

Tutorial to perform a Landraces ex-situ gap analysis using LGA toolbox.

Open the server.R file

Steps

1. Download LGA toolbox (DONE!)
2. Unzip LGA toolbox (DONE!)
3. Put your occurrence data in documents
4. Start your LGA toolbox
5. Create your outcomes folder
6. Preprocessing: Upload occurrences, Select areas, and curate coordinate
7. Modeling species distribution
8. Obtain gap collection metrics and maps
9. Validate approach

Please check your inputs

- LGA toolbox unzip
- Occurence data

The occurrences data file

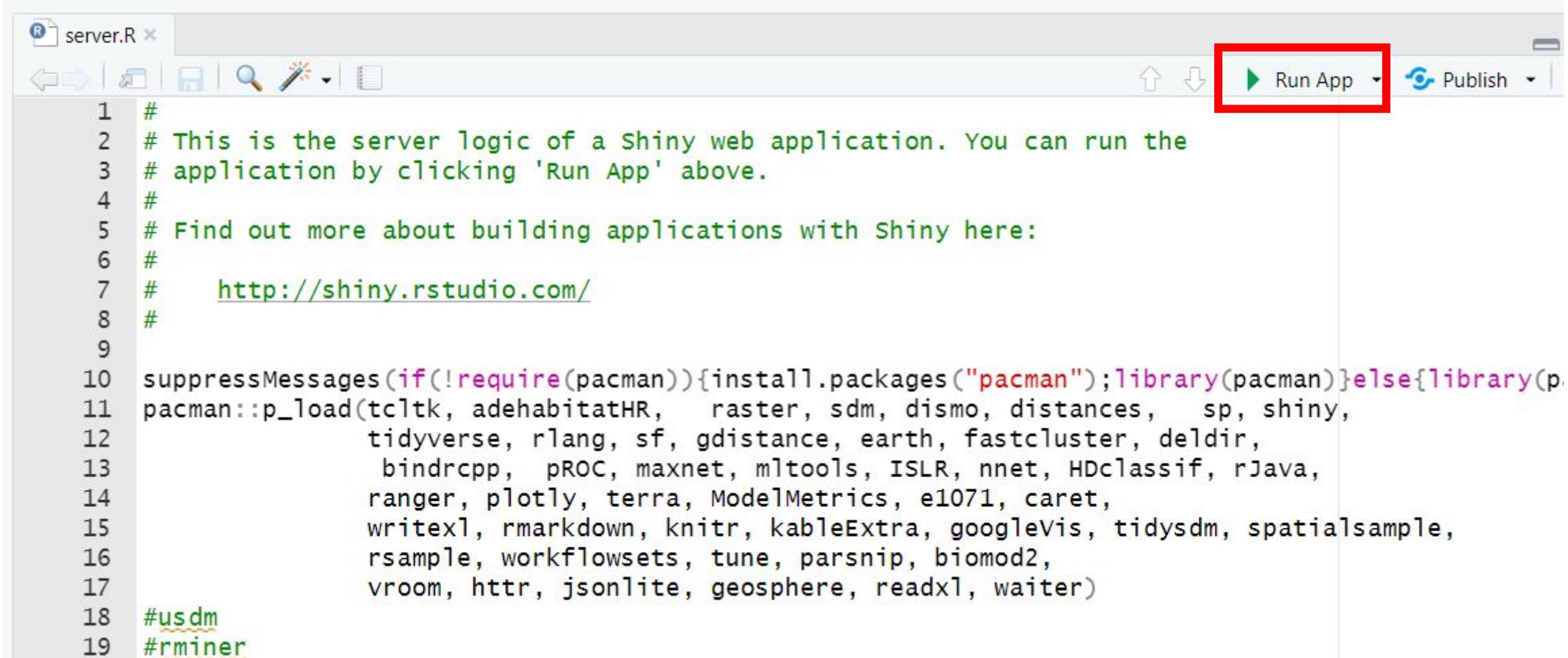
crop_name	Latitude	Longitude	status	source_db	database_id
triticum_turgidum	31	51	G	GENESYS	GENESYS_922
triticum_turgidum	31	51	G	GENESYS	GENESYS_935
triticum_turgidum	34.716667	32.533333	G	GENESYS	GENESYS_937
triticum_turgidum	34.783333	-5.6	G	GENESYS	GENESYS_948
triticum_turgidum	43.416389	10.583056	G	GENESYS	GENESYS_953
triticum_turgidum	44.066667	12.033056	G	GENESYS	GENESYS_954
triticum_turgidum	43.685278	12.633056	G	GENESYS	GENESYS_955
triticum_turgidum	40.693611	15.500278	G	GENESYS	GENESYS_958
triticum_turgidum	40.693611	15.500278	G	GENESYS	GENESYS_959
triticum_turgidum	40.733056	17.546667	G	GENESYS	GENESYS_960
triticum_turgidum	34.783333	-5.6	G	GENESYS	GENESYS_966

Species name	Decimal latitude	Decimal longitude	Germplasm:G Herbarium:H	DB Source	DB Source accession id
crop_name	Latitude	Longitude	status	source_db	database_id
triticum_turgidum	31	51	G	GENESYS	GENESYS_922
triticum_turgidum	31	51	G	GENESYS	GENESYS_935
triticum_turgidum	34.716667	32.533333	G	GENESYS	GENESYS_937
triticum_turgidum	34.783333		-5.6 G	GENESYS	GENESYS_948

