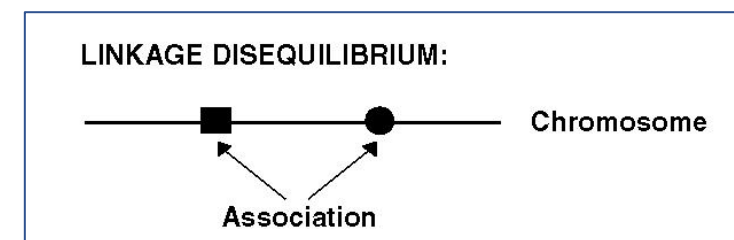
- **Tests** for different combinations of SNP data **quality filtering thresholds** (aka “Cases”).
- **Objective**: Maximize the quantity and quality of the resulting SNPs, while maintaining **accuracy**.
- **Imputation for non-model organisms**
  - No requires ordered markers or a reference panel
 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

### CASES

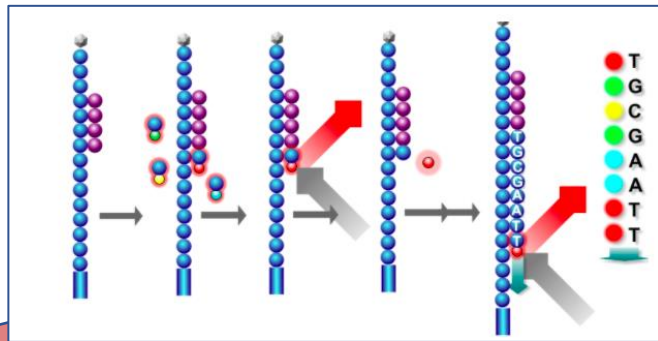
Name	Samples	Positions	Accuracy	Correlation	Filters Additional	Depth
Case 1	100	343652	0.9684	0.9428	PositionMiss(0.7),SampleMiss(0.7),MAF(0.01)	Depth(2)
Case 2	100	422531	0.9669	0.9425	PositionMiss(0.8),SampleMiss(0.8),MAF(0.01)	Depth(2)
Case 3	100	930660	0.9558	0.9189	PositionMiss(0.9),SampleMiss(0.9),MAF(0.01)	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



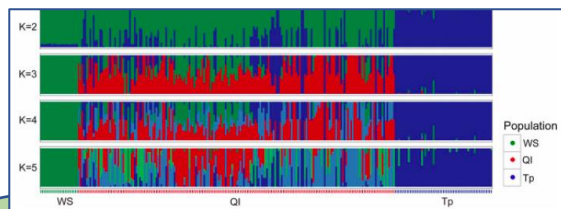
# FACTORS THAT CAN AFFECT THE ACCURACY OF THE GWAS/GEA RESULTS



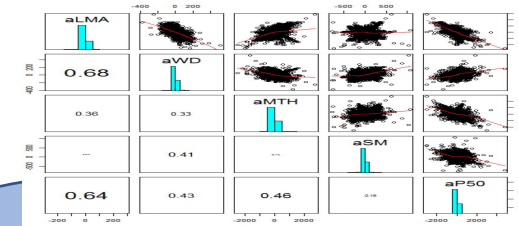
**GENOTYPING ERRORS**

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**

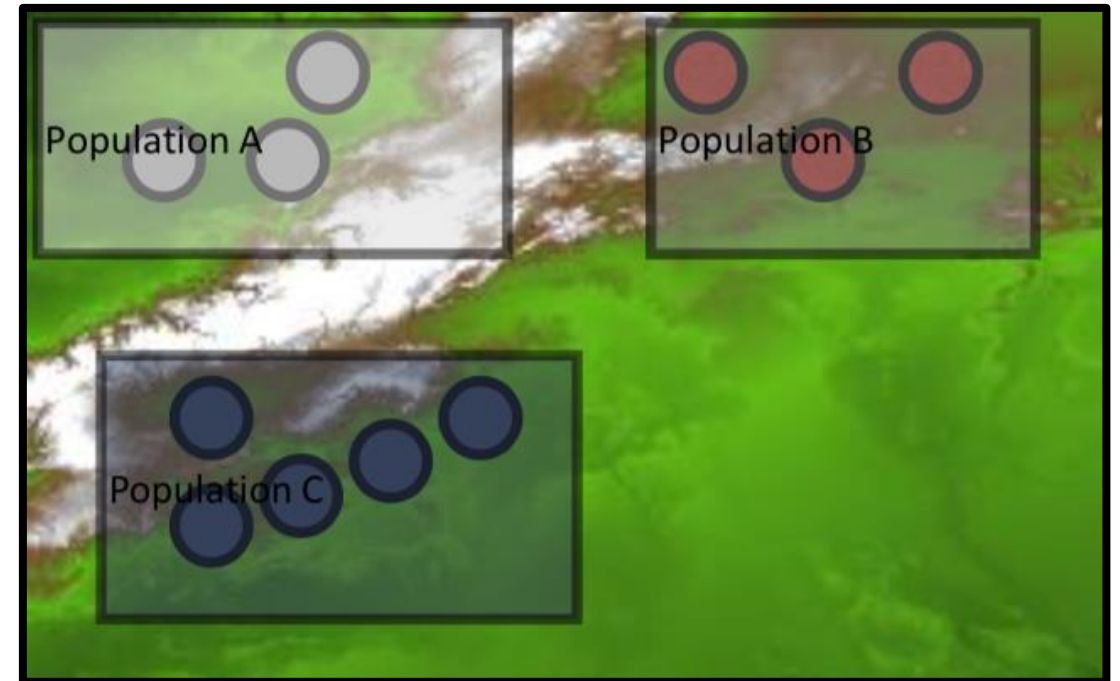


**MULTICOLLINEARITY**

# POPULATION STRUCTURE/KINSHIP CORRECTION

## SOLUTION:

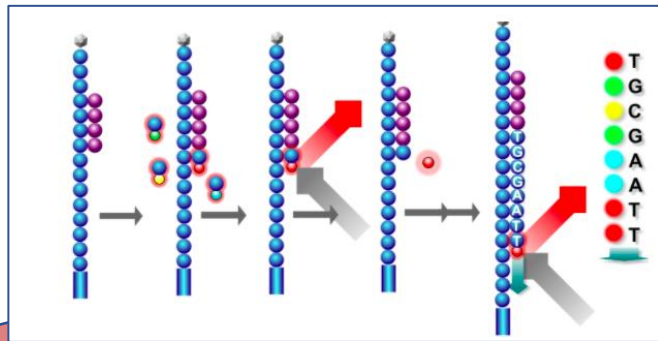
- Account for **confounding variables** in the model
- **Objective of GWAS/GEA:** To identify **alleles associated with the traits/environmental variables of interest.**
- Reduce **spurious associations or false positives** in GWAS/GEA
  - **Population structure:** patterns caused by by isolation by 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

