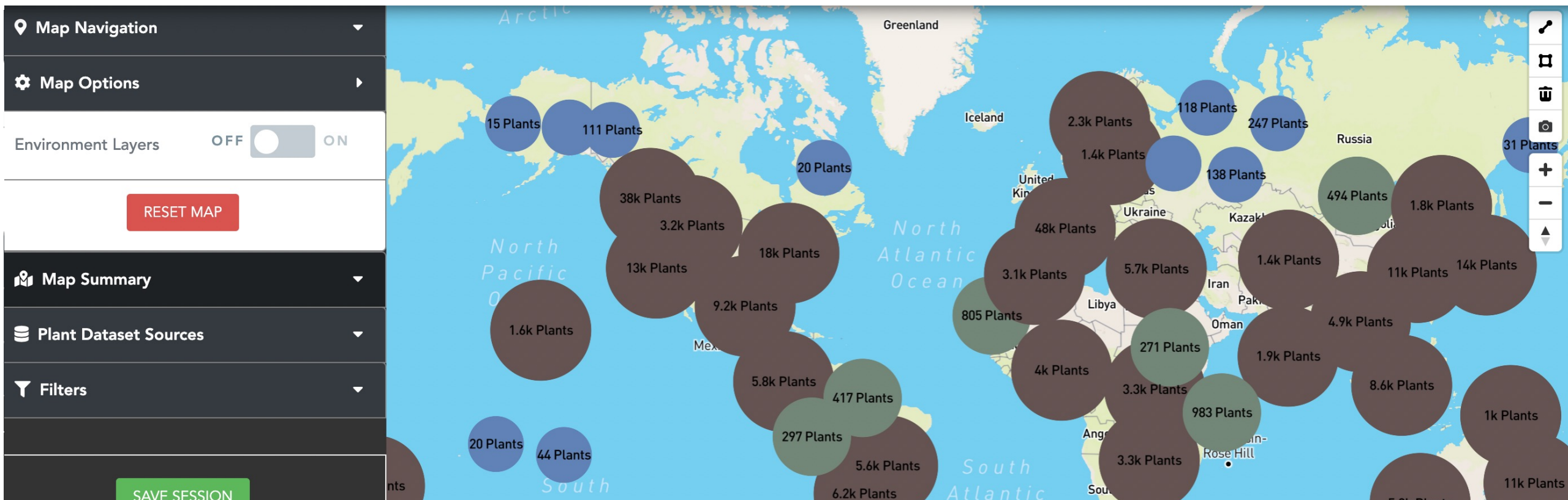


# Analytics with CartograPlant (GWAS and GEA): Demo

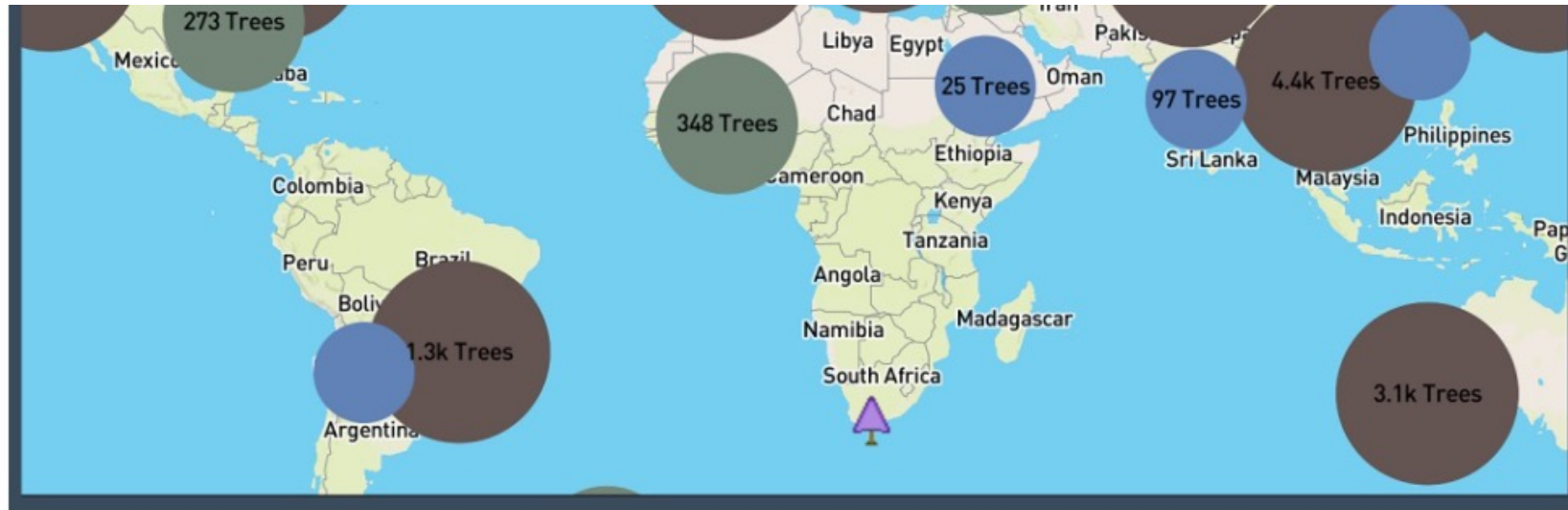


Irene Cobo Simón, PhD.