Step 4. Start your LGA toolbox

Open the server.R file

Click on Run App

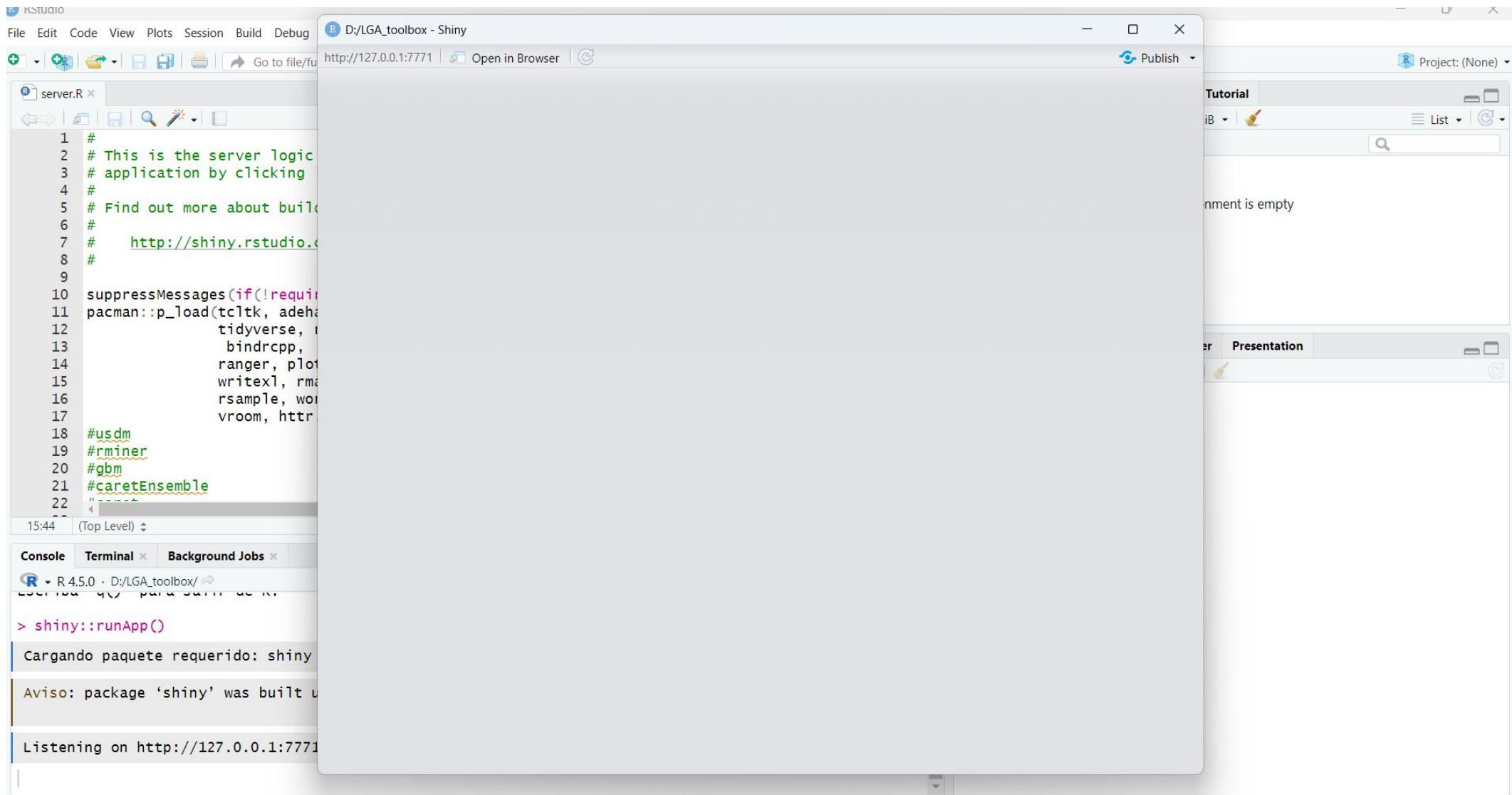


The screenshot shows the RStudio interface with the 'server.R' file open. The code in the editor is as follows:

```
1 #  
2 # This is the server logic of a Shiny web application. You can run the  
3 # application by clicking 'Run App' above.  
4 #  
5 # Find out more about building applications with Shiny here:  
6 #  
7 #     http://shiny.rstudio.com/  
8 #  
9  
10 suppressMessages(if(!require(pacman)){install.packages("pacman");library(pacman)}else{library(pac  
11 pacman::p_load(tcltk, adehabitatHR, raster, sdm, dismo, distances, sp, shiny,  
12                 tidyverse, rlang, sf, gdistance, earth, fastcluster, deldir,  
13                 bindrcpp, pROC, maxnet, mltools, ISLR, nnet, HDclassif, rJava,  
14                 ranger, plotly, terra, ModelMetrics, e1071, caret,  
15                 writexl, rmarkdown, knitr, kableExtra, googleVis, tidysdm, spatialsample,  
16                 rsample, workflowsets, tune, parsnip, biomod2,  
17                 vroom, httr, jsonlite, geosphere, readxl, waiter)  
18 #usdm  
19 #rminer
```

The 'Run App' button in the toolbar is highlighted with a red box.

A browser will be opened



Click on Open to charge the LGA toolbox in your browser

The screenshot shows the LGA Toolbox interface. At the top, there is a red box highlighting the 'Open in Browser' button. Below the header, there are tabs for 'Wizards', 'Gap scores', and 'Coord quality'. On the right, there are buttons for 'Working directory', 'Define study area', 'Covariates', and 'Passport data'. A central panel displays the 'Working directory set up' wizard. It has a title, a 'select folder' section with a 'No folder path' button and a 'Browse' button, and a 'Select' button at the bottom.

Working directory set up

The wizard will guide you through the process to set up working directories for the Gap analysis.

Landrace Gap Analysis App is an R-shiny application developed in R software version 4.2.2. This app uses local resources for processing, it will save and load file locally as it needs. To start the process user must select a folder where all inputs, intermediate-inputs and outputs will be stored.

1. Write the Crop name in the Crop name field.
2. Click the button “**Browse**” for the root folder in which the App directory system will be created.
3. Finally, click on “**Select**” button, this will create all necessary folder to store inputs, results, etc...

LGA Toolbox



Wizards

Gap scores

Coord quality

Analysis
menu

Working directory

Define study area

Covariates

Passport data

Working directory set up

select folder

No folder path selected

Browse

Select

Select folder where your input data is saved

Working directory set up

The wizard will guide you through the process to set up working directories for the Gap analysis.

Landrace Gap Analysis App is an R-shiny application developed in R software version 4.2.2. This app uses local resource processing, it will save and load file locally as it needs. To start the process user must select a folder where all inputs, intermediate-inputs and outputs will be stored.

1. Write the Crop name in the Crop name field.
 2. Click the button “**Browse**” for the root folder in which the App directory system will be created.
 3. Finally, click on “**Select**” button, this will create all necessary folder to store inputs, results, etc...
-

Instructions

User
options

Step 5. Create your outcomes folder

- Select a folder to save outcomes
- Write the crop name. This is the folder name with your outcomes