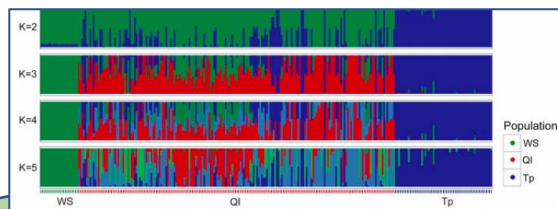
# FACTORS THAT CAN AFFECT THE ACCURACY OF THE GWAS/GEA RESULTS



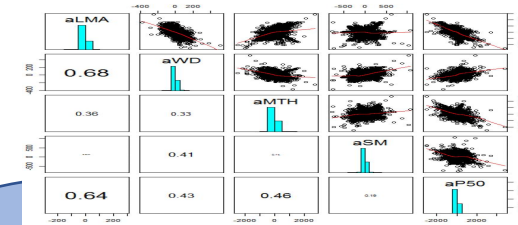
**GENOTYPING ERRORS**

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**



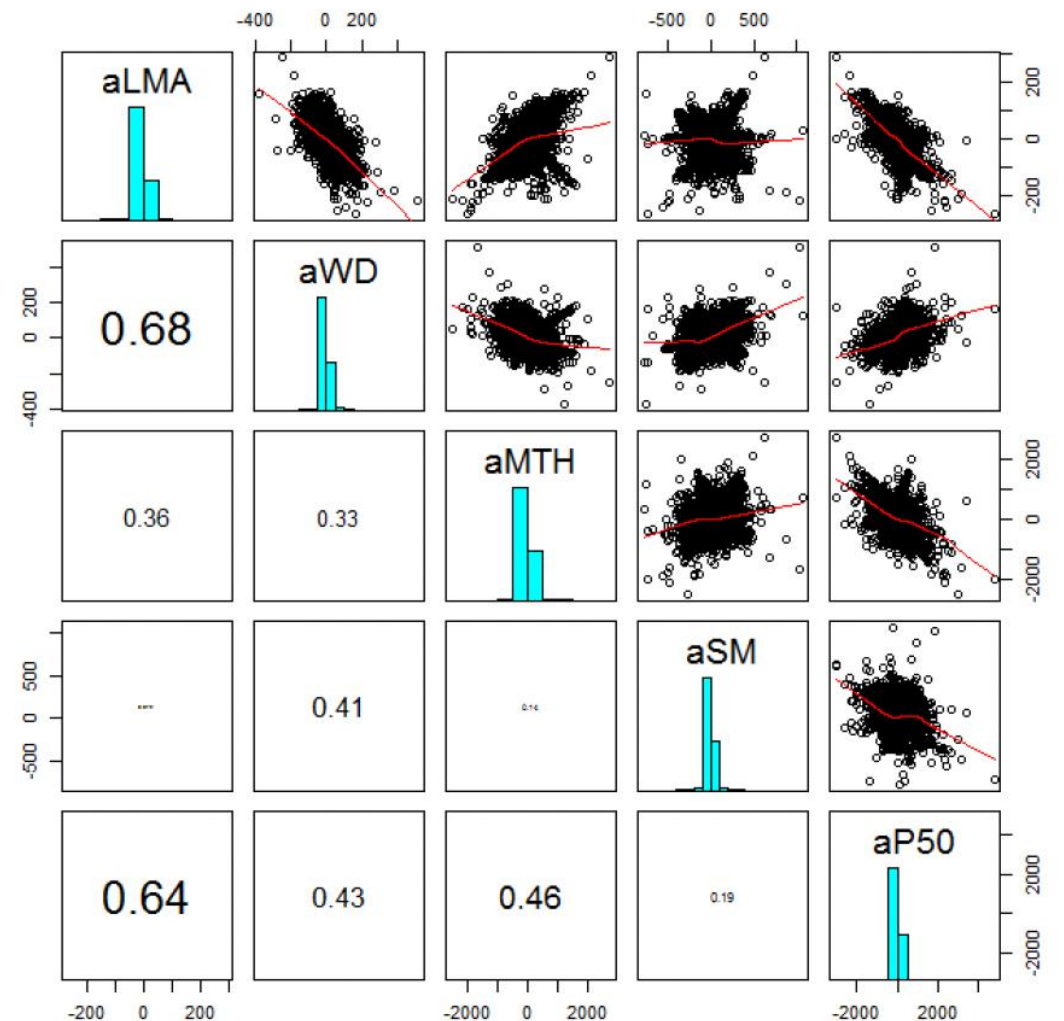
**MULTICOLLINEARITY**

# OTHER GxE ISSUES: MULTICOLLINEARITY

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

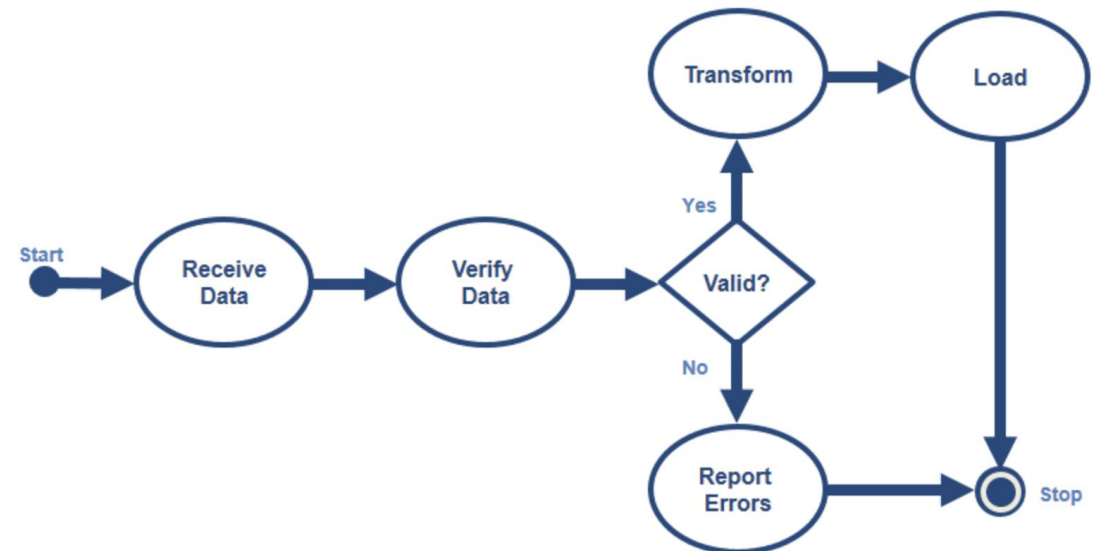
## SOLUTION

- **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**

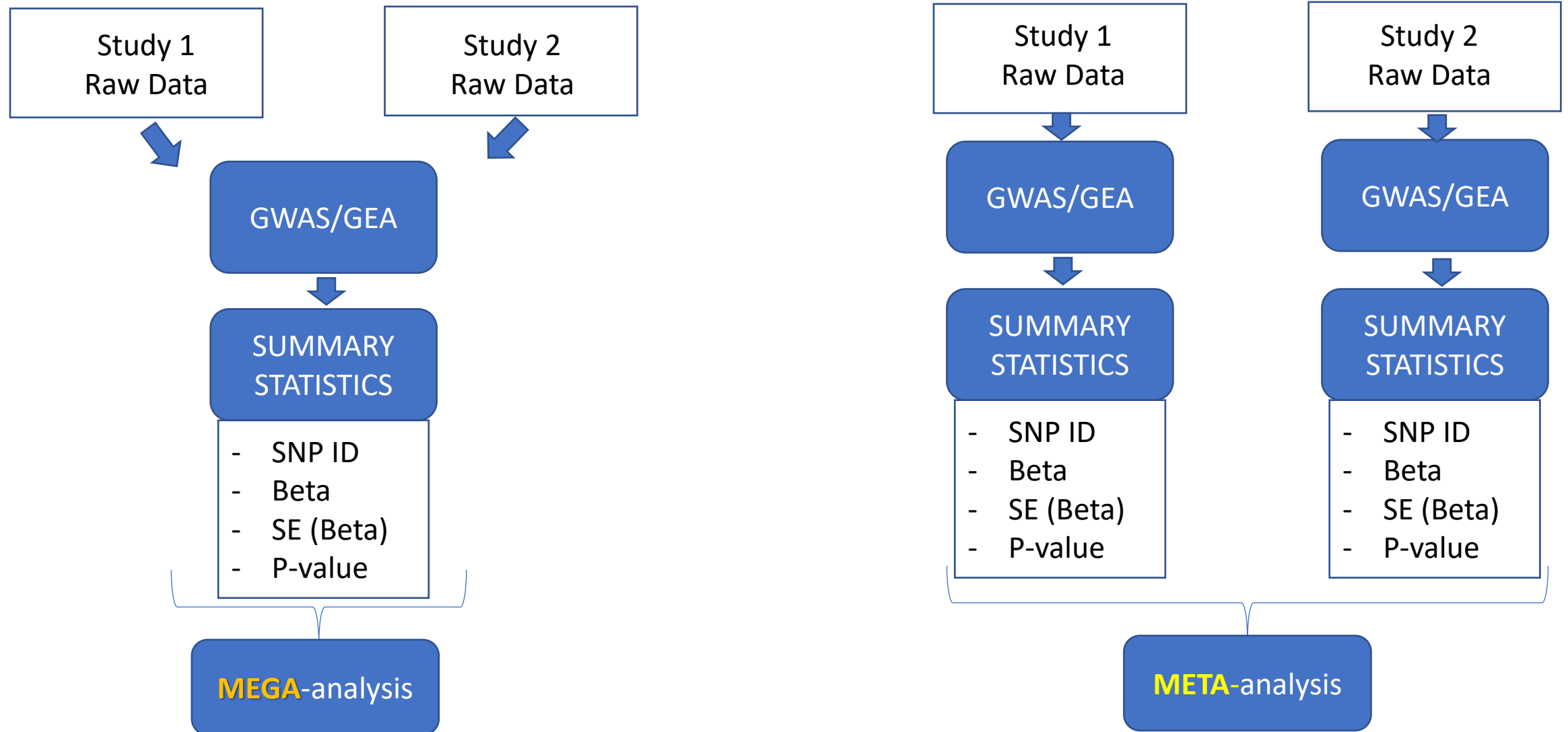


# CARTOGRAPLANT ANALYTIC WORKFLOWS OBJECTIVES

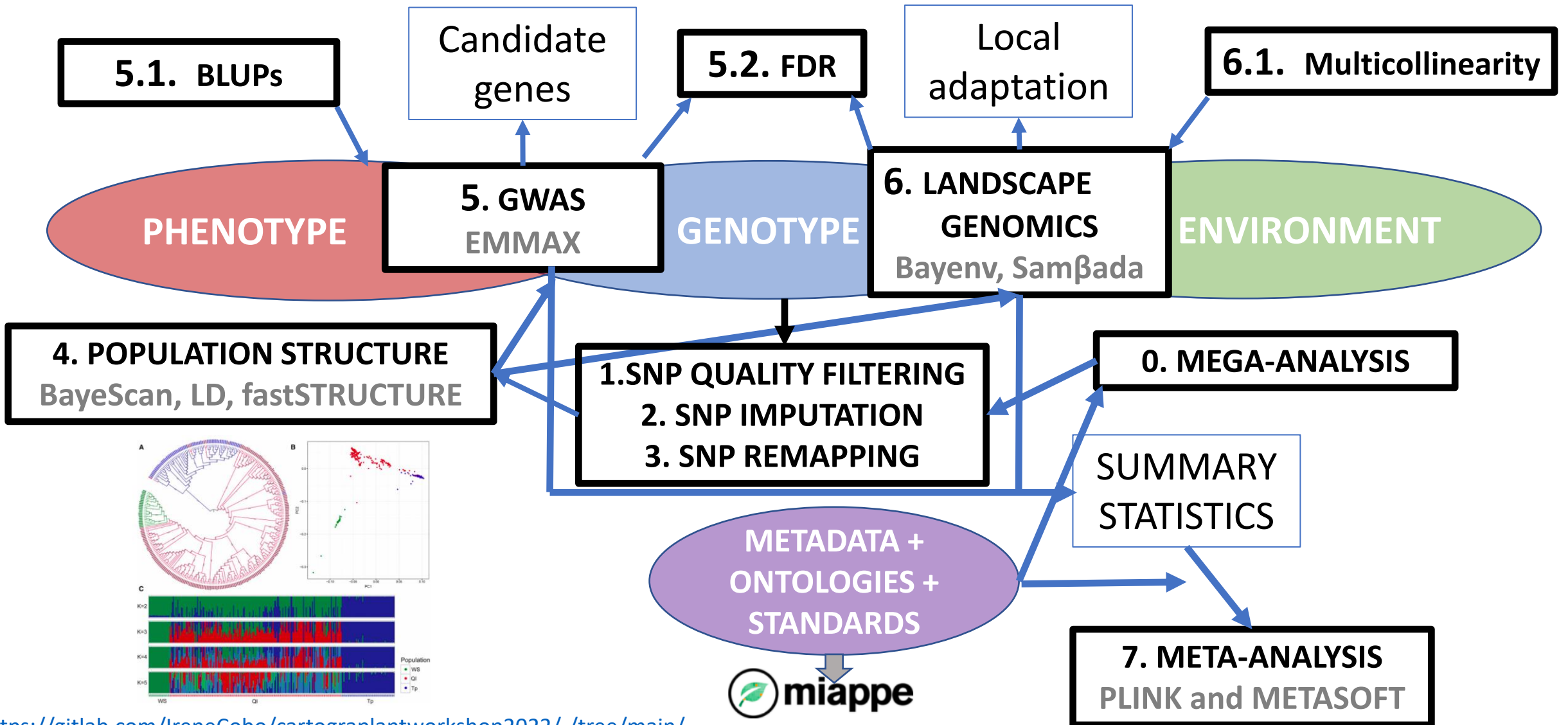
- Genome Wide Association Analysis (GWAS, G X P)
- Landscape genomics (G X E)
- Mega/meta-analysis