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

## 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



<https://cartograplant.org/>

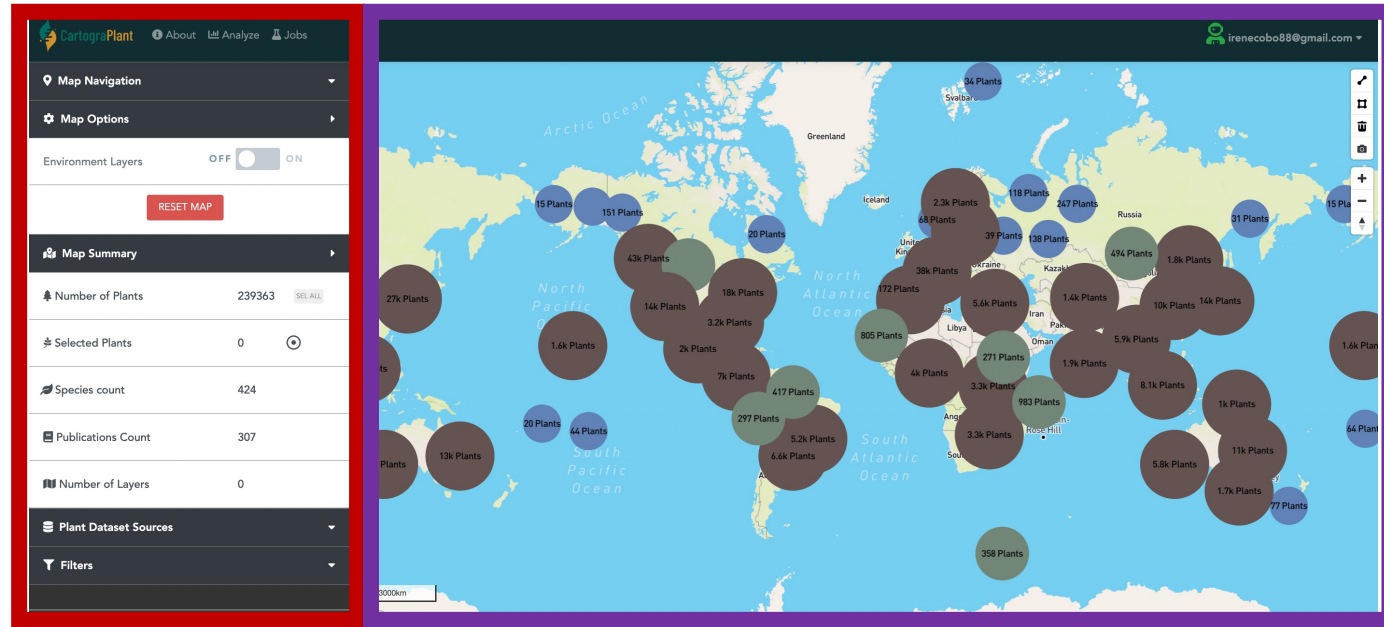
<https://gitlab.com/IreneCobo/cartograplantworkshop2022/>



# 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



## 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**

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', and a 'GO' button. This panel is highlighted with a red box.
- 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 toggle switches and opacity sliders. The 'Forest Fragmentation Risk (Contiguous US, 2002)' layer is currently active and highlighted with a green box.
- Map:** A map of North America showing plant distribution as bubbles of varying sizes and colors (blue, green, brown). A mouse cursor is pointing at a bubble in the central US.
- Environmental Data:** A table showing data for the selected layer: 'ENVIRONMENTAL LAYER: FOREST FRAGMENTATION RISK (CONTIGUOUS US, 2002)'. The table has four rows: RED BAND (0), GREEN BAND (185), BLUE BAND (92), and ALPHA BAND (255).
- Map Controls:** A vertical toolbar on the right side of the map, including zoom in (+), zoom out (-), and pan (arrow) buttons. This toolbar is highlighted with a red box.

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 summary panels. The main area is a map of the United States with plant locations marked by colored circles and polygons. A callout box on the right shows a detailed view of a plant specimen, including its name, coordinates, and a photograph. A table at the bottom of the callout provides plant genotypic data.

**Map Navigation**

**Map Options**

Environment Layers OFF ON

RESET MAP

**Map Summary**

Number of Plants 243623

**Selected plants**

Selected Plants 143

Species count 421

Publications Count 324

Number of Layers 0

**Plant Dataset Sources**

**Filters**

**Tree icon**

**Plant icon**

**TreeSnap icon**

**Polygon selecting plants**

**Polygon drawing button**

**Study info window**

TGDR300-LK1372  
31.51 Lat | -81.37 Lon  
Elevation: 0 m

Rosaceae  
Prunus serotina

**STUDY INFO**

**ADD PLANT**

**TGDR300-LK1419**

BIOME Temperate Broadleaf & Mixed Forests

Study Associated  
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/>



## DATA SELECTION FOR ANALYSIS



# INTERFACE AND DATA VISUALIZATION

<https://cartograplant.org/>

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

The screenshot displays the CartograPlant web interface. The top navigation bar includes the logo, 'About', 'Analyze', and 'Jobs' links, and a user profile for 'ireneCobo88@gmail.com'. The left sidebar is divided into sections: 'Map Summary', 'Plant Dataset Sources', and 'Filters'. The 'Map Summary' section shows statistics: Number of Plants (59916), Selected Plants (0), Species count (66), Publications Count (66), and Number of Layers (0). The 'Plant Dataset Sources' section is highlighted with a red box and contains five toggle switches: 'Internal submissions' (ON), 'TreeSnap' (OFF), 'Direct submissions' (OFF), 'BIEN' (OFF), and 'WFID' (OFF). A mouse cursor is pointing at the 'TreeSnap' toggle. The main area features a world map with circular data points of varying sizes and colors (blue, green, brown) representing plant counts in different regions. A scale bar for 3000km and a Mapbox logo are visible at the bottom of the map. The footer includes the text '© Mapbox © OpenStreetMap Improve this map'.

## 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 with the following data:

Category	Value	Action
Number of Plants	59916	<input type="button" value="UNSEL ALL"/>
Selected Plants	59916	<input type="radio"/>
Species count	66	
Publications Count	66	
Number of Layers	0	

Below the summary is the 'Plant Dataset Sources' section with toggle switches for: Internal submissions (ON), TreeSnap (OFF), Direct submissions (OFF), BIEN (OFF), and WFID (OFF). A 'Filters' section is partially visible at the bottom.

The main area features a world map with circular markers of varying sizes and colors (brown, green, blue) representing plant data points. Each marker is labeled with its count, such as '14k Plants', '11k Plants', '840 Plants', '6.9k Plants', '596 Plants', '1.6k Plants', '14k Plants', '11k Plants', '6k Plants', '296 Plants', '11k Plants', '106 Plants', '5.1k Plants', '1.8k Plants', and '18 Plants'. A scale bar for 3000km and a Mapbox logo are visible at the bottom left of the map. The URL <https://treegenesdb.org/cartogratre#tree-filter-options> is shown 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 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, a sidebar contains several toggle switches for data sources: TreeSnap (ON), Direct submissions (ON), BIEN (OFF), and WFID (OFF). Below these is a 'Filters' section with a dropdown menu set to 'Study Accession' and a value of 'TGDR674'. A search input field contains 'TGDR674' and a mouse cursor is hovering over it. A 'SAVE SESSION' button is visible at the bottom of the sidebar. The main area features a world map with numerous circular data points of varying sizes and colors (blue, green, brown) representing plant data. Each point is labeled with a count, such as '31 Plants', '117 Plants', '58k Plants', '983 Plants', and '358 Plants'. The map includes a 3000km scale bar and a 'mapbox' logo. The top navigation bar includes 'CartograPlant', 'About', 'Analyze', 'Jobs', and a user profile for 'ireneco88@gmail.com'.