Working directory set up

Folder selected

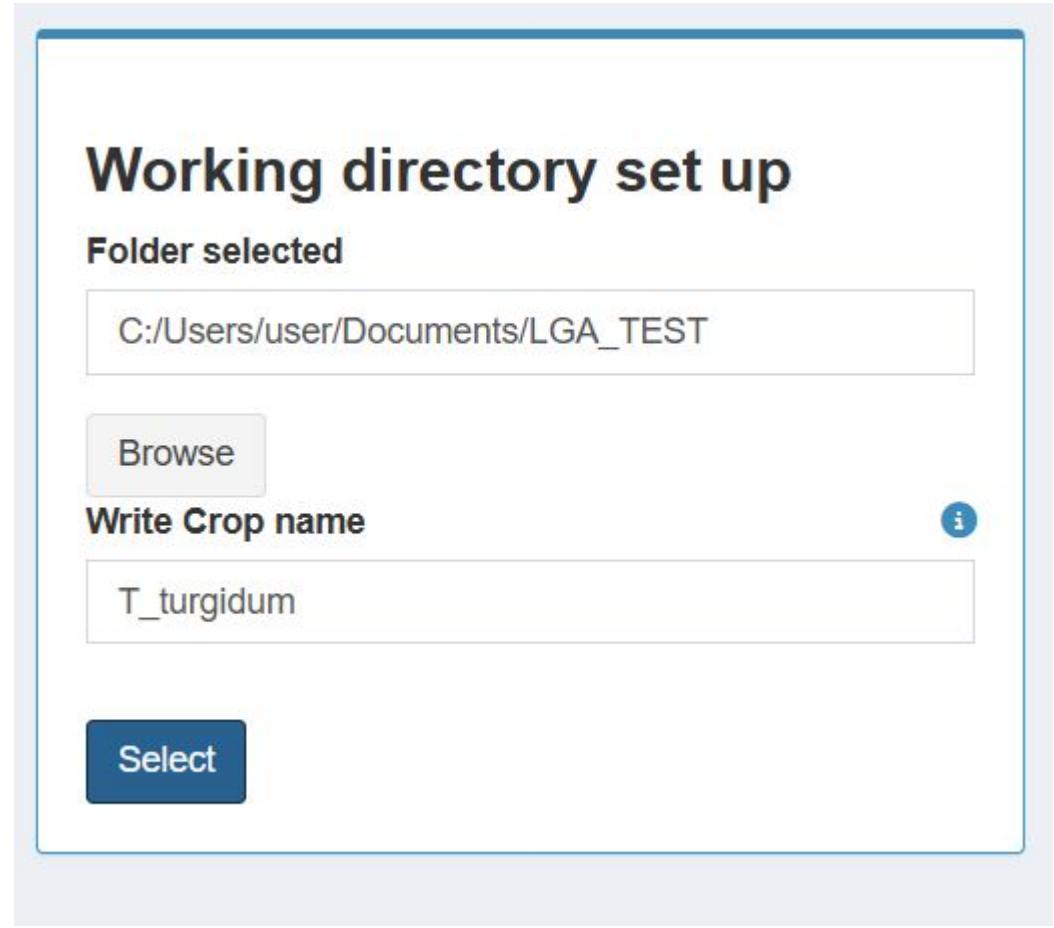
C:/Users/user/Documents/LGA_TEST

Browse

Write Crop name i

T_turgidum

Select



You created your folders!

Working directory set up

Folder selected

C:/Users/user/Documents/LGA_TE

Browse

Write Crop name

T_turgidum

Selected

Working directory set up

Set up working directories for the Gap analysis.



Success !!

Main directories were successfully created.

Ok

Your canvas



Working directory

Define study area

Covariates

Passport data

Working directory set up

Folder selected

C:/Users/user/Documents/LGA_TEST

Browse

Write Crop name

T_turgidum

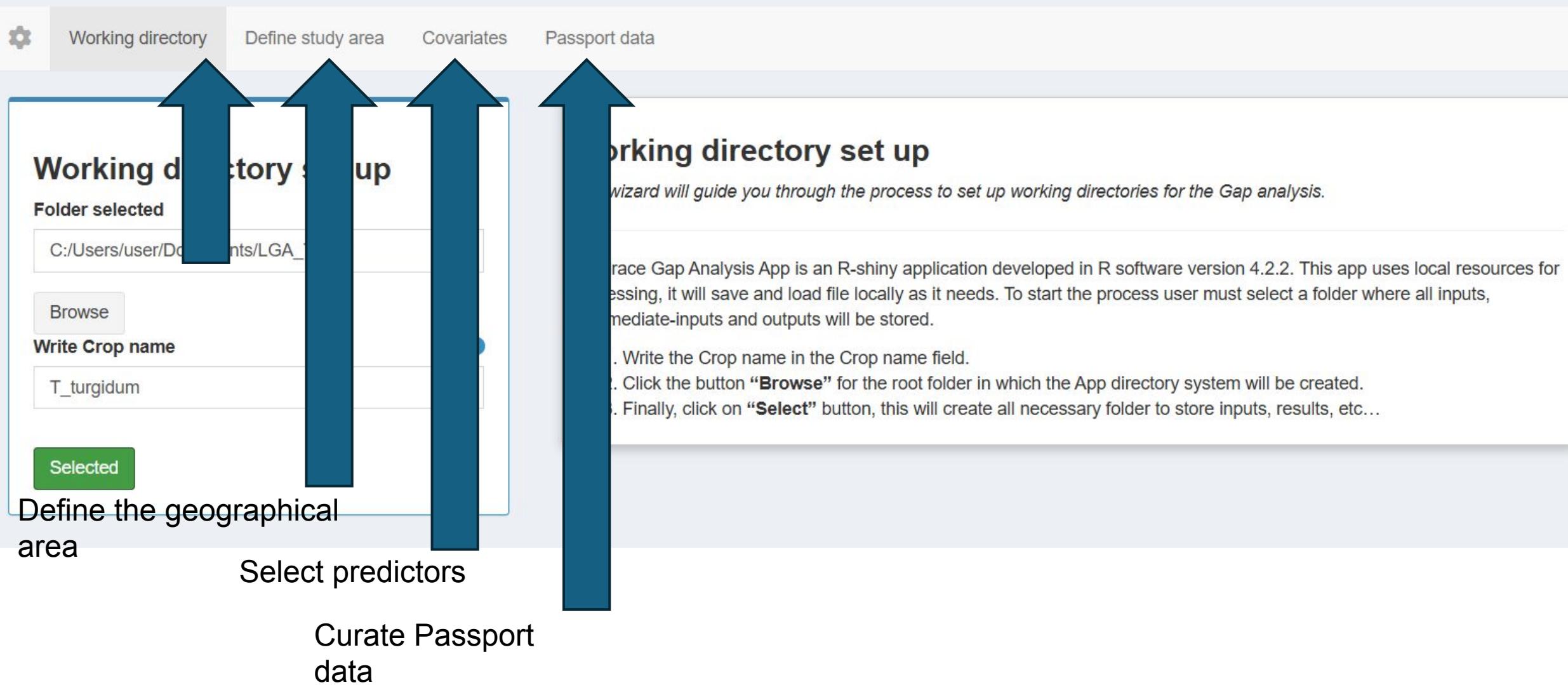
Selected

Working directory set up

The wizard will guide you through the process to set up working directories for the Gap analysis.

Landrace Gap Analysis App is an R-shiny application developed in R software version 4.2.2. This app uses local resources for processing, it will save and load file locally as it needs. To start the process user must select a folder where all inputs, intermediate-inputs and outputs will be stored.