# WHAT ARE META-ANALYSIS AND MEGA-ANALYSIS?

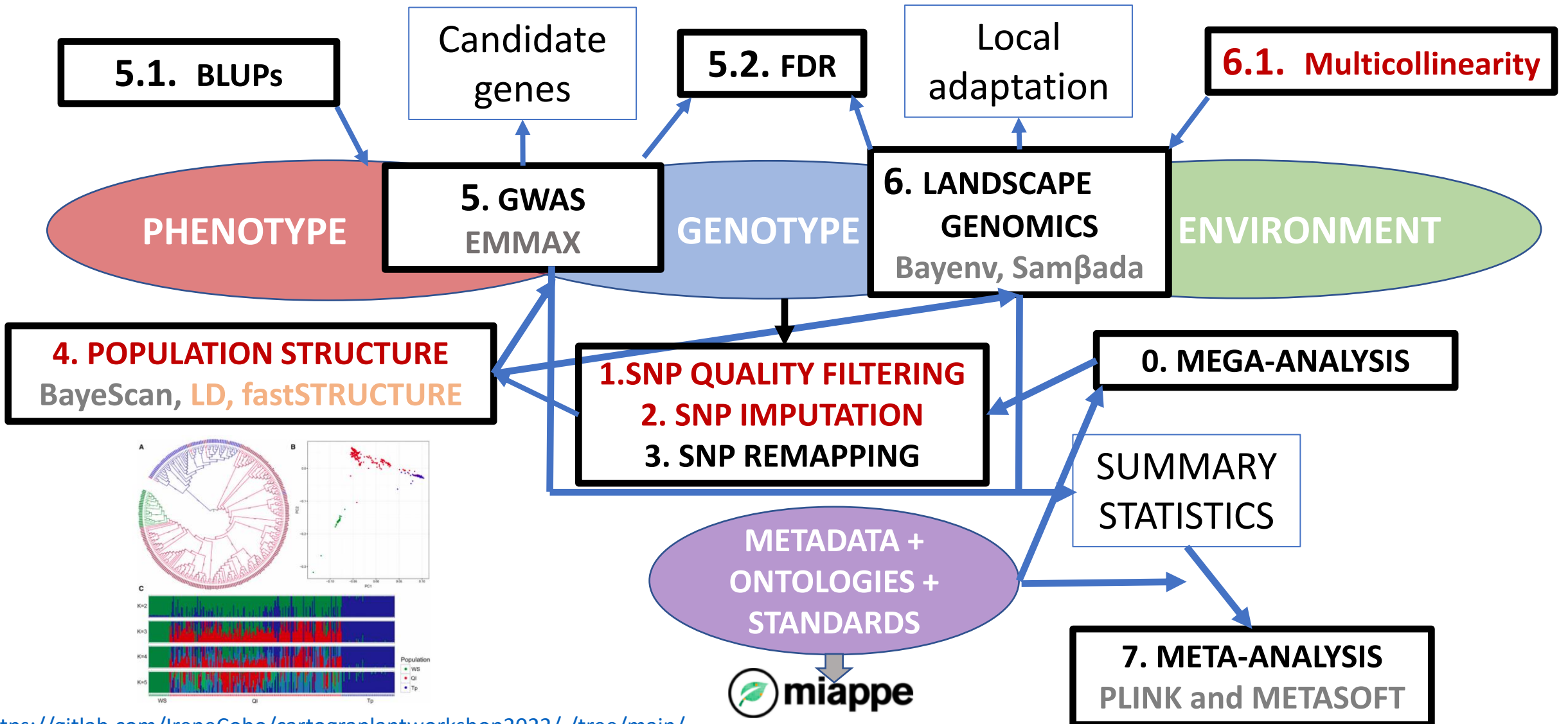


# CARTOGRAPLANT WORKFLOWS OVERVIEW



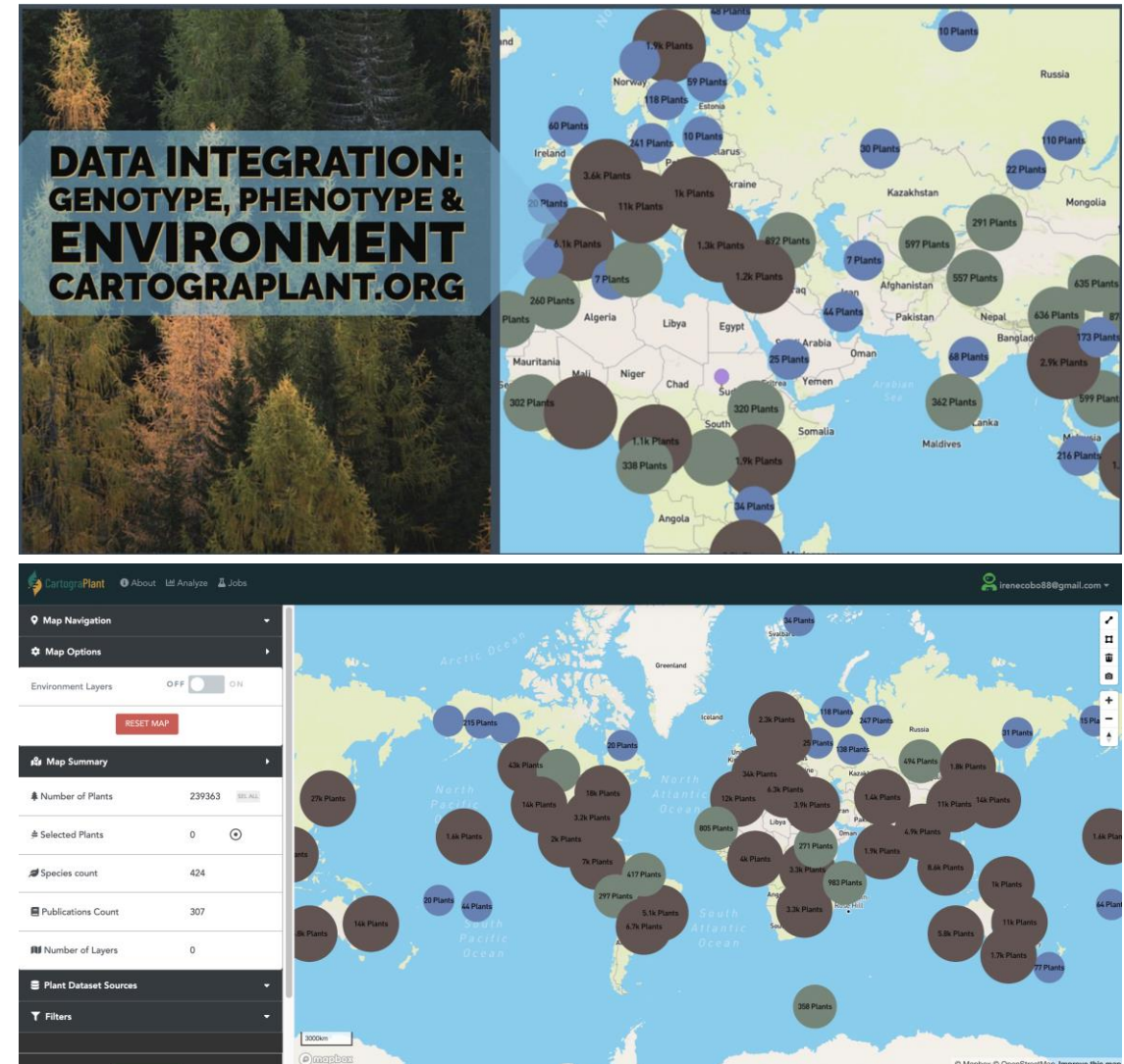


# CARTOGRAPLANT WORKFLOWS OVERVIEW



# CONCLUSIONS

- Cartogrplant 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



# 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

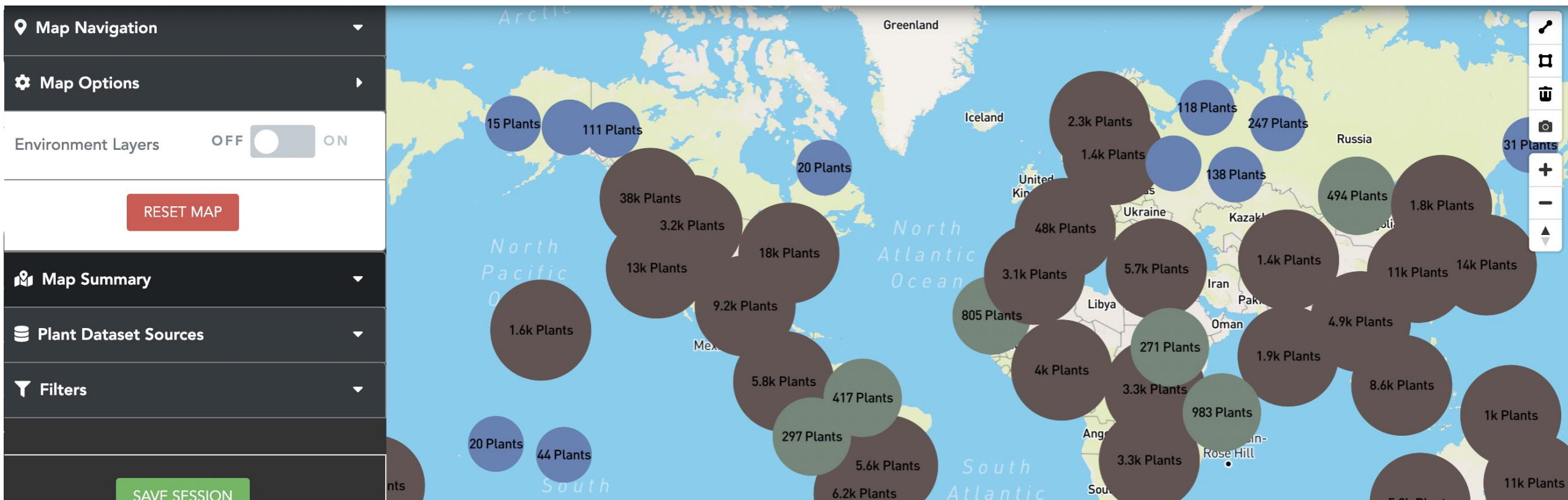
12:15-12:35 Behind the Scenes of CartograPlant - Environmental Layers and Data - Risharde Ramnath

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

# 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

# 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

Submit data for analysis



User Guide

Detailed documentation, tutorials and resources

<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**

The screenshot displays the CartograPlant web application interface. The top navigation bar includes the logo, 'About', 'Analyze', and 'Jobs' links, and a user profile for 'ireneCobo88@gmail.com'. The interface is divided into several panels:

- Map Navigation:** Features input fields for 'Lat' and 'Lon', a 'GO' button, and a red box highlighting these elements.
- Map Options:** Includes a toggle for 'Environment Layers' (currently ON) and a 'RESET MAP' button.
- Map Summary:** Provides a summary of data: Number of Plants (243623), Selected Plants (0), Species count (421), Publications Count (324), and Number of Layers (1).
- Environmental Layers:** A list of layers with checkboxes and sliders. The 'Forest Fragmentation Risk (Contiguous US, 2002)' layer is checked and highlighted with a red box. A blue arrow points from this layer to the data table.
- Environmental Data Table:** A table showing data for the selected layer: ENVIRONMENTAL LAYER: FOREST FRAGMENTATION RISK (CONTIGUOUS US, 2002).

Band	Count
RED BAND	0
GREEN BAND	185
BLUE BAND	92
ALPHA BAND	255
- Map:** A map of North America with numerous circular markers of varying sizes and colors (blue, green, brown) representing plant data points. A mouse cursor is hovering over a point labeled '26 Plants'. A red box highlights the zoom controls on the right side of the map.

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)

# INTERFACE AND DATA VISUALIZATION

The screenshot displays the CartograPlant web application interface. On the left, a sidebar contains navigation and map controls, including a 'Map Summary' section with the following data:

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

The main map area shows a geographical view of the United States with plant distribution data. Callouts identify several key UI features:

- Tree icon:** A green tree icon on the map.
- Plant icon:** A green plant icon on the map.
- TreeSnap icon:** A purple tree icon on the map.
- Polygon selecting plants:** A blue polygon on the map.
- Polygon drawing button:** A red-bordered button in the top right corner of the map area.
- Study info window:** A detailed window for study TGDR300-LK1419, showing associated study information and a table of plant genotypic data.
- Study info window (right):** A window for study TGDR300-LK1372, showing coordinates, elevation, and a photo of the plant.

The 'Study info window' for TGDR300-LK1419 includes the following data:

**Study Associated**  
A rangewide herbarium-derived dataset indicates high levels of gene flow in black cherry (*Prunus serotina*)  
Konrado, 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 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/>

The screenshot displays the CartograPlant web application interface. On the left, a sidebar contains a 'Map Summary' section with the following statistics:

- Number of Plants: 243623 (with a 'SEL ALL' button)
- Selected Plants: 0 (with a selection icon)
- Species count: 421
- Publications Count: 324
- Number of Layers: 0

Below the summary is the 'Plant Dataset Sources' section, which is highlighted with a red border. It includes five toggle switches, all of which are currently turned 'ON':

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

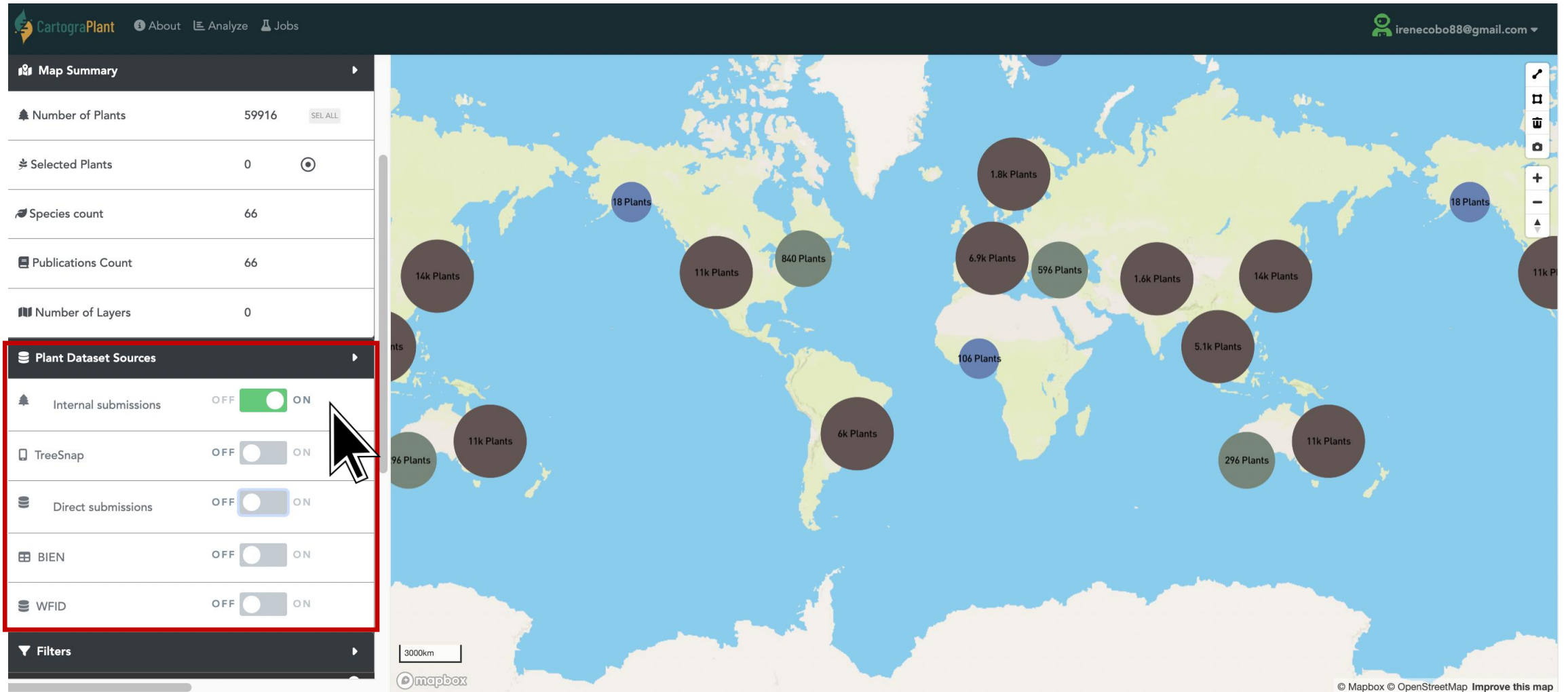
At the bottom of the sidebar is a 'Filters' section. The main area of the interface is a world map where data points are represented by circles of varying sizes and colors (blue, green, brown). Each circle is labeled with the number of plants it represents, such as '31 Plants', '58k Plants', '983 Plants', and '1.1k Plants'. The map includes a scale bar for 3000km and a 'mapbox' logo. The user's email address, 'ireneCobo88@gmail.com', is visible in the top right corner.

## DATA SELECTION FOR ANALYSIS

# INTERFACE AND DATA VISUALIZATION

<https://cartograplant.org/>

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



## DATA SELECTION FOR ANALYSIS

# INTERFACE AND DATA VISUALIZATION

<https://cartograplant.org/>

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

The screenshot displays the CartograPlant web interface. On the left, a sidebar contains a 'Map Summary' section with a red border. This section includes a table of statistics and a 'UNSEL ALL' button. Below this is the 'Plant Dataset Sources' section with toggle switches for various data sources. At the bottom of the sidebar is a 'Filters' section. The main area of the interface is a world map where data points are represented by circles of varying sizes and colors, each labeled with the number of plants (e.g., '1.8k Plants', '11k Plants', '840 Plants'). A mouse cursor is pointing at the 'UNSEL ALL' button in the sidebar. The top navigation bar includes the 'CartograPlant' logo and links for 'About', 'Analyze', and 'Jobs'. The user's email address 'irencobo88@gmail.com' is visible in the top right corner. A scale bar for 3000km and the Mapbox logo are located at the bottom left of the map area. The bottom right corner of the map area contains the text '© Mapbox © OpenStreetMap Improve this map'.