## DATA SELECTION FOR ANALYSIS





# INTERFACE AND DATA VISUALIZATION

<https://cartograplant.org/>

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

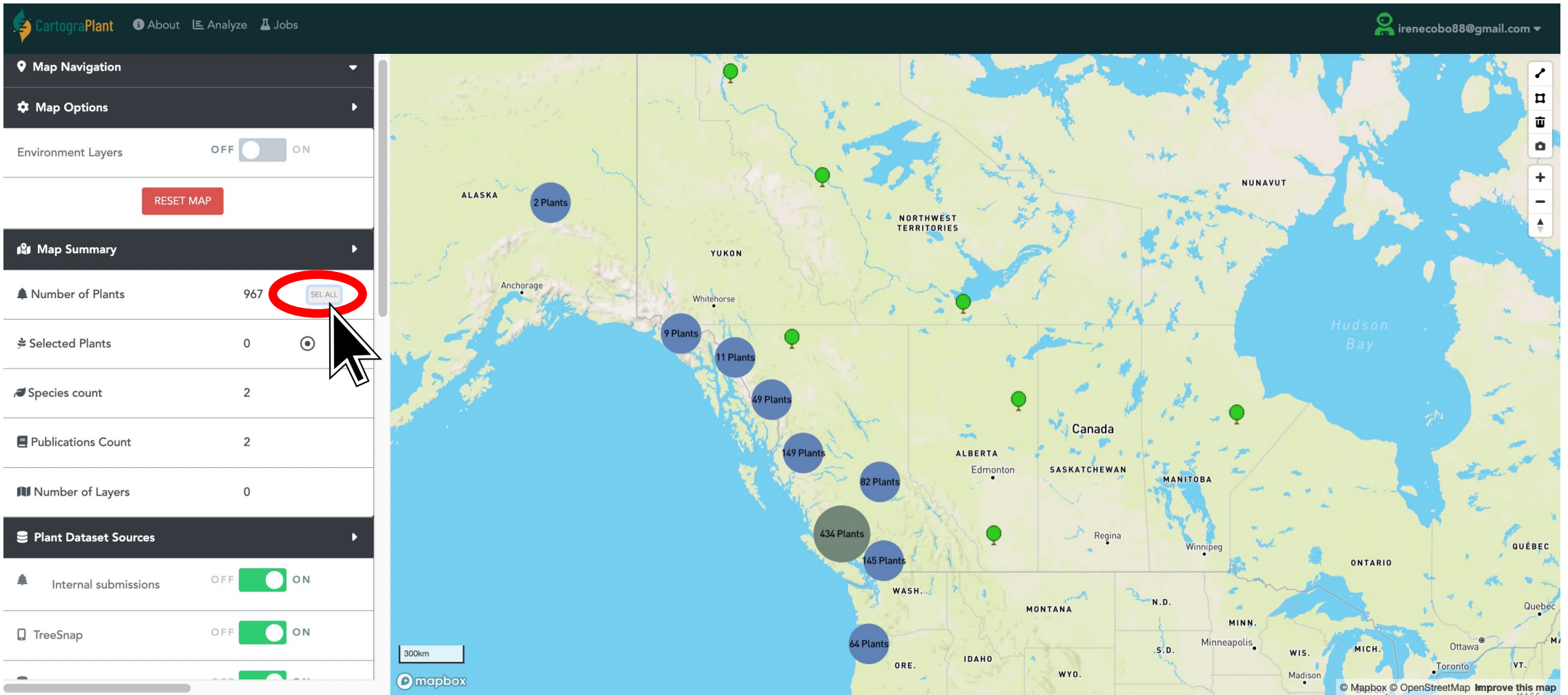
The screenshot displays the CartograPlant web interface. On the left, a 'Filters' panel is visible, containing two filter rules. The first rule is for 'Study Accession' TGDR674, with an 'equal' operator and '1 phenotypes found in study: frost free days'. The second rule is for 'Study Accession' TGDR675, with an 'equal' operator and '6 phenotypes found in study: leaf carbon to nitrogen ratio, leaf shape, maximum photosynthetic rate, nitrogen use efficiency, water use efficiency, wood carbon 12 content'. The 'APPLY FILTER' button at the bottom of the filter panel is circled in red, with a mouse cursor pointing to it. On the right, a world map shows the distribution of plants, with a blue circle labeled '17 Plants' over the Pacific Northwest and a larger grey circle labeled '880 Plants' over the United States. The map includes labels for major oceans, continents, and countries. The top navigation bar includes 'CartograPlant', 'About', 'Analyze', and 'Jobs' links, 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/>