1. Write the Crop name in the Crop name field.
2. Click the button “Browse” for the root folder in which the App directory system will be created.
3. Finally, click on “Select” button, this will create all necessary folder to store inputs, results, etc...



Step 6. Preprocessing:
Upload occurrences, Select
areas, and curate coordinate

Lets try select Northern Africa
countries for the gap analysis

Geographic area

Create Import Select

Select one region:

---Custom region---

Countries selected:

- Algeria
- Egypt
- Western Sahara
- Libyan Arab Jamahiriya
- Morocco
- Mauritania
- Tunisia

Write file name

Africa_test

Create

Geographic area selector wizard

the wizard will support you to define the strategic areas/regions to analyze.

Select a pre-defined continental mask from the, “Region Menu” or create a customized geographic area clicking over the countries in the map. Finally set a name for the defined region.

file output: input_data/“mask_name”.tif



Click on create (Be patient)

Geographic area

Create **Import** **Select**

Select one region:

---Custom region---

Countries selected:

- Algeria
- Egypt
- Western Sahara
- Libyan Arab Jamahiriya
- Morocco
- Mauritania
- Tunisia

Write file name

Africa_test

Mask created

Geographic area selector wizard

the wizard will support you to define the strategic areas/regions to analyze.

Select a pre-defined continental mask from the, “Region Menu” or create a customized geographic mask by selecting countries in the map. Finally set a name for the defined region.

file output: input_data/“mask_name”.tif

The map displays a yellow-shaded region covering parts of North Africa and the Middle East, specifically encompassing Algeria, Libya, and Egypt. The map also shows the outlines of other countries like Morocco, Mauritania, and the surrounding regions. Labels are provided in both English and Arabic for most countries shown.

Click on load

The screenshot shows a software interface with a top navigation bar containing tabs: 'Working directory', 'Define study area', 'Covariates', and 'Passport data'. The 'Covariates' tab is active. Below the navigation bar, there are two main sections. On the left, a blue-bordered box titled 'Auxiliary files' contains the text 'Add external/auxiliary raster files:' followed by a dropdown menu set to 'No' and a 'Load' button. On the right, a larger section titled 'Import external/auxiliary data' with the subtitle 'This wizard allows you to import external raster files relevant for the gap analysis' provides detailed information about the data sources used for the analysis. It lists three categories: Climate approach, Topography, and anthropogenic and/or socio-economic factors, each with a list of specific datasets and their sources.

Auxiliary files

Add external/auxiliary raster files:

No

Load

Import external/auxiliary data

This wizard allows you to import external raster files relevant for the gap analysis

Gridded data comes from different sources encompassing different perspectives, it has to be in a 2.5 arc-min spatial or 5km resolution, below are specified the source for each data:

- **Climate approach:** We have use 19 bioclimatic variables at 2.5 arc-min spatial resolution from [WoldClim](#). Raster files for an additional 16 climate variables were also used at the same spatial resolution from the Environmental Rasters for Ecological Modeling [ENVIREM](#) database.
- **Topography:** To characterize topography, we used elevation above sea level from the Shuttle Radar Topography Mission [SRTM](#), dataset of the CGIAR-Consortium on Geospatial Information (CSI) portal. Climate and Topography data will be downloaded in the path `.../input_data/generic_rasters/world/`
- **anthropogenic and/r socio-economic:** namely, crop yield, total harvested area, total crop production, percentage of area under irrigation, population accessibility, distance to rivers, and distance to ancient human settlements.
 - Crop yield, harvested area and crop production data were gathered from the, [MapSPAM](#) , 2005 database.
 - Percentage of area under irrigation from the, [FAO](#), Global Map of Irrigation Areas.
 - Population accessibility from the Global Map of Accessibility published by the, [Joint Research Center](#).
 - Distance to rivers map was created by computing the distance from each pixel (in a 2.5 arc-min grid) to the closest river.
 - Geographic distance to the primary gene pool wild relatives was created by computing the distance between every pixel (in a 2.5 arc-min grid) to the closest known observation of any wild relative within the primary gene pool.

All datasets will be downloaded in the path `.../global_data/`, After all main inputs are stored in their respective folder, you will be able to crop each input according to the raster mask already created.

Find your dataset

Data base set up

1. Select .cvs database

No file selected



2. Write response column number

1



Predict missing groups

Set up database

Description Parameters Preview data Results

Database set up Assistant

The assistant will support you to select, clean, and standardize your database (passport data) to a specific format supported by this application.

Important aspects:

- The database (passport data) need three or 4 columns specifying: races, latitude and longitude, and an optional column with the type of accession (H for Herbarium accession and G for Germplasm accessions).
- The only file type supported in this app is a CSV (comma separated values) format, once you have upload a valid file itn can be previewed below.
- Final output: CSV file with the class names, latitude, longitude and, all variables extracted from the raster files mentioned in “Download input”. Stored in
`.../results/“crop_name”/“region_name”/input_data/“crop_name_lv1_bd.csv”`

Data base set up

1. Select .csv database

triticum_turgidum_to_process.csv

Upload complete

2. Write response column number

1

Predict missing groups

[Set up database](#)

		Description	Parameters	Preview data	Results		
		Show 5 entries					
		crop_name	Latitude	Longitude	status	source_db	database_id
1	triticum_turgidum	31	51	G	GENESYS	GENESYS_922	
2	triticum_turgidum	31	51	G	GENESYS	GENESYS_935	
3	triticum_turgidum	34.716667	32.533333	G	GENESYS	GENESYS_937	
4	triticum_turgidum	34.783333	-5.6	G	GENESYS	GENESYS_948	
5	triticum_turgidum	43.416389	10.583056	G	GENESYS	GENESYS_953	