Map Summary	Value	Action
Number of Plants	59916	<a href="#">UNSEL ALL</a>
Selected Plants	59916	<input checked="" type="radio"/>
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**

<https://treegenesdb.org/cartogratre#tree-filter-options>

## DATA SELECTION FOR ANALYSIS

# INTERFACE AND DATA VISUALIZATION

<https://cartogrplant.org/>

<https://gitlab.com/IreneCobo/cartogrplantworkshop2022/>

The screenshot displays the CartograPlant web interface. On the left, there are control panels for 'TreeSnap', 'Direct submissions', 'BIEN', and 'WFID', each with an 'OFF' and 'ON' toggle. Below these is a 'Filters' section with a red border around the active filter rule: 'Family' (selected), 'equal', and 'Anacardiaceae'. Buttons for 'APPLY FILTER', 'RESET FILTER', and 'SAVE SESSION' are visible. A dropdown menu is open, listing various filter criteria: Family (checked), Genus, Species, Markers, Plant Structure, Phenotype Attribute, Study Title, Study First Author, and Study Accession (highlighted with a mouse cursor). The main area is a world map with circular markers of varying sizes and colors (brown, green, blue) representing plant data points, each labeled with a count (e.g., '11k Plants', '1.8k Plants', '840 Plants'). The top navigation bar includes 'CartograPlant', 'About', 'Analyze', 'Jobs', and a user profile 'ireneCobo88@gmail.com'. A scale bar for 3000km and a 'mapbox' logo are at the bottom left, and a copyright notice '© Mapbox © OpenStreetMap Improve this map' is at the bottom right.

## DATA SELECTION FOR ANALYSIS



# INTERFACE AND DATA VISUALIZATION

<https://cartograplant.org/>

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

The screenshot displays the CartograPlant web application interface. On the left is a sidebar with various settings and filters. At the top of the sidebar, there are four toggle switches: 'TreeSnap' (ON), 'Direct submissions' (ON), 'BIEN' (OFF), and 'WFID' (OFF). Below these is a 'Filters' section with 'AND' and 'OR' options. A red circle highlights the 'ADD RULE' and 'ADD GROUP' buttons. The filter rules section shows a dropdown for 'Study Accession' set to 'equal' and a text input containing 'TGDR674'. A red box highlights a filter rule: '1 phenotypes found in study: frost free days'. At the bottom of the sidebar are buttons for 'APPLY FILTER', 'RESET FILTER', and 'SAVE SESSION'. The main area is a world map with numerous circular data points of varying sizes and colors (blue, green, brown) representing plant data. Each point is labeled with the number of plants, such as '31 Plants', '117 Plants', '3.8k Plants', etc. The map includes a scale bar for 3000km and a 'mapbox' logo at the bottom left.

## DATA SELECTION FOR ANALYSIS



# INTERFACE AND DATA VISUALIZATION

<https://cartograplant.org/>

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

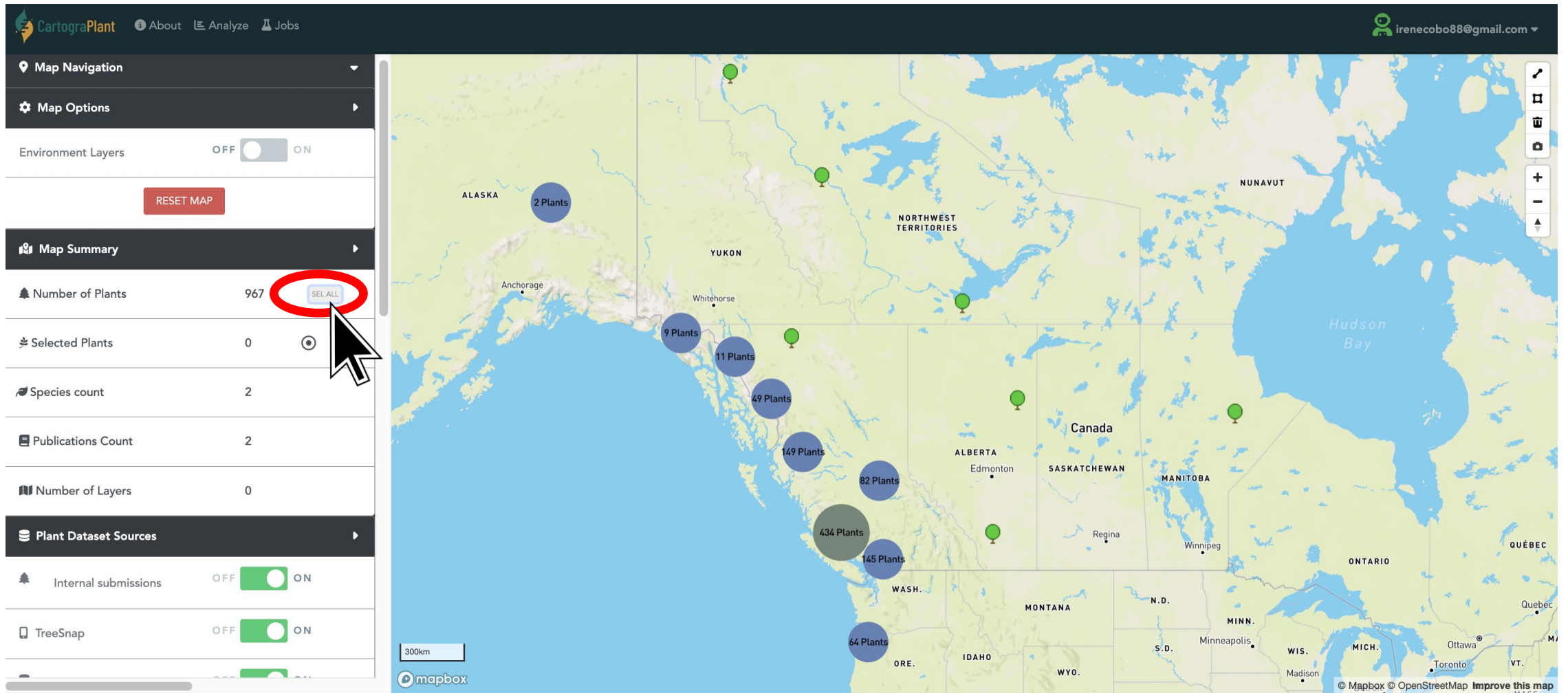
The screenshot displays the CartograPlant web application interface. On the left, a sidebar contains a filter configuration panel. The filter is set to 'AND' mode and includes two rules. The second rule, for 'Study Accession' 'TGDR682', is highlighted with a red border and lists 17 phenotypes: bud break, 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, and wood carbon 13 content. Below the filter rules, the 'APPLY FILTER' button is circled in red with a mouse cursor pointing to it. Other buttons include 'RESET FILTER' and 'SAVE SESSION'. The main area features a world map with numerous circular data points of varying sizes and colors (blue, green, brown) representing different plant studies. Each point is labeled with the number of plants, such as '31 Plants', '58k Plants', and '983 Plants'. The map includes a scale bar for 3000km and a Mapbox logo. The top navigation bar shows 'CartograPlant', 'About', 'Analyze', and 'Jobs' menus, along with a user profile for 'ireneCobo88@gmail.com'.