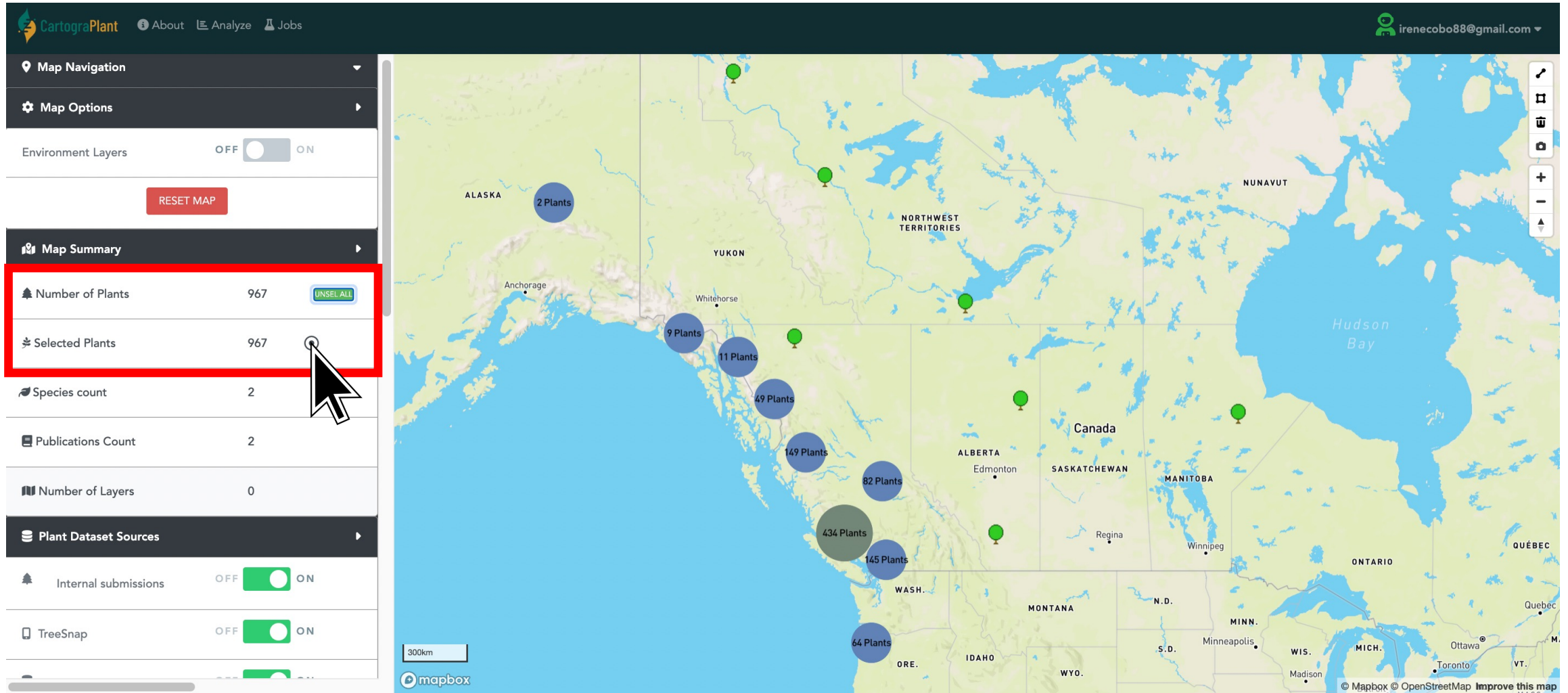
## DATA SELECTION FOR ANALYSIS



# INTERFACE AND DATA VISUALIZATION

<https://cartograplant.org/>

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



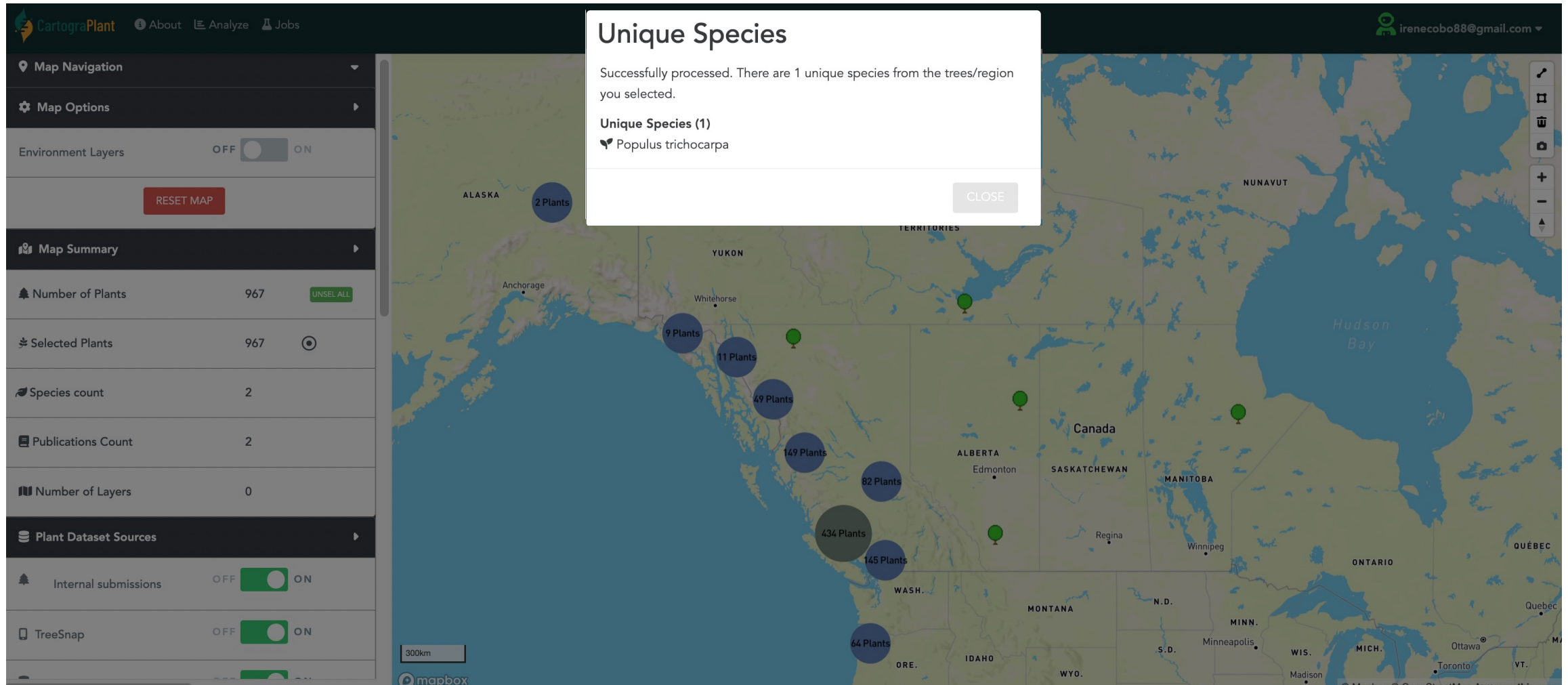
**DATA SELECTION FOR ANALYSIS**



# INTERFACE AND DATA VISUALIZATION

<https://cartograplant.org/>

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

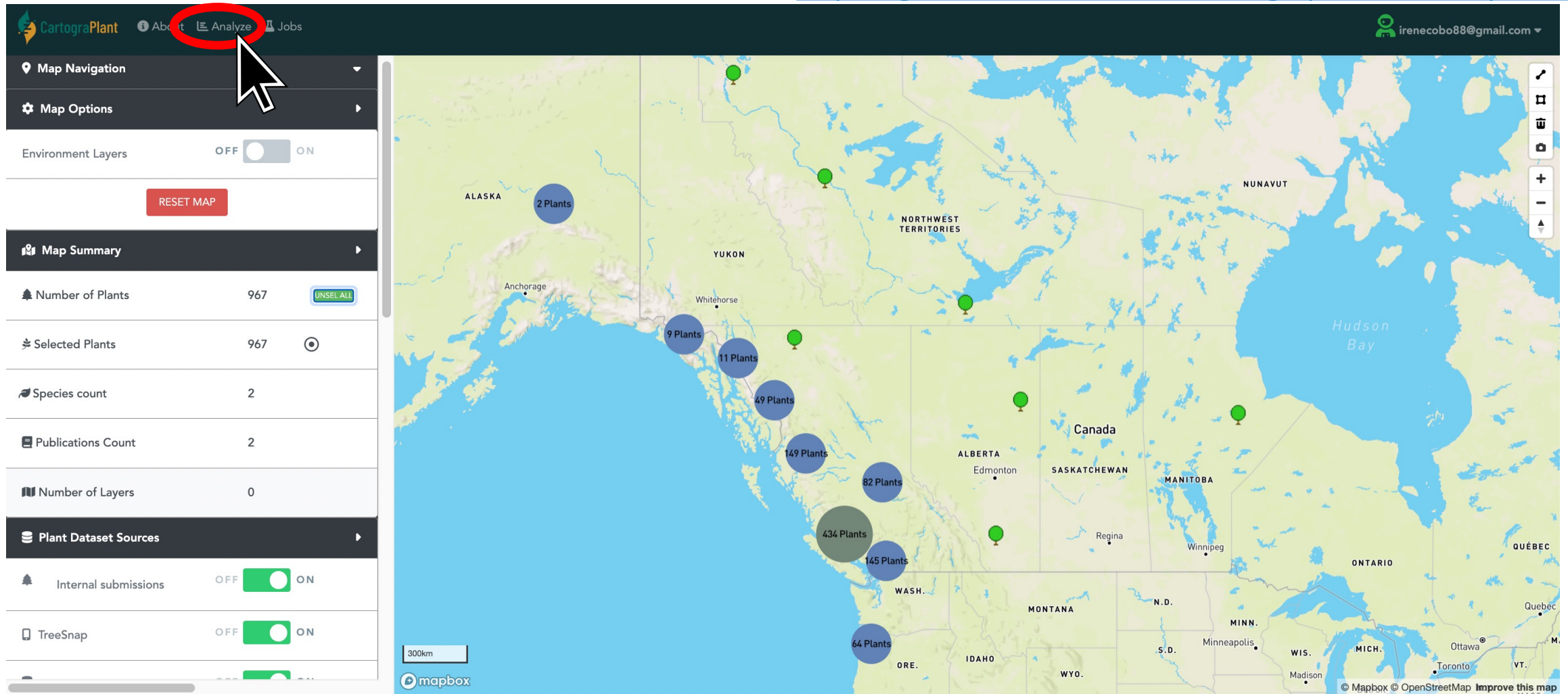


## DATA SELECTION FOR ANALYSIS

# 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 'Hello irenecobo88@gmail.com'. A central modal window is open, titled 'CartograPlant', and displays the following information:

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

The modal contains a series of tabs: Manage, Filter By Traits, Filter By Genotypes, Filtering & Imputation, Population Structure, Add environmental data, Run Analysis, and Summary and Confirm. Below the tabs, a 'Welcome!' message states: '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 message follows: '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.' Below this, there is a 'Select workspace' dropdown menu with 'Test' selected and a 'CREATE NEW WORKSPACE' button. The 'MANAGE WORKSPACE FILES' section indicates 'This workspace currently contains no files'. At the bottom right of the modal are 'NEXT' and 'CLOSE' buttons. The background shows a map of Canada with plant locations marked: 263 Plants near Seattle, 50 Plants near WASH., and 2 Plants near Regina. The left sidebar contains various filters and counts, including 'Plant Dataset Sources', 'Filters', 'Number of Plants', 'Selected Plants', 'Species count', 'Publications Count', and 'Number of Layers'.

## 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 contains the following text and buttons:

- Welcome!
- Your unique analysis ID: **1015**
- 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**
- This workspace currently contains no files
- 
- 

The background of the interface shows a map with several plant locations marked by circles, with labels such as "263 Plants", "50 Plants", and "2 Plants". The map also shows geographical features like "Canada" and "SASKATCHEWAN".

## DATA ANALYSIS: WORKSPACE CREATION



The screenshot displays the CartograPlant 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 modal window is open, showing the analysis details for ID 1015. 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 main content area of the modal includes a welcome message, the analysis ID (1015), and the analysis name 'PAG30 workshop test' (highlighted with a red box). The analysis type is set to 'Genotype x Phenotype x Environmental'. There is an 'UPDATE' button. Below this, there is a section for selecting a workspace, with a 'Test' dropdown and a 'CREATE NEW WORKSPACE' button. The modal also contains 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 background of the application shows a map of Canada with plant locations marked by circles, such as '263 Plants' near Seattle and '2 Plants' near Regina.