Showing 1 to 5 of 20 entries

Previous [1](#) [2](#) [3](#) [4](#) Next

23933

Rows

1

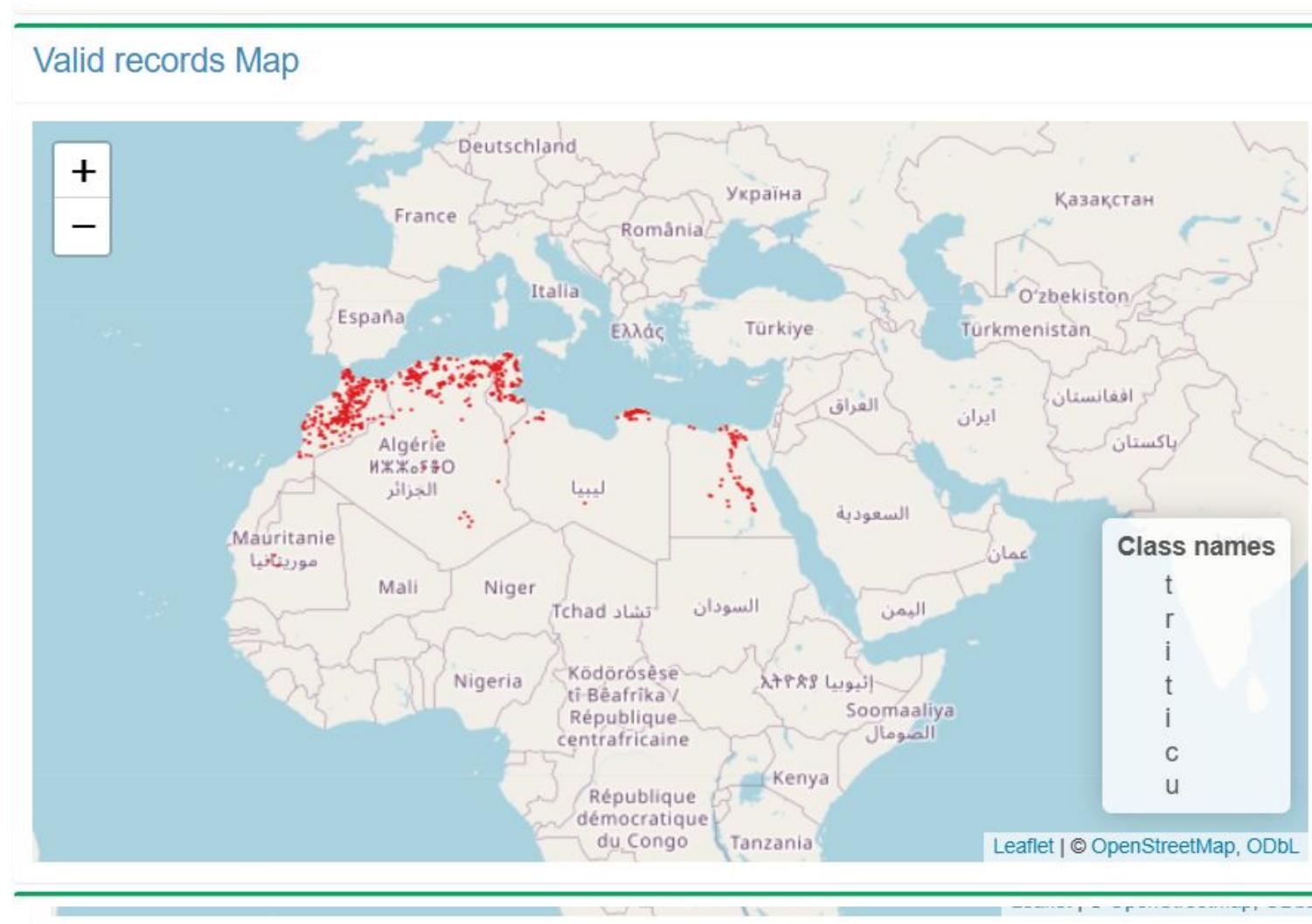
Groups

0%

Missing Values

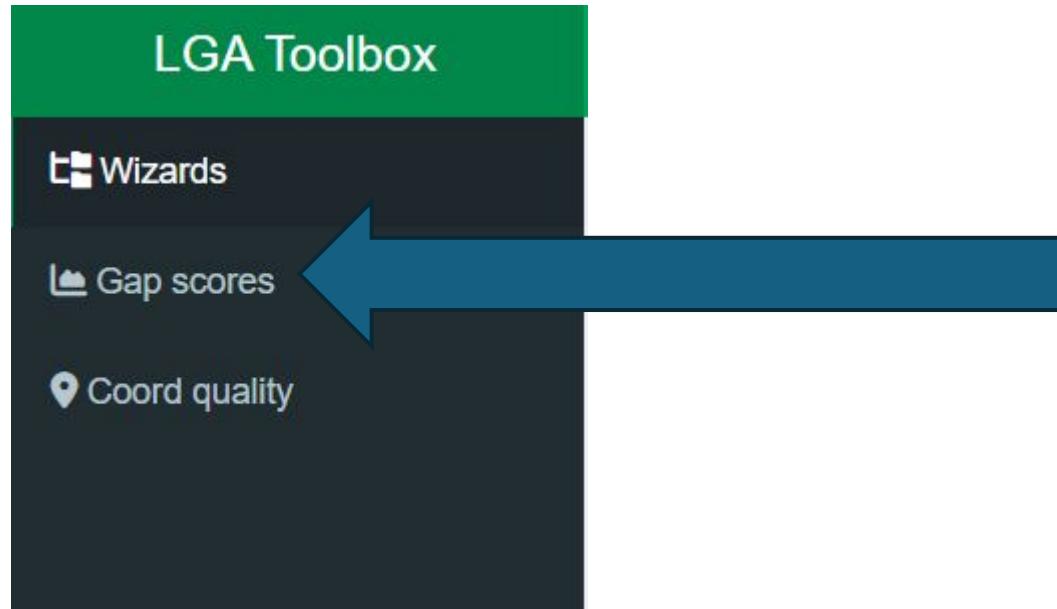
Description	Parameters	Preview data	Results
Data base output			
Show 5 entries			
Y	Latitude	Longitude	Accessibility
1 triticum_turgidum	34.783333	-5.6	36.4970703125
2 triticum_turgidum	34.816667	-5.55	4.96386957168579
3 triticum_turgidum	33.916667	-5.4	46.7060623168945
4 triticum_turgidum	34.066667	-5.6	52.9687538146973
5 triticum_turgidum	34.016667	-5.15	37.474609375
◀ ▶			
Showing 1 to 5 of 716 entries		Previous	1 2 3 4 5 ... 144 Next
Database Summary			
Counts			
Valid records Map			
ML performance metrics			

See your records



Step 7. Modeling species distribution

At this point you perform all preprocessing steps



Lets try perform SDM and obtain the gaps metrics

Steps

1. Download LGA toolbox (DONE!)
2. Unzip LGA toolbox (DONE!)
3. Put your occurrence data in documents
4. Start your LGA toolbox
5. Create your outcomes folder
6. Preprocessing: Upload occurrences, Select areas, and curate coordinate
7. Modeling species distribution
8. Obtain gap collection metrics and maps
9. Validate approach

This is your new menu



SDM modelling

Geo-Scores

Geo score assessment

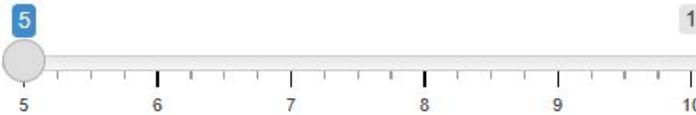
Landrace spatial distribution

Select Group/Class/Race to process:

triticum_turgidum

Model settings

Number of cross validation folds:



Run model

Description

Results

Spatial model distribution

To predict the probability of geographic occurrence for each landrace group, we generated MaxEnt models using the 'maxnet' R package. Group-specific spatial predictors were selected using a combination of the variance inflation factor (VIF) and a principal component analysis (PCA) to control for excessive model complexity and variable collinearity. We removed variables that did not contribute significantly to the variance in the PCA, defined as contributing less than 15% to the first component, and we further discarded variables with a VIF>10.