## DATA SELECTION FOR ANALYSIS

# INTERFACE AND DATA VISUALIZATION

<https://cartograplant.org/>

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

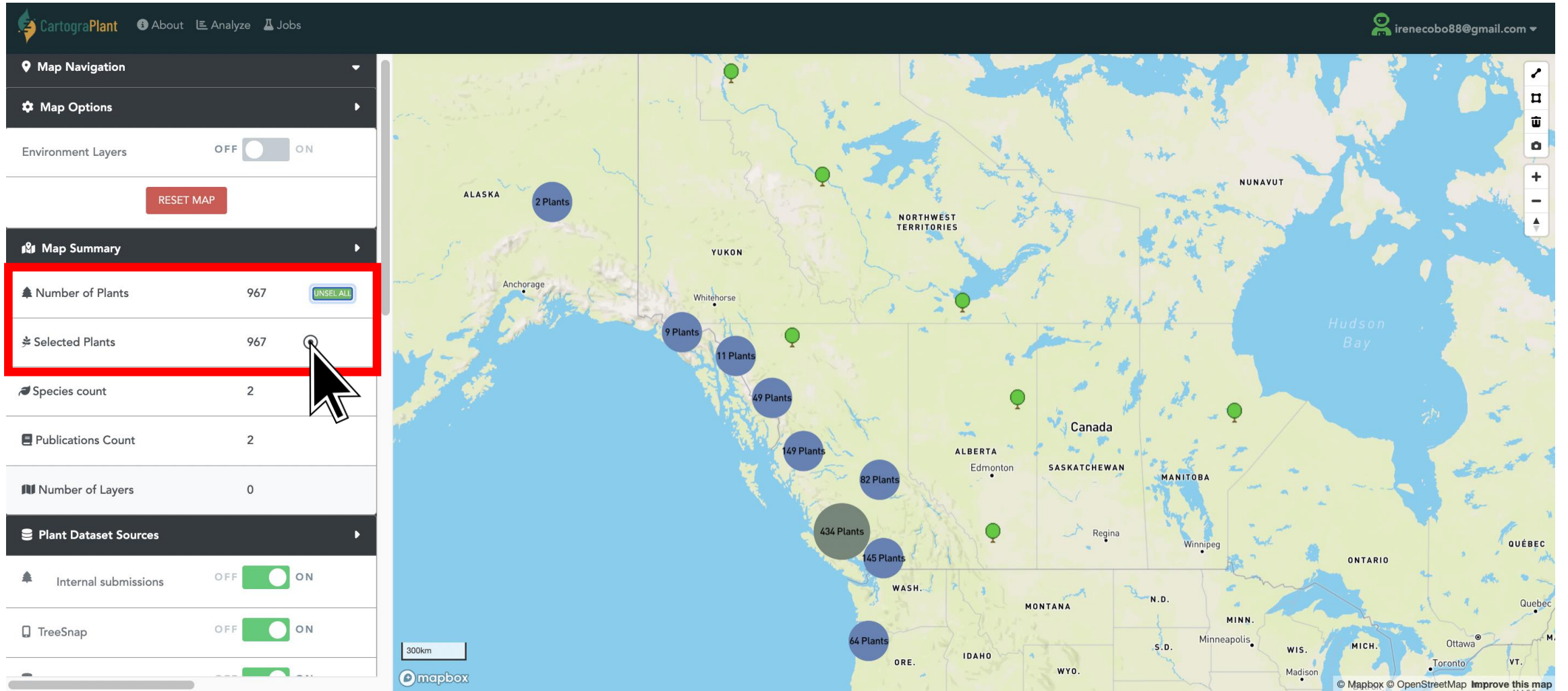


## DATA SELECTION FOR ANALYSIS

# INTERFACE AND DATA VISUALIZATION

<https://cartograplant.org/>

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

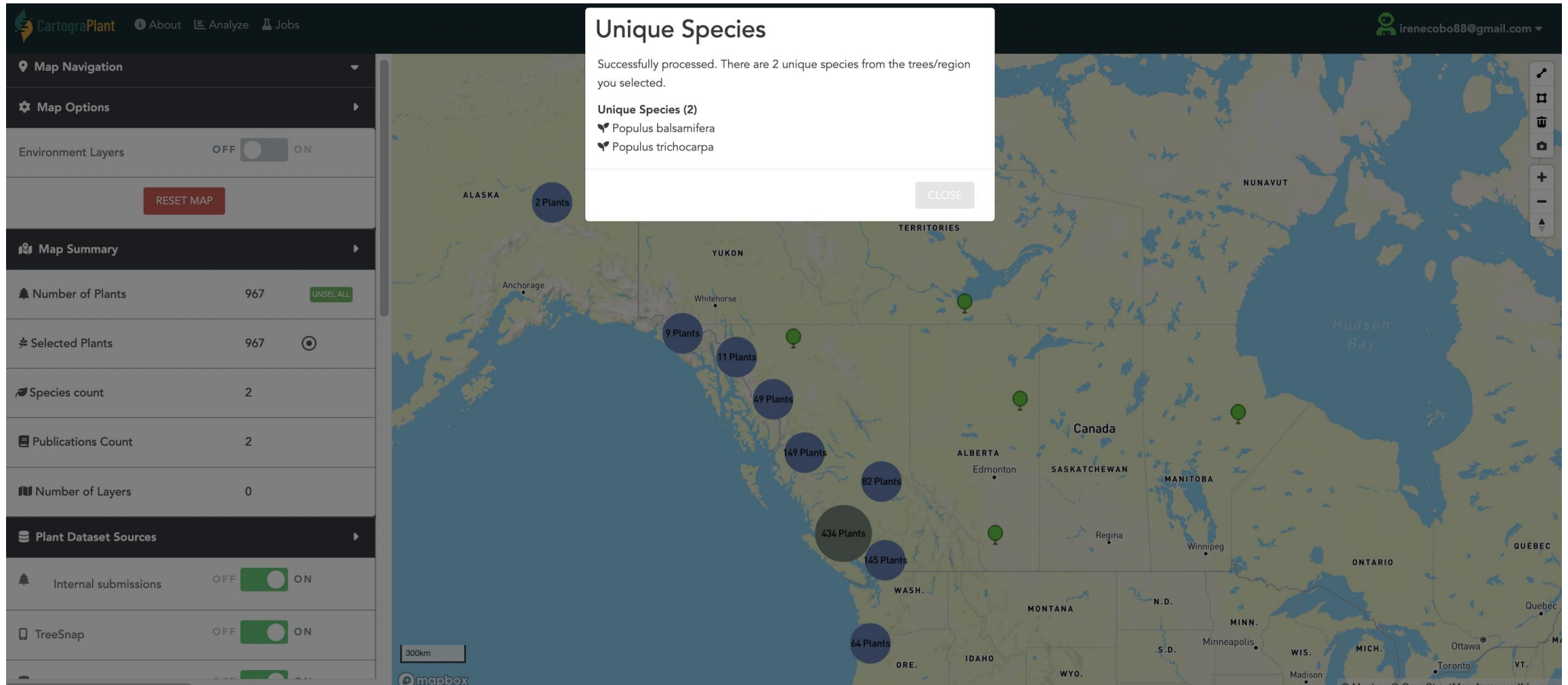


**DATA SELECTION FOR ANALYSIS**

# INTERFACE AND DATA VISUALIZATION

<https://cartograplant.org/>

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

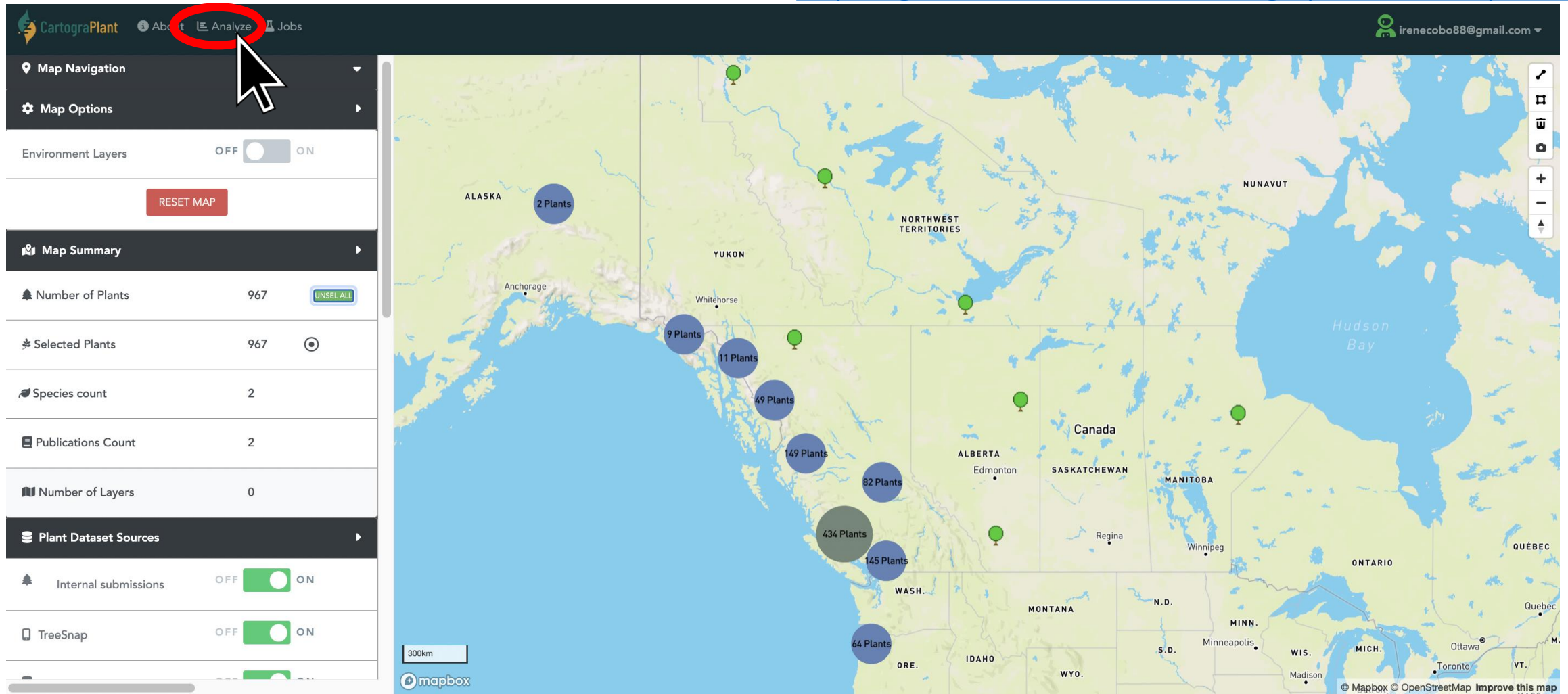


**DATA SELECTION FOR ANALYSIS**

# INTERFACE AND DATA VISUALIZATION

<https://cartograplant.org/>

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



## DATA SELECTION FOR ANALYSIS

The screenshot displays the CartograPlant web interface. At the top, a navigation bar includes links for Dashboard, Content, Structure, Tripal, Appearance, People, Modules, Mainlab, CartograTree Admin, Configuration, TG Gus, Reports, and Help. The user is logged in as irenecobo88@gmail.com. A central modal window is open, titled 'CartograPlant', showing analysis details: Analysis ID: 1015, Phenotypes: 0, Genotypes: 0, and Environmental layers: 0. Below this, a series of tabs are visible: Manage, Filter By Traits, Filter By Genotypes, Filtering & Imputation, Population Structure, Add environmental data, Run Analysis, and Summary and Confirm. The 'Manage' tab is active, displaying a welcome message and the unique analysis ID: 1015. The analysis name is 'Untitled' and the type is 'Genotype x Phenotype x Environmental'. A 'Test' dropdown menu is present, and a 'CREATE NEW WORKSPACE' button is highlighted. The 'MANAGE WORKSPACE FILES' section indicates that the workspace currently contains no files. A 'NEXT' button is visible at the bottom right of the modal. The background shows a map of Canada with plant distribution markers: 263 Plants in the Pacific Northwest, 50 Plants near Seattle, and 2 Plants in the Saskatchewan region.

## DATA ANALYSIS: WORKSPACE CREATION

The screenshot shows the CartograPlant web application interface. A modal window is open, displaying the following information:

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

The modal also includes a navigation bar with buttons: Manage, Filter By Traits, Filter By Genotypes, Filtering & Imputation, Population Structure, Add Environmental data, Run Analysis, and Summary and Confirm. Below this, a welcome message states: "Welcome! Your unique analysis ID: 1015". The analysis name is "Untitled" and the analysis type is "Genotype x Phenotype x Environmental". An "UPDATE" button is present. A section titled "MANAGE WORKSPACE FILES" indicates that the workspace currently contains no files. A "CREATE NEW WORKSPACE" button is visible. The background shows a map of Canada with plant locations marked as circles: 263 Plants, 50 Plants, and 2 Plants.

## DATA ANALYSIS: WORKSPACE CREATION

The screenshot shows the CartograPlant web application interface. At the top, there is a navigation bar with various menu items like Dashboard, Content, Structure, Tripal, Appearance, People, Modules, Mainlab, CartograTree Admin, Configuration, TG Gus, Reports, and Help. The user is logged in as 'Hello irenecobo88@gmail.com'. The main content area displays the analysis details for ID 1015, including 'Phenotypes: 0', 'Genotypes: 0', and 'Environmental layers: 0'. Below this, there are several tabs: 'Manage', 'Filter By Traits', 'Filter By Genotypes', 'Filtering & Imputation', 'Population Structure', 'Add environmental data', 'Run Analysis', and 'Summary and Confirm'. The 'Manage' tab is active, showing a 'Welcome!' message and the analysis ID '1015'. The analysis name is 'PAG30 workshop test' (highlighted with a red box), and the analysis type is 'Genotype x Phenotype x Environmental'. There is an 'UPDATE' button next to the analysis type. Below this, there is a section for selecting a workspace, with a 'Test' dropdown menu and a 'CREATE NEW WORKSPACE' button. The 'MANAGE WORKSPACE FILES' section shows that the workspace currently contains no files. At the bottom right of the workspace management area, there are 'NEXT' and 'CLOSE' buttons. The background shows a map of Canada with plant locations marked by circles, such as '263 Plants' near Seattle, '50 Plants' near WASH., and '2 Plants' near Regina.

## DATA ANALYSIS: WORKSPACE CREATION



The screenshot displays the CartograPlant web application interface. At the top, there is a navigation menu with options like Dashboard, Content, Structure, Tripal, Appearance, People, Modules, Mainlab, CartograTree Admin, Configuration, TG Gus, Reports, and Help. The user is logged in as irenecobo88@gmail.com. The main workspace management area is titled 'CartograPlant' and shows analysis details: Analysis ID: 1015, Phenotypes: 0, Genotypes: 0, and Environmental layers: 0. Below this, there are several tabs: Manage, Filter By Traits, Filter By Genotypes, Filtering & Imputation, Population Structure, Add environmental data, Run Analysis, and Summary and Confirm. The 'Manage' tab is active, showing a 'Welcome!' message and the unique analysis ID: 1015. The analysis name is 'PAG30 workshop test' and the analysis type is 'Genotype x Phenotype x Environmental'. A red box highlights the 'Workspace name' input field containing 'Test2' and the 'Create' button. A mouse cursor is pointing at the 'CREATE NEW WORKSPACE' button. The 'MANAGE WORKSPACE FILES' section shows that the workspace currently contains no files. The background shows a map of Canada with plant locations marked.

## DATA ANALYSIS: WORKSPACE CREATION

The screenshot shows the CartograPlant web interface. At the top, there is a navigation menu with items like Dashboard, Content, Structure, Tripal, Appearance, People, Modules, Mainlab, CartograTree Admin, Configuration, TG Gus, Reports, and Help. The user is logged in as irenecobo88@gmail.com. A modal window is open, displaying the workspace creation process. The modal has a title bar with the CartograPlant logo and a close button. Below the title bar, there are several tabs: Manage, Filter By Traits, Filter By Genotypes, Filtering & Imputation, Population Structure, Add environmental data, Run Analysis, and Summary and Confirm. The 'Manage' tab is active. The main content of the modal includes a welcome message, the analysis ID (1015), and the analysis name (PAG30 workshop test). There is a section for selecting a workspace, with a dropdown menu showing 'Test' and a 'CREATE NEW WORKSPACE' button. Below this, there is a field for the workspace name (Test2) and a 'Create' button. The modal also has a 'MANAGE WORKSPACE FILES' section, which currently shows 'This workspace currently contains no files'. At the bottom right of the modal, there are 'NEXT' and 'CLOSE' buttons. The 'NEXT' button is highlighted with a red box and a mouse cursor. The background of the modal shows a map of Canada with plant locations marked by circles and labels like '263 Plants', '50 Plants', and '2 Plants'.

## DATA ANALYSIS: WORKSPACE CREATION

The screenshot displays the CartograPlant web interface. On the left is a sidebar with navigation options: Dashboard, Content, Add content, Add Tripal Co, Environment Layers, Map Summary, Number of Plants, Selected Plants, Species count, Publications Count, Number of Layers, Plant Dataset Source, and Filters. The main panel is titled 'Filter By Traits' and shows '2 studies detected based on the trees you selected on the map': TGDR022 and TGDR474. Below this, there are filter options for 'diameter at breast height' (20,284 phenotypes, centimeter), 'plant height' (20,284 phenotypes, meter), and 'elevation' (1,680 phenotypes, centimeter). A histogram for 'plant height' is shown with a distribution peaking around 10-12 units. A dropdown menu for the 'plant height' filter is open, showing 'meter' selected. A blue arrow points from the histogram to a 'frost free days' histogram, which has a 'Save adjustment' button highlighted in a red box. A 'PCA (3 phenotypes)' plot is also visible. A red box at the top right contains the text: 'Analysis ID: 1015', 'Phenotypes: 3', 'Genotypes: 0', and 'Environmental layers: 0'. A 'NEXT' button is at the bottom right.