## DATA ANALYSIS: WORKSPACE CREATION

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 Co

CartograPlant

Environment Layers

Map Summary

Number of Plants

Selected Plants

Species count

Publications Count

Number of Layers

Plant Dataset Source

Filters

AND OR ADD RULE ADD GROUP

CartograPlant

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

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

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

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

Select workspace Test

Workspace name Test2 Create

MANAGE WORKSPACE FILES  
This workspace currently contains no files

CREATE NEW WORKSPACE

NEXT

CLOSE

Canada  
SASKATCHEWAN  
Regina  
Seattle WASH.  
263 Plants  
50 Plants  
2 Plants

## DATA ANALYSIS: WORKSPACE CREATION

The screenshot displays 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 irenecobo88@gmail.com. The main workspace configuration area is titled 'CartograPlant' and shows the following details:

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

The workspace configuration area includes a 'Manage' tab and several other tabs: 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'. There is an 'UPDATE' button next to the analysis type dropdown.

Below this, there is a section for workspace creation. It states: '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.' There is a 'Select workspace' dropdown menu with 'Test' selected and a 'CREATE NEW WORKSPACE' button. Below this, there is a 'Workspace name' input field with 'Test2' entered and a 'Create' button.

At the bottom of the workspace configuration area, there is a 'MANAGE WORKSPACE FILES' section with the text 'This workspace currently contains no files'. A red box highlights the 'NEXT' button, which is being clicked by a mouse cursor. There is also a 'CLOSE' button below the 'NEXT' button.

## DATA ANALYSIS: WORKSPACE CREATION



The screenshot displays the CartograPlant web interface. At the top, a red box highlights the analysis details: Analysis ID: 1015, Phenotypes: 3, Genotypes: 0, and Environmental layers: 0. The main navigation bar includes options like 'Filter By Traits', 'Filter By Genotypes', and 'Run Analysis'. The 'Filter By Traits' section shows two studies, TGDR674 and TGDR675, with filters for 'diameter at breast height', 'plant height', and 'elevation'. A histogram for 'plant height' is shown with a blue arrow pointing to the 'Adjust thresholds' button. A dropdown menu for units (meter, centimeter) is open, with a green box indicating 'Overlaps with all studies'. Below this, a histogram for 'elevation' is shown with a red box around the 'Save adjustment' button. To the right, a PCA plot is displayed with 'standardized PC2 (82.1% explained var)' on the y-axis and 'standardized PC1 (63.1% explained var)' on the x-axis. A map of Canada is visible on the far right.