MaxEnt models were fitted through five-fold ($K=5$) cross-validation with 80% training and 20% testing. For each fold, we calculated the area under the receiving operating characteristic curve (AUC), sensitivity, specificity and Cohen's kappa as measures of model performance. To create a single prediction that represents the probability of occurrence for the landrace group, we computed the median across K models. Geographic areas in the form of pixels with probability values above the maximum sum of sensitivity and specificity were treated as the final area of predicted presence.





SDM modelling

Geo-Scores

Geo score assessment

Landrace spatial distribution

Select Group/Class/
Species to process:

triticum_turgidum

Model settings

Number of cross validation folds:

5

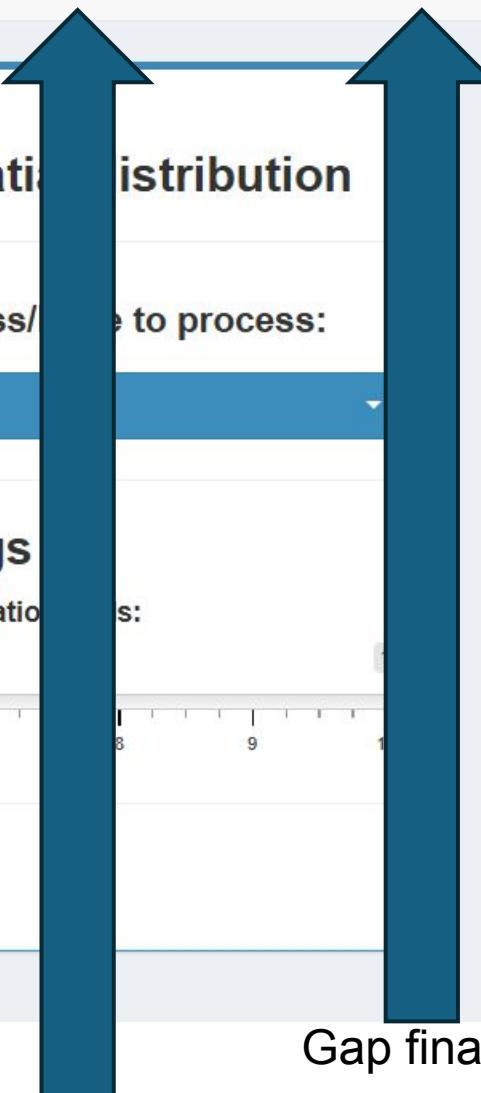


Run model

SDMs
parameters

Gap
scores

Gap final
scores



Description

Results

Spatial model distribution

To predict the probability of geographic occurrence for each landrace group, we generated MaxEnt models using the 'maxnet' R package. Group-specific spatial predictors were selected using a combination of the variance inflation factor (VIF) and a principal component analysis (PCA) to control for excessive model complexity and variable collinearity. We removed variables that did not contribute significantly to the variance in the PCA, defined as contributing less than 15% to the first component, and we further discarded variables with a VIF>10.

MaxEnt models were fitted through five-fold ($K=5$) cross-validation with 80% training and 20% testing. For each fold, we calculated the area under the receiving operating characteristic curve (AUC), sensitivity, specificity and Cohen's kappa as measures of model performance. To create a single prediction that represents the probability of occurrence for the landrace group, we computed the median across K models. Geographic areas in the form of pixels with probability values above the maximum sum of sensitivity and specificity were treated as the final area of predicted presence.

SDM workflow

Try 5 K-fold

- Background selection
- ENMEval calibration

Landrace spatial distribution

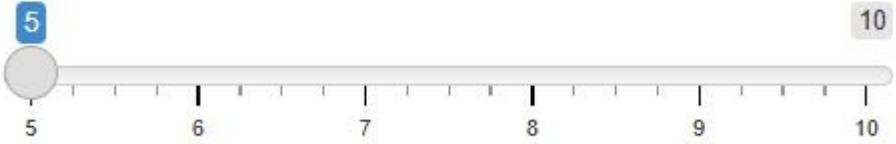
Select Group/Class/Race to process:

triticum_turgidum ▾

Model settings

Number of cross validation folds:

5 6 7 8 9 10



Run model

Wait for 15 minutes



SDM modelling

Geo-Scores

Geo score assessment

Landrace spatial distribution

Select Group/Class/Race to process:

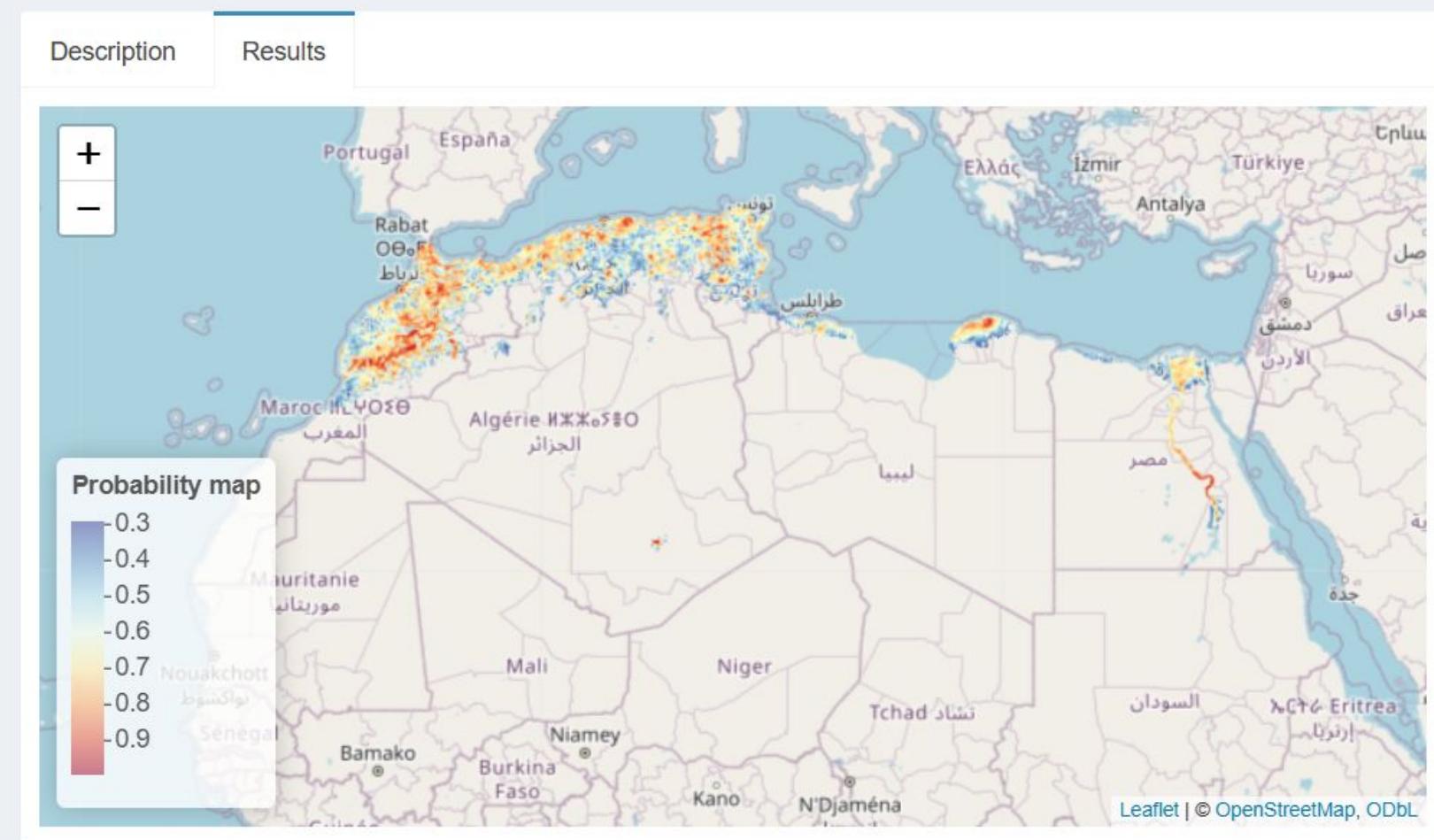
triticum_turgidum

Model settings

Number of cross validation folds:

5

Done





SDM modelling

Geo-Scores

Geo score assessment

Accessibility geo score

Select friction surface file:

friction.tif

Connectivity geo score

Select occurrence shp file:

occurrences.shp

Environmental geo score

Number of environmental cluster:

20

Coordinates to sample:



Calculate

Description

Results

Accessibility geo score

Number of coordinates to use

An accessibility gap score was calculated for each 2.5-arc-minute pixel in the distribution model by computing travel time from each pixel to its nearest genebank accession occurrence location based both on distance and the speed of travel, defined by a friction surface. Travel time scores were normalized by dividing pixel values by the longest travel time within the distribution model, with the final score ranging from 0 to 1. A high accessibility gap score value for a pixel reflects long travel times from existing genebank collection occurrences and, thus, represents a higher probability of the pixel location being a gap in existing ex situ collections.

connectivity gap score

Connectivity geo score

Connectivity gap score is calculated for each 2.5-arc-minute pixel within the distribution model by drawing a triangle around each pixel using the three closest genebank accession occurrence locations as vertices and then deriving normalized values for the pixel based on distance to the triangle centroid and vertices. The connectivity gap score of a pixel is high—closer to 1 on a scale of 0–1—when its corresponding triangle is large, when the pixel is close to the centroid of the triangle or when the distance to the vertices is large. A high connectivity gap score represents a greater probability of the pixel location being a gap in existing ex situ collections.

connectivity gap score

Environmental geo score

Step 8. Obtain gap collection
metrics and maps

Steps

1. Download LGA toolbox (DONE!)
2. Unzip LGA toolbox (DONE!)
3. Put your occurrence data in documents
4. Start your LGA toolbox
5. Create your outcomes folder
6. Preprocessing: Upload occurrences, Select areas, and curate coordinate
7. Modeling species distribution
8. Obtain gap collection metrics and maps
9. Validate approach