## DATA ANALYSIS: PHENOTYPIC TRAITS SELECTION/VISUALIZATION

The screenshot shows the CartograPlant web application interface. At the top, there is a navigation menu with items like Dashboard, Content, Structure, Tripal, Appearance, People, Modules, Mainlab, CartograTree Admin, Configuration, TG Gus, Reports, and Help. The user is logged in as 'Hello irenecobo88@gmail.com'. The main panel displays the 'Filter By Genotypes' step of an analysis. It shows two studies selected: TGDR674 and TGDR682. A Venn diagram visualizes the overlap: TGDR674 has 0 unique SNPs, TGDR682 has 8183 unique SNPs, and they share 24887 overlapping SNPs. Below the diagram, there are checkboxes for 'SNP overlaps between TGDR674, TGDR682: 24887', 'TGDR674 has 24887 SNPs', and 'TGDR682 has 33070 SNPs'. There are also summary statistics for non-overlaps: 'TGDR674 has 0 non-overlaps' and 'TGDR682 has 8183 non-overlaps'. A 'NEXT' button is highlighted by a mouse cursor, and a 'CLOSE' button is also visible. The sidebar on the left contains 'Map Options', 'Map Summary', 'Number of Plants', 'Selected Plants', 'Species count', 'Publications Count' (2), and 'Number of Layers' (0).

## DATA ANALYSIS: GENOTYPIC DATA SELECTION/VISUALIZATION

CartograPlant

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

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

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

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

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

TGDR674 0 24887 TGDR682 8183

NEXT  
CLOSE

## DATA ANALYSIS: GENOTYPIC DATA SELECTION/VISUALIZATION

The screenshot shows the CartograPlant web application interface. At the top, there is a navigation menu with options like Dashboard, Content, Structure, Tripal, Appearance, People, Modules, Mainlab, CartograTree Admin, Configuration, TG Gus, Reports, and Help. The user is logged in as 'Hello irenecobo88@gmail.com'. The main analysis window displays 'Analysis ID: 1053', 'Phenotypes: 3', 'Genotypes: 0', and 'Environmental layers: 0'. Below this, there are several tabs: Manage, Filter By Traits, Filter By Genotypes (selected), Filtering & Imputation, Population Structure, Add environmental data, Run Analysis, and Summary and Confirm. The 'Filter By Genotypes' section shows '2 studies detected based on the trees you selected on the map' and lists 'TGDR674' and 'TGDR682'. It states 'Overlapping genotypes: 24887'. A Venn diagram visualizes the overlap: TGDR674 has 0 unique SNPs, TGDR682 has 8183 unique SNPs, and they share 24887 overlapping SNPs. A red box highlights the 'Insights' section, which contains the following text: 'We analyzed 2 studies and discovered the following: There are a total of 24887 SNP overlaps across all 2 studies'. Below this, there are four bullet points: 'SNP overlaps between TGDR674, TGDR682: 24887', 'TGDR674 has 24887 SNPs', 'TGDR682 has 33070 SNPs', and 'TGDR674 has 0 non-overlaps' and 'TGDR682 has 8183 non-overlaps'. At the bottom right of the analysis window, there are 'NEXT' and 'CLOSE' buttons. A mouse cursor is pointing at the 'CLOSE' button. The sidebar on the left contains 'Map Options', 'Environment Layers', 'Map Summary', 'Number of Plants', 'Selected Plants', 'Species count', 'Publications Count' (2), and 'Number of Layers' (0).

## DATA ANALYSIS: GENOTYPIC DATA SELECTION/VISUALIZATION

The screenshot displays the CartograPlant web application interface. At the top, there is a navigation menu with options like Dashboard, Content, Structure, Tripal, Appearance, People, Modules, Mainlab, CartograTree Admin, Configuration, TG Gus, Reports, and Help. The user is logged in as 'Hello irenecobo88@gmail.com'. The main content area shows the 'Filter By Genotypes' section, which has detected 2 studies: TGDR674 and TGDR682. A Venn diagram visualizes the overlap between the two studies, showing 24887 overlapping genotypes. The 'Insights' section, highlighted with a red box, provides a summary of the analysis: 'We analyzed 2 studies and discovered the following: There are a total of 24887 SNP overlaps across all 2 studies'. Below this, a list of findings is shown: 'SNP overlaps between TGDR674, TGDR682: 24887', 'TGDR674 has 24887 SNPs', 'TGDR682 has 33070 SNPs', 'TGDR674 has 0 non-overlaps', and 'TGDR682 has 8183 non-overlaps'. The interface also includes a 'Map Summary' section on the left and a 'Map Options' section at the top. A 'NEXT' button is visible at the bottom right of the main content area.

## DATA ANALYSIS: GENOTYPIC DATA SELECTION/VISUALIZATION

# DATA ANALYSIS

<https://cartograplant.org/>

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

The screenshot displays the CartograPlant web application interface. At the top, a navigation bar includes links for Dashboard, Content, Structure, Tripal, Appearance, People, Modules, Mainlab, CartograTree Admin, Configuration, TG Gus, Reports, and Help. A user profile for 'Hello irenecobo88@gmail.com' is visible in the top right corner. The main content area features a 'Filter By Genotypes' panel with a summary of the analysis. A red box highlights the analysis details: Analysis ID: 1053, Phenotypes: 3, Genotypes: 24887, and Environmental layers: 0. Below this, a Venn diagram shows the overlap between two studies: TGDR674 (0 unique genotypes) and TGDR682 (8183 unique genotypes), with 24887 overlapping genotypes. The interface also includes a 'Map Summary' section with a list of insights, such as 'SNP overlaps between TGDR674, TGDR682: 24887'. A 'NEXT' button is highlighted with a mouse cursor, and a 'CLOSE' button is also visible. A white text box at the bottom of the screenshot reads: 'Reference genome information and SNP remapping analysis available soon!!!'

## DATA ANALYSIS: GENOTYPIC DATA SELECTION/VISUALIZATION



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

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

✔ VCF files successfully uploaded to workspace

✔ Found Genotype VCF for TGDR674  
✔ Found Genotype VCF for TGDR682

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

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

Missingness per individual, filtering

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

Select the file to use.

Workspace available files: [ ]

Output from the step 1 (part 3)

Select the file to use.

✔ LinkImputeR Step 1 - Accuracy Mode  
LinkImputeR Step 2 - Imputation  
SNP Quality Filtering Step 1.1: Missingness per individual list  
SNP Quality Filtering Step 1.2: Missingness per individual sorted list  
SNP Quality Filtering Step 1.3: Missingness per individual, individual selection  
SNP Quality Filtering Step 1.4: Missingness per individual, filtering  
SNP Quality Filtering Step 2: Minor allele count per marker  
SNP Quality Filtering Step 3: Minimum quality score  
SNP Quality Filtering Step 4: Minimum reads per marker  
SNP Quality Filtering Step 5: Missingness per marker  
SNP Quality Filtering Step 6: Minimum allele frequency  
SNP Quality Filtering Step 7: Hardy-Weinberg equilibrium  
SNP Quality Filtering Step 8: Mendelian errors (for family-based data only)  
SNP Quality Filtering Step 9: Linkage disequilibrium (retain only independent SNPs f

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

## DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

Dashboard Content Structure Tripal Appearance People Modules Mainlab CartograTree Admin Configuration TG Gus Reports Help Hello irenecobo88@gmail.com Log out

Add content Add Tripal Cor

CartograPlant

Map Options

Environment Layers

Map Summary

Number of Plants

Selected Plants

Species count

Publications Count

Number of Layers

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

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

VCF files successfully uploaded to workspace

Found Genotype VCF for TGDR674  
Found Genotype VCF for TGDR682

Quality filtering method SNP Quality Filtering Step 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:

INITIATE ANALYSIS JOB

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

## DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

Dashboard Content Structure Tripal Appearance People Modules Mainlab CartograTree Admin Configuration TG Gus Reports Help Hello irenecobo88@gmail.com Log out

Add content Add Tripal Cor

CartograPlant

Map Options

Environment Layers

Map Summary

Number of Plants

Selected Plants

Species count

Publications Count

Number of Layers

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

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

🌟 VCF files successfully uploaded to workspace

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

Quality filtering method: SNP Quality Filtering Step 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: Analysis 1053: TGDR674 Filtered SNPs VCF (a minute ago)

INITIATE ANALYSIS JOB

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

## DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

The screenshot shows the CartograPlant web interface. The top navigation bar includes links for Dashboard, Content, Structure, Tripal, Appearance, People, Modules, Mainlab, CartograTree Admin, Configuration, TG Gus, Reports, and Help. The user is logged in as irenecobo88@gmail.com. The main content area features a toolbar with buttons for Manage, Filter By Traits, Filter By Genotypes, Filtering & Imputation (selected), Population Structure, Add environmental data, Run Analysis, and Summary and Confirm. A notification bar indicates that VCF files were successfully uploaded to the workspace. Below this, two green checkmarks confirm the presence of genotype VCF files for TGDR674 and TGDR682. The 'Quality filtering method' is set to 'SNP Quality Filtering Step 3: Minimum quality score'. The current step is 'SNP QUALITY FILTERING STEP 3: MINIMUM QUALITY SCORE ANALYSIS CONFIGURATION'. The 'Output of Step 2' section shows a file selection interface with a message: 'Select the file to use.' Below this, a list of workspace files is shown, with the first file, 'Analysis 1053: TGDR674 Filtered SNPs VCF (a minute ago)', highlighted in blue and enclosed in a red box. A mouse cursor is pointing at this file. A blue 'INITIATE ANALYSIS JOB' button is visible at the bottom left of the configuration area. A refresh button is present on the right side of the file selection area.

## DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

The screenshot displays the CartograPlant web interface. At the top, a navigation bar includes 'Dashboard', 'Content', 'Structure', 'Tripal', 'Appearance', 'People', 'Modules', 'Mainlab', 'CartograTree Admin', 'Configuration', 'TG Gus', 'Reports', and 'Help'. The user is logged in as 'Hello irenecobo88@gmail.com'. The main content area is titled 'Analysis ID: 1053' and shows 'Phenotypes: 3', 'Genotypes: 24887', and 'Environmental layers: 0'. A series of tabs at the top of the main panel includes 'Manage', 'Filter By Traits', 'Filter By Genotypes', 'Filtering & Imputation' (which is active), 'Population Structure', 'Add environmental data', 'Run Analysis', and 'Summary and Confirm'. A green notification bar states 'VCF files successfully uploaded to workspace'. Below this, two green checkmarks indicate 'Found Genotype VCF for TGDR674' and 'Found Genotype VCF for TGDR682'. A dropdown menu for 'Quality filtering method' is set to 'SNP Quality Filtering Step 3: Minimum quality score'. The configuration section is titled 'SNP QUALITY FILTERING STEP 3: MINIMUM QUALITY SCORE ANALYSIS CONFIGURATION'. It shows 'Minimum quality score' and 'Output of Step 2' with a message 'Select the file to use.' and a dropdown menu showing 'Workspace available files: Analysis 1053: TGDR674 Filtered SNPs VCF (a minute ago)'. A blue button labeled 'INITIATE ANALYSIS JOB' is highlighted with a red box and a mouse cursor. A refresh button and a message 'If files don't appear, please try refreshing the workflow' are also visible.

## DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

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 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**

Number of SNPs

Minimum quality score

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

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

RETURN TO QUALITY FILTERING METHODS

NEXT

## DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

The screenshot displays the CartograPlant web interface. At the top, a navigation menu includes Dashboard, Content, Structure, Tripal, Appearance, People, Modules, Mainlab, CartograTree Admin, Configuration, TG Gus, Reports, and Help. The user is logged in as irenecobo88@gmail.com. The main content area shows the 'Filtering & Imputation' step of an analysis. A summary box at the top right indicates: Analysis ID: 1053, Phenotypes: 3, Genotypes: 24887, and Environmental layers: 0. Below this, a series of tabs includes Manage, Filter By Traits, Filter By Genotypes, Filtering & Imputation (active), Population Structure, Add environmental data, Run Analysis, and Summary and Confirm. A green notification bar states 'VCF files successfully uploaded to workspace'. Below that, two green checkmarks confirm 'Found Genotype VCF for TGDR674' and 'Found Genotype VCF for TGDR682'. A dropdown menu is open, listing various quality filtering steps, with 'SNP Quality Filtering Step 9: Linkage disequilibrium (retain only independent SNPs for further population structure analysis)' highlighted in blue. The interface also shows a 'Quality filtering method' section with 'SNP QUALITY FILTERING' selected, and a 'Workspace available files' section with a 'Select the file to use.' button. A map of Alaska is visible on the right side of the interface.

## DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

The screenshot displays the CartograPlant web interface. At the top, a navigation bar includes 'Dashboard', 'Content', 'Structure', 'Tripal', 'Appearance', 'People', 'Modules', 'Mainlab', 'CartograTree Admin', 'Configuration', 'TG Gus', 'Reports', and 'Help'. The user is logged in as 'Hello irenecobo88@gmail.com'. The main content area shows an analysis summary for Analysis ID: 1053, with 3 Phenotypes, 24887 Genotypes, and 0 Environmental layers. A progress bar indicates the current step is 'Filtering & Imputation'. A green notification box states 'VCF files successfully uploaded to workspace'. Below this, two green checkmarks confirm 'Found Genotype VCF for TGDR674' and 'Found Genotype VCF for TGDR682'. The 'Quality filtering method' is set to 'SNP Quality Filtering Step 9: Linkage disequilibrium (retain only independent SNPs for further population structure analysis)'. The configuration for this step is shown, including the output of Step 8 and a dropdown menu for 'Workspace available files' with the selected option 'QCfiltering\_3: Minimum quality score'. A red box highlights this dropdown menu. A blue button labeled 'INITIATE ANALYSIS JOB' is visible at the bottom left of the main content area. A sidebar on the left contains various map and analysis options, and a map is visible on the right side.

## DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION



The screenshot displays the CartograPlant web interface. At the top, a navigation bar includes 'Dashboard', 'Content', 'Structure', 'Tripal', 'Appearance', 'People', 'Modules', 'Mainlab', 'CartograTree Admin', 'Configuration', 'TG Gus', 'Reports', and 'Help'. The user is logged in as 'ireneCobo88@gmail.com'. The main content area shows the 'Filtering & Imputation' step of an analysis workflow. A summary box at the top right indicates: Analysis ID: 1053, Phenotypes: 3, Genotypes: 24887, and Environmental layers: 0. Below this, a series of tabs includes 'Manage', 'Filter By Traits', 'Filter By Genotypes', 'Filtering & Imputation' (active), 'Population Structure', 'Add environmental data', 'Run Analysis', and 'Summary and Confirm'. A green notification bar states 'VCF files successfully uploaded to workspace'. Below this, two green checkmarks confirm 'Found Genotype VCF for TGDR674' and 'Found Genotype VCF for TGDR682'. The 'Quality filtering method' is set to 'SNP Quality Filtering Step 9: Linkage disequilibrium (retain only independent SNPs for further population structure analysis)'. A large grey box contains the text 'SNP QUALITY FILTERING STEP 9: LINKAGE DISEQUILIBRIUM (RETAIN ONLY INDEPENDENT SNPs FOR FURTHER POPULATION STRUCTURE ANALYSIS) ANALYSIS CONFIGURATION'. Below this, the text 'Linkage disequilibrium (retain only independent SNPs for further population structure analysis)' is displayed. A gear icon indicates 'Output of Step 8 (family-based studies) or 7 (non-family-based studies)'. A message says 'Select the file to use.' and a dropdown menu shows 'Workspace available files: QCfiltering\_3: Minimum quality score'. A blue button with a refresh icon is labeled 'If files don't appear, please try refreshing the workflow'. At the bottom left, a blue button labeled 'INITIATE ANALYSIS JOB' is highlighted with a red box and a mouse cursor.

## DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

Dashboard Content

Add content Add Tripal Cor

CartograPlant

Map Options

Environment Layers

Map Summary

Number of Plants

Selected Plants

Species count

Publications Count

Number of Layers

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) 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:

INITIATE ANALYSIS JOB

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

NEXT

CLOSE

## DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

The screenshot displays the CartograPlant web interface. On the left is a dark sidebar with navigation options: Dashboard, Content, Add content, Add Tripal Cor, CartograPlant, Map Options, Environment Layers, Map Summary, Number of Plants, Selected Plants, Species count, Publications Count, and Number of Layers. The main content area shows a success message: "VCF files successfully uploaded to workspace". Below this are two green checkmarks: "Found Genotype VCF for TGDR674" and "Found Genotype VCF for TGDR682". A dropdown menu for "Quality filtering method" is set to "SNP Quality Filtering Step 9: Linkage disequilibrium (retain only independent SNPs for further population structure analysis)". A grey box contains the text "SNP QUALITY FILTERING STEP 9: LINKAGE DISEQUILIBRIUM (RETAIN ONLY INDEPENDENT SNPS FOR FURTHER POPULATION STRUCTURE ANALYSIS) ANALYSIS CONFIGURATION". Below this, the same text is repeated. A gear icon indicates "Output of Step 8 (family-based studies) or 7 (non-family-based studies)". A message says "Select the file to use." with a note "If files don't appear, please try refreshing the workflow" and a refresh button. The "Workspace available files:" dropdown shows "QCfiltering\_3: Minimum quality score". A blue button "INITIATE ANALYSIS JOB" is present. Below it, a message states "Successfully submitted job to Galaxy server" with "Invocation ID: 49ec47cc9a1bdc41" and "Status: Submitted". A red box highlights the "Output Results (running)" section, which shows "Awaiting" for "File QCfiltering\_9: Linkage disequilibrium (0 bytes)". A "NEXT" button is visible at the bottom right, with a mouse cursor pointing to it.

## DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

The screenshot displays the CartograPlant web interface. At the top, a navigation menu includes Dashboard, Content, Structure, Tripal, Appearance, People, Modules, Mainlab, CartograTree Admin, Configuration, TG Gus, Reports, and Help. The user is logged in as 'Hello IreneCobo88@gmail.com'. A central modal window is open, titled 'CartograPlant', showing analysis details: Analysis ID: 1053, Phenotypes: 3, Genotypes: 24887, and Environmental layers: 0. The modal has a tabbed interface with 'Population Structure' selected. Below the tabs, there are input fields for 'Number of populations' (set to 2) and 'Select VCF file' (with a dropdown menu showing 'SELECT VCF FILE'). A blue button labeled 'GENERATE FAST STRUCT' is visible, with a mouse cursor hovering over it. To the right of the modal are 'NEXT' and 'CLOSE' buttons. In the background, a world map is visible with several countries highlighted, including Peru, Brazil, Chile, São Paulo, Uruguay, Angola, Namibia, Madagascar, South Africa, Tanzania, Indonesia, Papua New Guinea, Australia, Brisbane, Perth, Melbourne, and New Zealand. On the left side of the interface, a sidebar shows a summary of plant data: Number of Plants (1934), Selected Plants (967), Species count (2), Publications Count (2), and Number of Layers (0).

## DATA ANALYSIS: POPULATION STRUCTURE CALCULATION/VISUALIZATION

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

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

Number of populations:   
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**

NEXT | CLOSE

Number of Plants: 1934 [UNSEL ALL]  
Selected Plants: 967  
Species count: 2  
Publications Count: 2  
Number of Layers: 0

## DATA ANALYSIS: POPULATION STRUCTURE CALCULATION/VISUALIZATION

Dashboard Content Structure Tripal Appearance People Modules Mainlab CartograTree Admin Configuration TG Gus Reports Help Hello irenecobo88@gmail.com Log out

CartograPlant

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

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

Number of populations  
2

Select VCF file  
QCfiltering\_9: Linkage disequilibrium

GENERATE FAST STRUCTURE

Step 1 - Performing PLINK process on VCF file... ( )

This can take minutes, please be patient...

NEXT

CLOSE

Map Options  
Environment Layers  
Map Summary  
Number of Plants  
Selected Plants  
Species count 2  
Publications Count 2  
Number of Layers 0

## DATA ANALYSIS: POPULATION STRUCTURE CALCULATION/VISUALIZATION

The screenshot displays the CartograPlant web application interface. At the top, a navigation menu includes 'Dashboard', 'Content', 'Structure', 'Tripal', 'Appearance', 'People', 'Modules', 'Mainlab', 'CartograTree Admin', 'Configuration', 'TG Gus', 'Reports', and 'Help'. The user is logged in as 'irene88@gmail.com'. A central panel shows analysis statistics: 'Analysis ID: 1015', 'Phenotypes: 3', 'Genotypes: 15326', 'Individuals: 426', and 'Environmental layers: 0'. Below this, a 'Population Structure' tab is active, featuring a 'Number of populations' input set to '2' and a 'Select VCF file' dropdown menu showing 'Analysis 1016: TGDR725 Filtered SNPs VCF (14 minutes ago)'. A 'GENERATE FAST STRUCTURE' button is prominently displayed. To the left, a PCA plot shows PC2 vs PC1 with points colored by population. To the right, a stacked bar chart shows the results for K=2, 3, 4, and 5, with a legend for populations WS (green), QI (red), and Tp (blue). A map of Australia and New Zealand is visible in the background. A mouse cursor is pointing at the 'NEXT' button.

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

# DATA ANALYSIS: POPULATION STRUCTURE CALCULATION/VISUALIZATION

The screenshot shows the CartograPlant web application interface. A modal window is open, displaying analysis statistics and options for selecting environmental layers. The modal title is "CartograPlant" and it includes the following statistics: Analysis ID: 1015, Phenotypes: 3, Genotypes: 15326, Individuals: 426, and Environmental layers: 0. The modal contains a navigation bar with buttons: Manage, Filter By Traits, Filter By Genotypes, Filtering & Imputation, Population Structure, Add environmental data, Run Analysis, and Summary and Confirm. Below the navigation bar, there are two yellow buttons: "PRECACHE VALUES" and "GATHER AND UPLOAD TO WORKSPACE". The "Choose environmental layers" section has two radio button options: "CATEGORY US" and "CATEGORY WORLD". A mouse cursor is pointing at the "CATEGORY US" option. Below the radio buttons are two empty text input fields. At the bottom right of the modal, there are "NEXT" and "CLOSE" buttons. The background shows a map of Canada with plant distribution data points: 65 Plants in British Columbia, 43 Plants in Alberta, 263 Plants in Saskatchewan, and 2 Plants in Regina. The left sidebar shows various navigation options like "Environment Layers", "Map Summary", "Number of Plants", "Selected Plants", "Species count", "Publications Count", "Number of Layers", and "Plant Dataset Sources".

## DATA ANALYSIS: ENVIRONMENTAL DATA SELECTION/ MULTICOLLINEARITY



The screenshot displays the 'Choose environmental layers' section of the CartograPlant application. At the top, a navigation bar includes tabs for 'Manage', 'Filter By Traits', 'Filter By Genotypes', 'Filtering & Imputation', 'Population Structure', 'Add environmental data', 'Run Analysis', and 'Summary and Confirm'. The 'Add environmental data' tab is active. Below the navigation, there are two buttons: 'PRECACHE VALUES' and 'GATHER AND UPLOAD TO WORKSPACE'. The main content area is titled 'Choose environmental layers' and features a search bar with a magnifying glass icon and a 'SELECT ALL' button. A mouse cursor is pointing at the search bar. The list of layers is organized into categories: 'CATEGORY US' (checked), 'GROUP Biotic Damage (North America)' (unchecked), and 'GROUP Climatic variables (World, ClimateWNA)' (checked). Under the 'Climatic variables' group, there are three classifications: 'CLASSIFICATION DAY' (unchecked), 'CLASSIFICATION DAYS' (unchecked), and 'CLASSIFICATION MOISTURE' (unchecked). Each classification has associated properties: 'DAY' has 'Day when FFP begins CWNA (BFFP)' and 'The day when FFP ends CWNA (EFFP)'; 'DAYS' has 'Cooling degree-days CWNA (DD18)', 'Growing Degree Days Above 5°C CWNA (DD5)', 'Chilling Days Below 0 °C CWNA (DD0)', and 'Heating degree-days CWNA (DD\_18)'; 'MOISTURE' has 'Summer Heat Moisture Index (CWNA) (SHM)', 'Hargreaves climatic moisture deficit CWNA (CMD)', and 'Annual Heat Moisture Index CWNA (AHM)'. On the right side, a map of Canada is visible, showing the province of Saskatchewan and the city of Regina.

## DATA ANALYSIS: ENVIRONMENTAL DATA SELECTION/ MULTICOLLINEARITY

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

The screenshot displays the CartograPlant web interface. On the left is a dark sidebar with navigation options: Dashboard, Content, Add content, Add Tripal Co, Environment Layers, Map Summary, Number of Plants, Selected Plants, Species count, Publications Count, Number of Layers, Plant Dataset Source, and Filters. The main area has a top navigation bar with tabs: Manage, Filter By Traits, Filter By Genotypes, Filtering & Imputation, Population Structure, Add environmental data, Run Analysis, and Summary and Confirm. Below this is a 'Choose environmental layers' section. It features a search bar and a 'SELECT ALL' button. A list of environmental layers is shown, each with a checkbox and a category label (GROUP, CLASSIFICATION, or PROPERTY). The selected layers are: 'US' (CATEGORY), 'Climatic variables (World, ClimateWNA)' (GROUP), and 'Chilling Days Below 0 °C CWNA (DD0)' (PROPERTY). Other layers include 'Biotic Damage (North America)', 'Day when FFP begins CWNA (BFFP)', 'The day when FFP ends CWNA (EFFF)', 'Cooling degree-days CWNA (DD18)', 'Growing Degree Days Above 5°C CWNA (DD5)', 'Heating degree-days CWNA (DD\_18)', 'Summer Heat Moisture Index (CWNA) (SHM)', 'Hargreaves climatic moisture deficit CWNA (CMD)', and 'Annual Heat Moisture Index CWNA (AHM)'. On the right side of the main area, there are two orange buttons: 'PRECACHE VALUES' and 'GATHER AND UPLOAD TO WORKSPACE'. Below the second button, it says 'Found 145 Mean Annual Precipitation CWNA (MEAN\_ANNUAL\_PREC) values.' and there are two empty input fields. At the bottom right, a map shows a portion of Canada with 'SASKATCHEWAN' and 'Regina' labeled.