## DATA ANALYSIS: PHENOTYPIC TRAITS SELECTION/VISUALIZATION

**CartograPlant**

Analysis ID: 1360  
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** **TGDR675**  
Overlapping genotypes: 24834

**Insights**  
We analyzed 2 studies and discovered the following:  
There are a total of 24834 SNP overlaps across all 2 studies  
 TGDR674 and TGDR675 have shared datasets

- SNP overlaps between TGDR674, TGDR675: 24834
- TGDR674 has 0 SNPs
- TGDR675 has 160 SNPs

TGDR674 has 0 non-overlaps  
TGDR675 has 0 non-overlaps

**TGDR674 0** | **24834** | **TGDR675 0**

NEXT

## 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 right, analysis statistics are shown: Analysis ID: 1360, Phenotypes: 3, Genotypes: 0, and Environmental layers: 0. A navigation bar contains tabs for 'Manage', 'Filter By Traits', 'Filter By Genotypes', 'Filtering & Imputation', 'Population Structure', 'Add environmental data', and 'Run Analysis'. The 'Filter By Genotypes' tab is active and highlighted with a red box. Below this tab, the text reads: 'Filter By Genotypes', '2 studies detected based on the trees you selected on the map', and two buttons labeled 'TGDR674' and 'TGDR675'. Below the buttons, it states 'Overlapping genotypes: 24834'. To the right of this text is a Venn diagram with two overlapping circles. The left circle is light blue and labeled 'TGDR674 0'. The right circle is light orange and labeled 'TGDR675 0'. The overlapping area in the center is shaded brown and labeled '24834'. Below the Venn diagram is a 'NEXT' button. On the left side of the interface, there is a sidebar with various menu items: 'Dashboard', 'Content', 'Map Options', 'Environment Layers', 'Map Summary', 'Number of Plants', 'Selected Plants', 'Species count', 'Publications Count', and 'Number of Layers'. The CartograPlant logo is visible at the top left of the main content area.

**DATA ANALYSIS: GENOTYPIC DATA SELECTION/VISUALIZATION**



# DATA ANALYSIS

<https://cartograplant.org/>

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

CartograPlant

Analysis ID: 1360  
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 TGDR675  
Overlapping genotypes: 24834

**Insights**  
We analyzed 2 studies and discovered the following:  
There are a total of 24834 SNP overlaps across all 2 studies  
 TGDR674 and TGDR675 have shared datasets

- SNP overlaps between TGDR674, TGDR675: 24834
- TGDR674 has 0 SNPs
- TGDR675 has 160 SNPs

TGDR674 has 0 non-overlaps  
TGDR675 has 0 non-overlaps

TGDR674 0 | 24834 | TGDR675 0

NEXT

**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 left is the CartograPlant logo. A navigation sidebar on the left includes options like 'Dashboard', 'Content', 'Map Options', 'Environment Layers', 'Map Summary', 'Number of Plants', 'Selected Plants', 'Species count', 'Publications Count', and 'Number of Layers'. The main content area shows analysis results for two studies, TGDR674 and TGDR675. A summary box indicates '2 studies detected based on the trees you selected on the map' and 'Overlapping genotypes: 24834'. A Venn diagram visualizes the overlap, with 0 SNPs unique to TGDR674, 0 SNPs unique to TGDR675, and 24834 shared SNPs. Below the Venn diagram, an 'Insights' section states: 'We analyzed 2 studies and discovered the following: There are a total of 24834 SNP overlaps across all 2 studies. TGDR674 and TGDR675 have shared datasets. SNP overlaps between TGDR674, TGDR675: 24834. TGDR674 has 0 SNPs. TGDR675 has 160 SNPs. TGDR674 has 0 non-overlaps. TGDR675 has 0 non-overlaps.' A red box highlights the last four lines of the insights. At the top right, a summary box lists: 'Analysis ID: 1360', 'Phenotypes: 3', 'Genotypes: 24834', and 'Environmental layers: 0'. A 'NEXT' button is located at the bottom right of the main content area.

CartograPlant

Analysis ID: 1360  
Phenotypes: 3  
Genotypes: 24834  
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 TGDR675  
Overlapping genotypes: 24834

Insights  
We analyzed 2 studies and discovered the following:  
There are a total of 24834 SNP overlaps across all 2 studies  
TGDR674 and TGDR675 have shared datasets  
SNP overlaps between TGDR674, TGDR675: 24834  
TGDR674 has 0 SNPs  
TGDR675 has 160 SNPs  
TGDR674 has 0 non-overlaps  
TGDR675 has 0 non-overlaps

TGDR674 0 24834 TGDR675 0

NEXT

**DATA ANALYSIS: GENOTYPIC DATA SELECTION/VISUALIZATION**

**CartograPlant**

Analysis ID: 1360  
Phenotypes: 3  
Genotypes: 24834  
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 TGDR675  
Overlapping genotypes: 24834

**Insights**  
We analyzed 2 studies and discovered the following:  
There are a total of 24834 SNP overlaps across all 2 studies  
TGDR674 and TGDR675 have shared datasets  
 SNP overlaps between TGDR674, TGDR675: 24834  
 TGDR674 has 0 SNPs  
 TGDR675 has 160 SNPs  
TGDR674 has 0 non-overlaps  
TGDR675 has 0 non-overlaps

TGDR674 0 24834 TGDR675 0

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

NEXT

## 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

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 '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 main configuration section is titled 'SNP QUALITY FILTERING STEP 3: MINIMUM QUALITY SCORE ANALYSIS CONFIGURATION'. It includes a 'Minimum quality score' field, an 'Output of Step 2' section with a gear icon and the instruction 'Select the file to use.', and a 'Workspace available file:' dropdown menu. A blue button labeled 'INITIATE ANALYSIS JOB' is at the bottom left. On the right side, a red box highlights a refresh icon (a blue square with a white circular arrow) and a mouse cursor is pointing at it. A text note next to it says 'If files don't appear, please try refreshing the workflow'. The left sidebar contains various map-related options like 'Map Options', 'Environment Layers', 'Map Summary', 'Number of Plants', 'Selected Plants', 'Species count', 'Publications Count', and 'Number of Layers'. A map of the world is visible on the right side of the interface.

## 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

CartograPlant

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

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

🌟 VCF files successfully uploaded to workspace

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

Quality filtering method: SNP Quality Filtering Step 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

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

☀️ 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 ago)

INITIATE ANALYSIS JOB

## DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION



The screenshot shows the CartograPlant web interface. At the top, there is a navigation bar with various menu items like Dashboard, Content, Structure, etc. The main content area is titled 'Filtering & Imputation' and displays analysis statistics: Analysis ID: 1053, Phenotypes: 3, Genotypes: 24887, and Environmental layers: 0. Below this, there are several tabs: 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 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' and includes a 'Minimum quality score' field. Under 'Output of Step 2', there is 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 is also visible with the text 'If files don't appear, please try refreshing the workflow'.