## DATA ANALYSIS: ENVIRONMENTAL DATA SELECTION/ MULTICOLLINEARITY

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

The screenshot displays the CartograPlant web interface. On the left is a navigation sidebar with options like 'Dashboard', 'Map Summary', and 'Filters'. The main area shows a 'Choose environmental layers' section with various categories and properties, such as 'Climatic variables (World, ClimateWNA)' and 'Chilling Days Below 0 °C CWNA (DD0)'. A 'PRECACHE VALUES' button is visible. Below this, a 'GATHER AND UPLOAD TO WORKSPACE' button is shown, followed by a message: 'Found 145 Mean Annual Precipitation CWNA (MEAN\_ANNUAL\_PREC) values.' To the right, there are two plots: a histogram of 'Annual Temperature CWNA, MEAN\_ANNUAL\_TEMP' and a scatter plot of 'Annual Temperature CWNA, MEAN\_ANNUAL\_TEMP' vs 'Annual Precipitation CWNA, MEAN\_ANNUAL\_PREC'. A correlation coefficient of 0.29 is displayed. The background shows a map of Canada with 'SASKATCHEWAN' and 'Regina' labeled.

## DATA ANALYSIS: ENVIRONMENTAL DATA SELECTION/ MULTICOLLINEARITY

# DATA ANALYSIS

<https://cartograplant.org/>

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

CartograPlant

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

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

Step 1 - Select workflow  
GWAS with EMMAX

REFRESH

Step 2 - Setup analysis

NEXT

CLOSE

25 Plants  
65 Plants  
43 Plants  
263 Plants  
50 Plants

BRITISH COLUMBIA

CartograPlant + Galaxy

**DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS**

# DATA ANALYSIS

<https://cartograplant.org/>

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

The screenshot displays the CartograPlant web application interface. At the top, a navigation bar includes links for Dashboard, Content, Structure, Tripal, Appearance, People, Modules, Mainlab, CartograTree Admin, Configuration, TG Gus, Reports, and Help. The user is logged in as irenecobo88@gmail.com. A central panel shows analysis details: Analysis ID: 1015, Phenotypes: 3, Genotypes: 32597, and Environmental layers: 0. Below this, a series of tabs includes Manage, Filter By Traits, Filter By Genotypes, Filtering & Imputation, Population Structure, Add environmental data, Run Analysis, and Summary and Confirm. A dropdown menu is open under 'Run Analysis', listing several workflows: GWAS with EMMAX (highlighted), Landscape genomics with Sambada, Landscape genomics with Bayenv, Multiple testing correction (FDR), BLUP calculation (phenotypes), Meta-analysis with METASOFT, and Meta-analysis with PLINK. A 'REFRESH' button is visible next to the dropdown. The background features a map of British Columbia with plant locations marked by circles and labels: 25 Plants, 65 Plants, 43 Plants, 263 Plants, and 50 Plants. A sidebar on the left contains various filters and statistics like Species count (2), Publications Count (2), and Number of Layers (0).



**DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS**

# DATA ANALYSIS

<https://cartograplant.org/>

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

The screenshot shows the CartograPlant web interface. At the top, there is a navigation bar with various menu items. Below it, a modal window displays analysis details: Analysis ID: 1015, Phenotypes: 3, Genotypes: 32597, and Environmental layers: 0. A series of tabs are visible: Manage, Filter By Traits, Filter By Genotypes, Filtering & Imputation, Population Structure, Add environmental data, Run Analysis, and Summary and Confirm. The 'Run Analysis' tab is active, showing a 'REFRESH' button and a 'NEXT' button. A dropdown menu is open, listing several analytic workflows: GWAS with EMMAX (highlighted), Landscape genomics with Sambada, Landscape genomics with Bayenv, Multiple testing correction (FDR), BLUP calculation (phenotypes), Meta-analysis with METASOFT, and Meta-analysis with PLINK. A mouse cursor is pointing at the 'GWAS with EMMAX' option. In the background, a map of British Columbia is visible with plant locations marked by blue circles and labeled with plant counts (e.g., 25 Plants, 65 Plants, 43 Plants, 50 Plants). A white text box at the bottom of the map area contains the text 'The other analytic workflows available soon!'.



## DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS

The screenshot shows the CartograPlant web interface. At the top right, there is a summary box for Analysis ID: 1053, Phenotypes: 3, Genotypes: 24887, and Environmental layers: 0. Below this is a navigation bar with buttons: Manage, Filter By Traits, Filter By Genotypes, Filtering & Imputation, Population Structure, Add environmental data, Run Analysis (highlighted with a red box), and Summary and Confirm. The main content area is titled 'GWAS WITH EMMAX ANALYSIS CONFIGURATION'. It includes a dropdown menu set to 'GWAS with EMMAX' and a 'REFRESH' button (also highlighted with a red box and a mouse cursor). The text describes the GWAS pipeline using EMMAX software. Below the text are two configuration sections: 'Genotype file containing SNPs in vcf or vcf.gz format' and 'LD filtered vcf'. Each section has a 'Select the file to use.' instruction, a 'Seleccionar archivo' button, a text input field with 'Ninguno ...hivo selec.', and an 'Upload to workspace' button. A workspace file list shows 'Analysis 1053: Filtered Phenotypes - bud set - ADJUSTED (10 minutes ago)'.

## DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS

**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) ✓

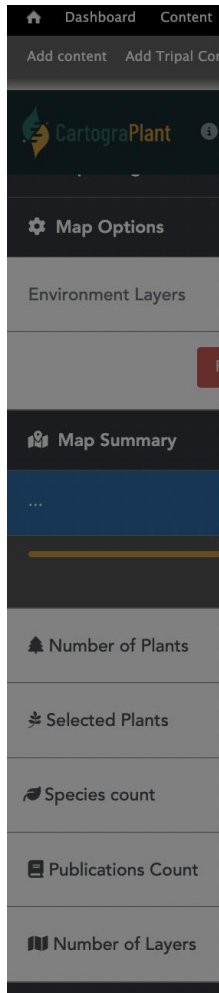
**Population Structure:** First two space delimited columns are family name and individual name in the same order as in the vcf file. The third column is recommended to be 1. The forth column contains the population number the individual is a member of  
**Select the file to use.**  
Seleccionar archivo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





Dashboard Content

Add content Add Tripal Cor

CartograPlant

Map Options

Environment Layers

Map Summary

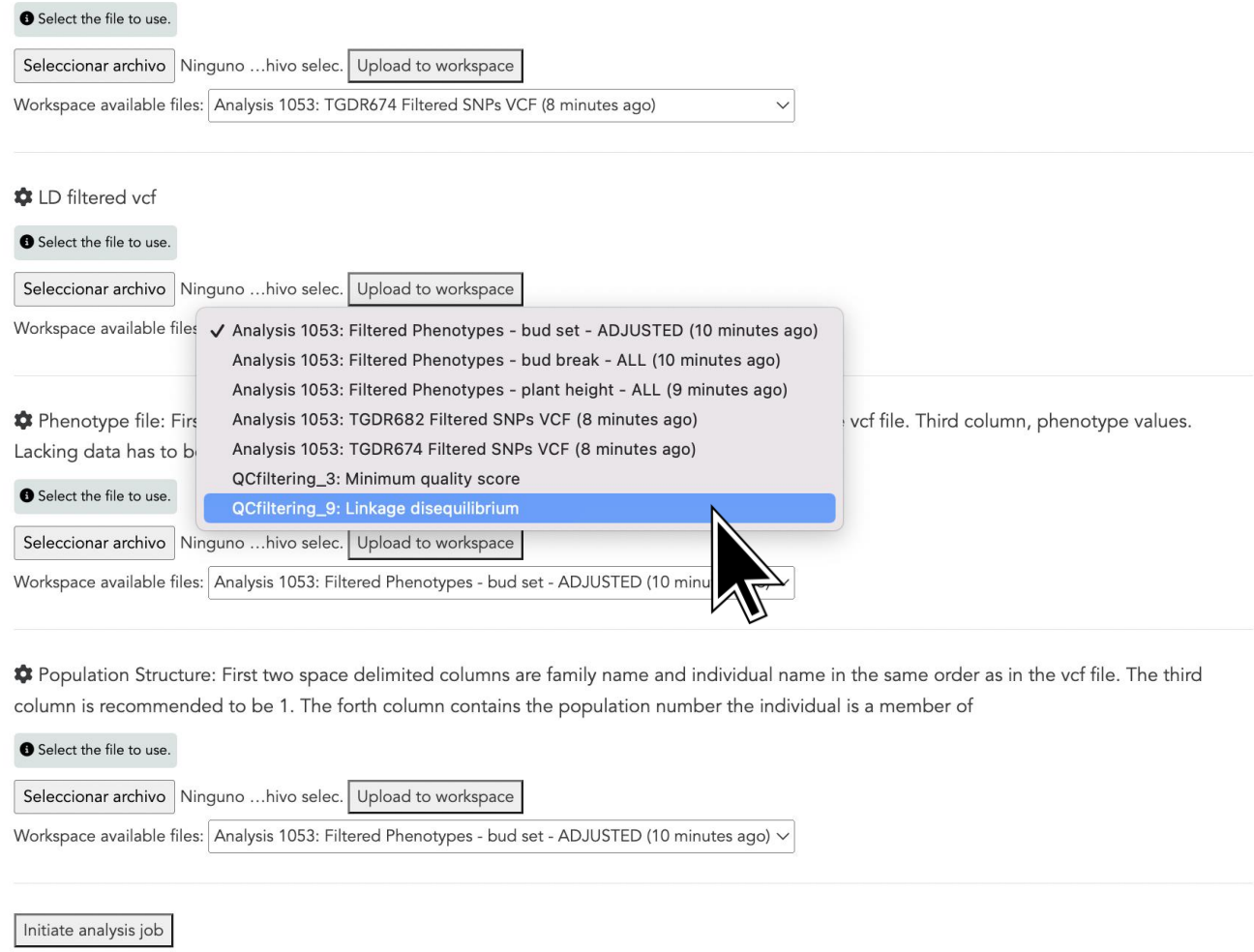
Number of Plants

Selected Plants

Species count

Publications Count

Number of Layers



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: First two space delimited columns are family name and individual name in the same order as in the vcf file. Third column, phenotype values.

Lacking data has to be

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 fourth 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



8@gmail.com Log out

Edit shortc

hecobo88@gmail.com

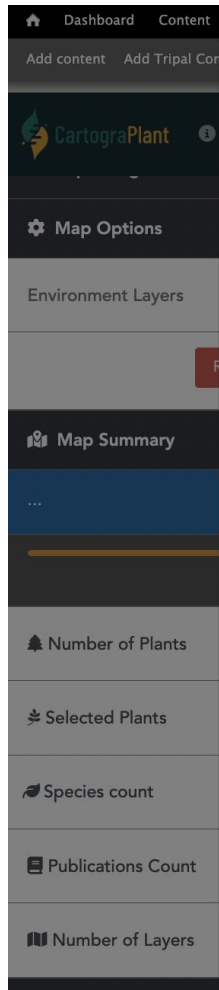
Map interface showing a geographical map with a toolbar on the right side.

## DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS

# DATA ANALYSIS

<https://cartograplant.org/>

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



Dashboard Content  
Add content Add Tripal Cor  
CartograPlant  
Map Options  
Environment Layers  
Map Summary  
Number of Plants  
Selected Plants  
Species count  
Publications Count  
Number of Layers

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



8@gmail.com Log out  
Edit shortcu  
hecobo88@gmail.com  
Map interface showing a map of the world with a sidebar on the right containing navigation controls.

**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: 693-1673367112-Panel1assignedPopDAPC.txt

Initiate analysis job

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)



NEXT

CLOSE

**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: 693-1673367112-Panel1assignedPopDAPC.txt

Initiate analysis job

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)

NEXT

CLOSE

**DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS**

Analysis ID: 1054  
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: **1054**  
Analysis name:  Analysis type:

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

### 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



- Stephen P. Ficklin



- Nic Herndon



- 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



- Margaret Staton
- Abdullah Almsaeed

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



@Irene\_Cobo  
@TreeGenes