## 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 a 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'. 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 links for 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 '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 are visible: 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)'. The main configuration area is titled 'SNP QUALITY FILTERING STEP 9: LINKAGE DISEQUILIBRIUM (RETAIN ONLY INDEPENDENT SNPS FOR FURTHER POPULATION STRUCTURE ANALYSIS) CONFIGURATION'. It includes a sub-header 'Linkage disequilibrium (retain only independent SNPs for further population structure analysis)'. A section for 'Output of Step 8 (family-based studies) or 7 (non-family-based studies)' contains a 'Select the file to use.' button and a dropdown menu labeled 'Workspace available files: QCfiltering\_3: Minimum quality score'. A red rectangle highlights this dropdown menu. To the right of the dropdown, there is a note: 'If files don't appear, please try refreshing the workflow' with a refresh icon. At the bottom left of the configuration area, there is a blue button labeled 'INITIATE ANALYSIS JOB'. The left sidebar contains various navigation options like 'Map Options', 'Environment Layers', 'Map Summary', 'Number of Plants', 'Selected Plants', 'Species count', 'Publications Count', and 'Number of Layers'. A map 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 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 shows the 'Filtering & Imputation' step of an analysis job (ID: 1053) with 3 phenotypes and 24,887 genotypes. A notification indicates that VCF files were successfully uploaded. Below this, two green checkmarks confirm the presence of genotype VCF files for TGD674 and TGD682. The 'Quality filtering method' is set to 'SNP Quality Filtering Step 9: Linkage disequilibrium (retain only independent SNPs for further population structure analysis)'. A configuration box for this step is visible, with a sub-header 'SNP QUALITY FILTERING STEP 9: LINKAGE DISEQUILIBRIUM (RETAIN ONLY INDEPENDENT SNPs FOR FURTHER POPULATION STRUCTURE ANALYSIS) CONFIGURATION'. The configuration includes a gear icon for settings, a note about the output of Step 8, and a file selection dropdown menu showing 'QCfiltering\_3: Minimum quality score'. A blue 'INITIATE ANALYSIS JOB' button is highlighted with a red box and a mouse cursor. A refresh button is also present with the text 'If files don't appear, please try refreshing the workflow'.

## 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

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)

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  
Invocation ID: 49ec47cc9a1bdc41  
Status: Submitted

**Output Results (running)**

- Awaiting File QCfiltering\_9: Linkage disequilibrium (0 bytes)

NEXT

## DATA ANALYSIS: QUALITY FILTERING AND SNP IMPUTATION

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 '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 map of the Southern Hemisphere is visible, with labels for South Pacific Ocean, South Atlantic Ocean, and Indian Ocean. A sidebar on the left 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**

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

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

Species count	2
Publications Count	2
Number of Layers	0

## DATA ANALYSIS: POPULATION STRUCTURE CALCULATION/VISUALIZATION



Analysis ID: 1360  
Phenotypes: 3  
Genotypes: 24834  
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

GENERATE FAST STRUCTURE

Population Structure successfully completed! You can find the outputs under the Manage tab.

NEXT

CLOSE

Species count	2
Publications Count	2
Number of Layers	0



## DATA ANALYSIS: POPULATION STRUCTURE CALCULATION/VISUALIZATION

The screenshot shows the CartograPlant web 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 irenecobo88@gmail.com. The main content area displays the analysis ID 1015, with 3 phenotypes, 24834 genotypes, and 0 environmental layers. 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 'Untitled' and the analysis type is 'Genotype x Phenotype x Environmental'. There is an 'UPDATE' button. Below this, a message states: '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.' There is a 'Select workspace' dropdown menu with 'Test' selected and a 'CREATE NEW WORKSPACE' button. Below this, there is a section titled 'MANAGE WORKSPACE FILES' with a list of files. The file 'Analysis 1360: PopStruct Panel Final (a few seconds ago)' is highlighted with a red box and a mouse cursor. To the right of the file list, there are 'NEXT' and 'CLOSE' buttons. The background shows a map of Canada with 'SASKATCHEWAN' and 'Regina' labeled.

## DATA ANALYSIS: WORKSPACE CREATION



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'. The main content area is titled 'Analysis ID: 1015' and shows 'Phenotypes: 3', 'Genotypes: 24834', and 'Environmental layers: 0'. A sidebar on the left contains various filters and counts, including 'Species count', 'Publications Count', 'Number of Layers', and 'Plant Dataset Sources'. The main panel has several tabs: 'Manage', 'Filter By Traits', 'Filter By Genotypes', 'Filtering & Imputation', 'Population Structure' (selected), 'Add environmental data', 'Run Analysis', and 'Summary and Confirm'. Under the 'Population Structure' tab, there is a 'Number of populations' input field set to '2', a 'Select VCF file' dropdown menu showing 'Analysis 1016: TGDR725 Filtered SNPs VCF (14 minutes ago)', and a 'GENERATE FAST STRUCTURE' button. Below this, there are four horizontal bar charts for K=2, K=3, K=4, and K=5, showing the genetic composition of individuals across three populations: WS (green), QI (red), and Tp (blue). A legend on the right identifies these populations. A 'NEXT' button is visible to the right of the charts. A mouse cursor is pointing at the 'NEXT' button. At the bottom of the screenshot, a white text box contains the message: 'Population structure calculation (PCA) and visualization (fastSTRUCTURE, PCA) available soon!!!'.

# DATA ANALYSIS: POPULATION STRUCTURE CALCULATION/VISUALIZATION

The screenshot displays the CartograPlant web application interface. A modal window is open for configuring an analysis. The modal title bar shows 'Analysis ID: 1015', 'Phenotypes: 3', 'Genotypes: 24834', 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'. A section titled 'Choose environmental layers' 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 are 'NEXT' and 'CLOSE' buttons. The background shows a map of Canada with several blue circular markers indicating plant locations, labeled with counts: '65 Plants', '43 Plants', '263 Plants', '50 Plants', and '2 Plants'. The map also shows provincial boundaries for 'BRITISH COLUMBIA', 'ALBERTA', and 'SASKATCHEWAN', and cities like 'Edmonton' and 'Regina'. The top navigation bar of the application 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'.

## DATA ANALYSIS: ENVIRONMENTAL DATA SELECTION/ MULTICOLLINEARITY

The screenshot displays the 'Choose environmental layers' section of the CartograPlant application. The interface includes 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'. The 'Add environmental data' tab is active. Below the navigation bar, there are buttons for 'PRECACHE VALUES' and 'GATHER AND UPLOAD TO WORKSPACE'. The main content area lists various environmental layers with checkboxes and labels:

- 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)

A search bar with a magnifying glass icon and a 'SELECT ALL' button are also visible. A mouse cursor is pointing at the search bar. On the right side, a map of Canada is partially 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 'PROPERTY' label. The 'PROPERTY' labels are color-coded: green for 'PROPERTY' and orange for 'CLASSIFICATION'. The selected 'PROPERTY' 'Chilling Days Below 0 °C CWNA (DD0)' is highlighted with a mouse cursor. To the right of the selection list is a 'PRECACHE VALUES' button and a 'GATHER AND UPLOAD TO WORKSPACE' button. Below the latter button, it states 'Found 145 Mean Annual Precipitation CWNA (MEAN\_ANNUAL\_PREC) values.' and provides two empty input fields. At the bottom right, 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 navigation sidebar with options like 'Dashboard', 'Map Summary', and 'Filters'. The main area features a top navigation bar with tabs: 'Manage', 'Filter By Traits', 'Filter By Genotypes', 'Filtering & Imputation', 'Population Structure', 'Add environmental data' (active), 'Run Analysis', and 'Summary and Confirm'. Below this, a 'Choose environmental layers' section lists various categories and properties, such as 'Climatic variables (World, ClimateWNA)' and 'Chilling Days Below 0 °C CWNA (DD0)'. A 'PRECACHE VALUES' button is visible. To the right, a message states: 'Found 145 Mean Annual Precipitation CWNA (MEAN\_ANNUAL\_PREC) values.' Below this message are four small plots: a histogram of 'Annual Temperature CWNA, MEAN\_ANNUAL\_TEMP', a scatter plot of 'Annual Temperature CWNA, MEAN\_ANNUAL\_TEMP' vs 'Annual Precipitation CWNA, MEAN\_ANNUAL\_PREC', and two histograms for 'Annual Precipitation CWNA, MEAN\_ANNUAL\_PREC'. A large number '0.29' is displayed in the top right plot area. A map of Canada is visible on the far right.

## 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

Step 1 - Select workflow  
GWAS with EMMAX Multiple Testing Correction REFRESH



Step 2 - Setup analysis  
✓ GWAS with EMMAX NEXT

CLOSE

Species count 2  
Publications Count 2  
Number of Layers 0  
Plant Dataset Sources  
Filters

AND OR ADD RULE ADD GROUP

25 Plants  
65 Plants  
43 Plants  
50 Plants  
BRITISH COLUMBIA

 + 

The other analytic workflows available soon!

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

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

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

Step 1 - Select workflow  
Step 2 - Setup analysis

GWAS with EMMAX

**GWAS WITH EMMAX ANALYSIS CONFIGURATION**

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

Genotype file containing SNPs in vcf or vcf.gz format

Select the file to use.

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

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

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)

## DATA ANALYSIS: SELECT/RUN ANALYTIC WORKFLOWS

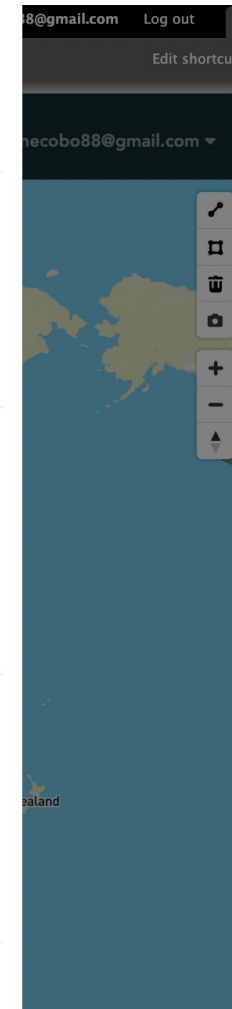
**Select the file to use.**  
Seleccionar archivo Ninguno ...hivo selec. Upload to workspace  
Workspace available files: Analysis 1053: TGDR674 Filtered SNPs VCF (8 minutes ago) ✓

**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) ✓

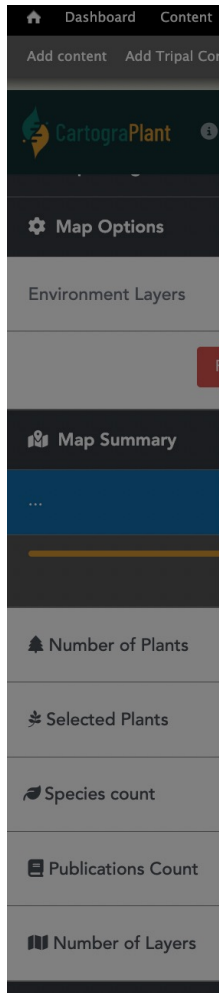
**Phenotype file:** First and second column, family and individual name in the same order as in the vcf file. Third column, phenotype values. Lacking data has to be typed as "NA"  
**Select the file to use.**  
Seleccionar archivo Ninguno ...hivo selec. Upload to workspace  
Workspace available files: 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



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) v

LD filtered vcf  
Select the file to use.  
Seleccionar archivo Ninguno ...hivo selec. Upload to workspace  
Workspace available files: Analysis 1053: Filtered Phenotypes - bud set - ADJUSTED (10 minutes ago) v

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 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) v

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

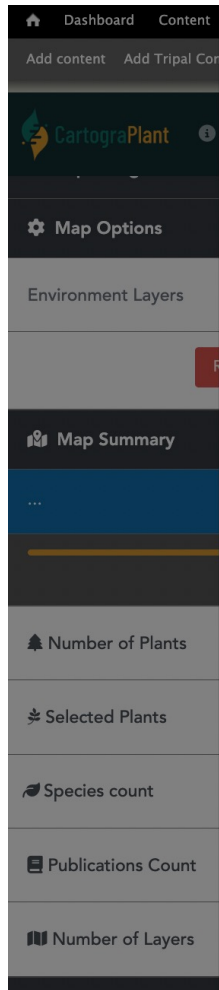
Select the file to use.  
Seleccionar archivo Ninguno ...hivo selec. Upload to workspace  
Workspace available files: Analysis 1053: Filtered Phenotypes - bud set - ADJUSTED (10 minutes ago) v

Initiate analysis job

*Note: A dropdown menu is open for the second 'LD filtered vcf' section, showing a list of files with 'QCfiltering\_9: Linkage disequilibrium' selected. A mouse cursor is pointing at the selected item.*



## 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

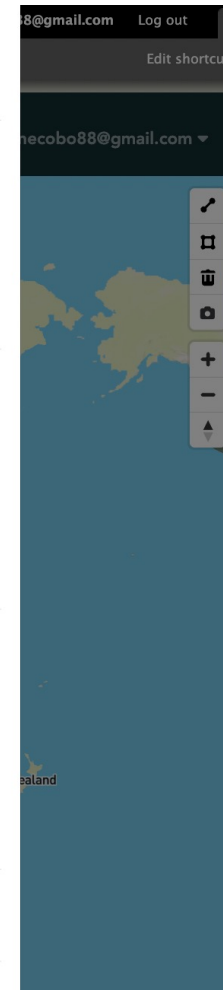
Seleccionar archivo Ninguno ...hivo selec. Upload to workspace  
Workspace available files: Analysis 1053: TGDR674 Filtered SNPs VCF (8 minutes ago)

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**

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