Wait for 20 minutes

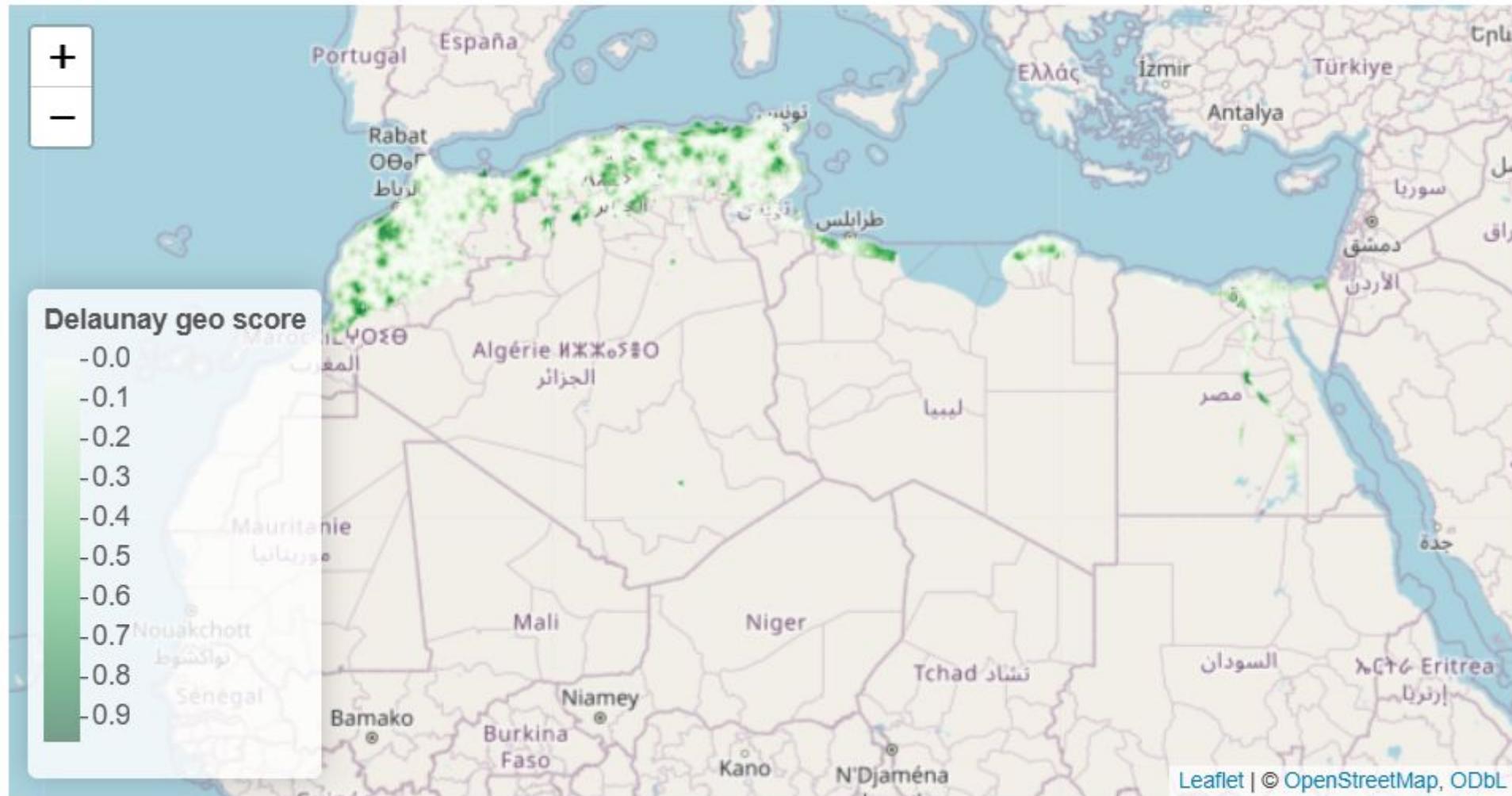
Description

Results

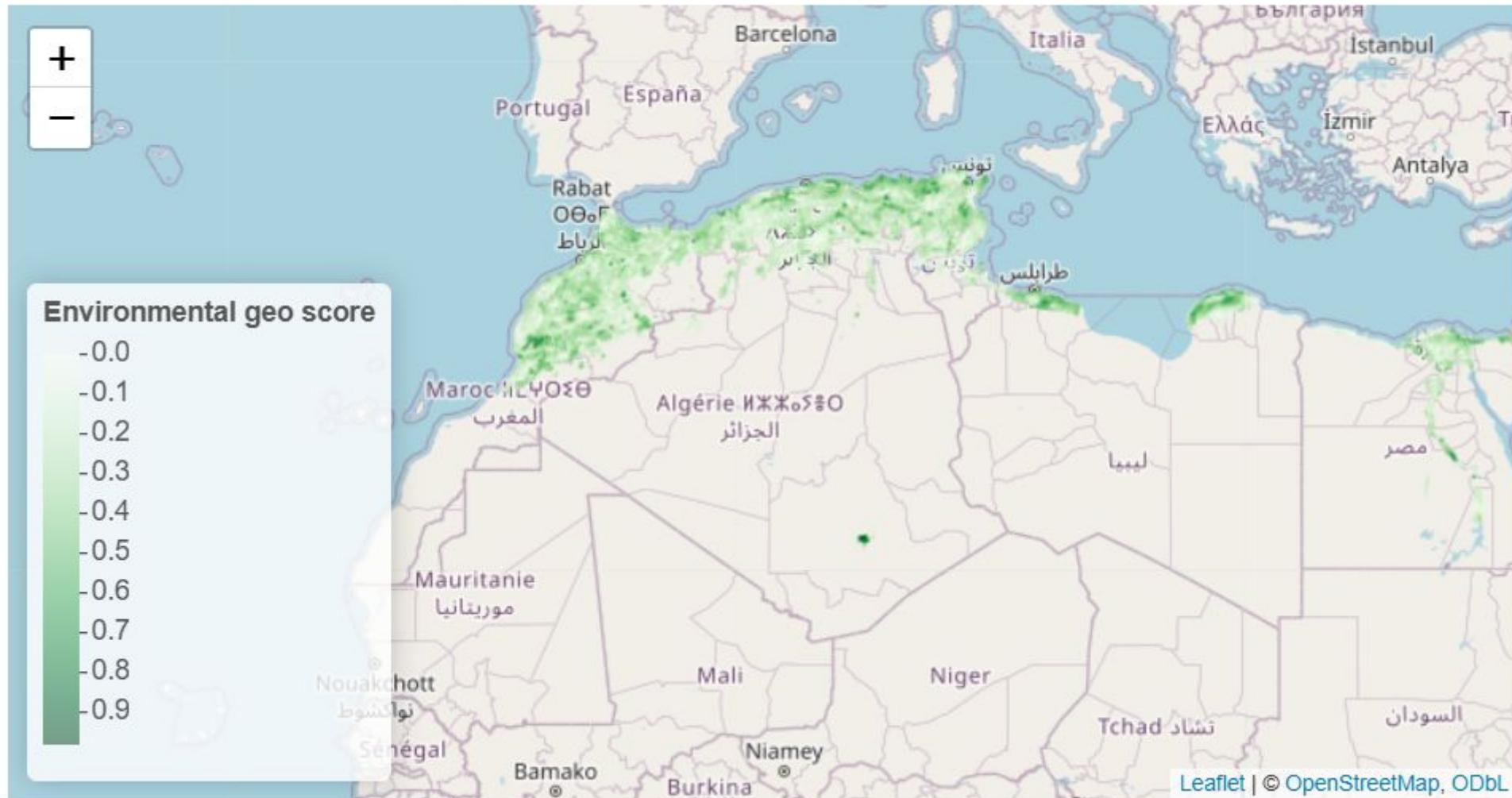
Map: Accessibility geo score



Map: Connectivity geo score



Map: Environmental geo score



Step 9. Validate approach

Number of simulations

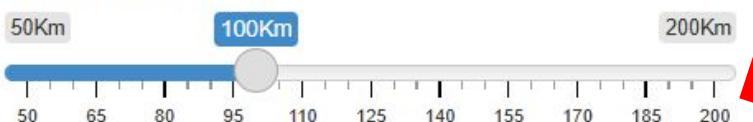
Buffer size for validation

Simulation:

Number of simulations:



Buffer size (in Km):



Progress:

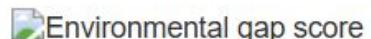
0%

Progress bar

Calculate

Mapping ex situ conservation gaps

Spatial ex situ conservation gaps were calculated from the conservation gap scores using a cross-validation procedure to derive a threshold for each landrace group and each of the gap scores (accessibility gap score, Connectivity gap score, Environmental gap score). To do so, we created synthetic (artificial) gaps by removing genebank occurrences in five randomly chosen circular areas of 100 km radius within the Landrace Distribution Model. We then tested whether these synthetic gaps could be predicted by our method and determined the threshold value of each gap score that would maximize the prediction of these synthetic gaps. Performance for each of the five synthetically created gaps was assessed using the AUC, sensitivity and specificity. Finally, the average threshold value of each gap score, maximizing the prediction of the synthetic gaps (balanced with minimizing false positives), was used to discretize the gap score datasets into areas with a high priority for further collecting (areas with gap score above the threshold, assigned a value of 1) as opposed to relatively well-conserved areas (areas with gap score below the threshold, assigned a value of 0).



Wait for 5 hours

Your validation is ready

SDM modelling Geo-Scores Geo score assessment

Simulation:

Number of simulations: 5 10

Buffer size (in Km): 50Km 100Km 200Km 50 65 80 95 110 125 140 155 170 185 200

Progress:

100%

Description **Results**

[Gap Map](#)

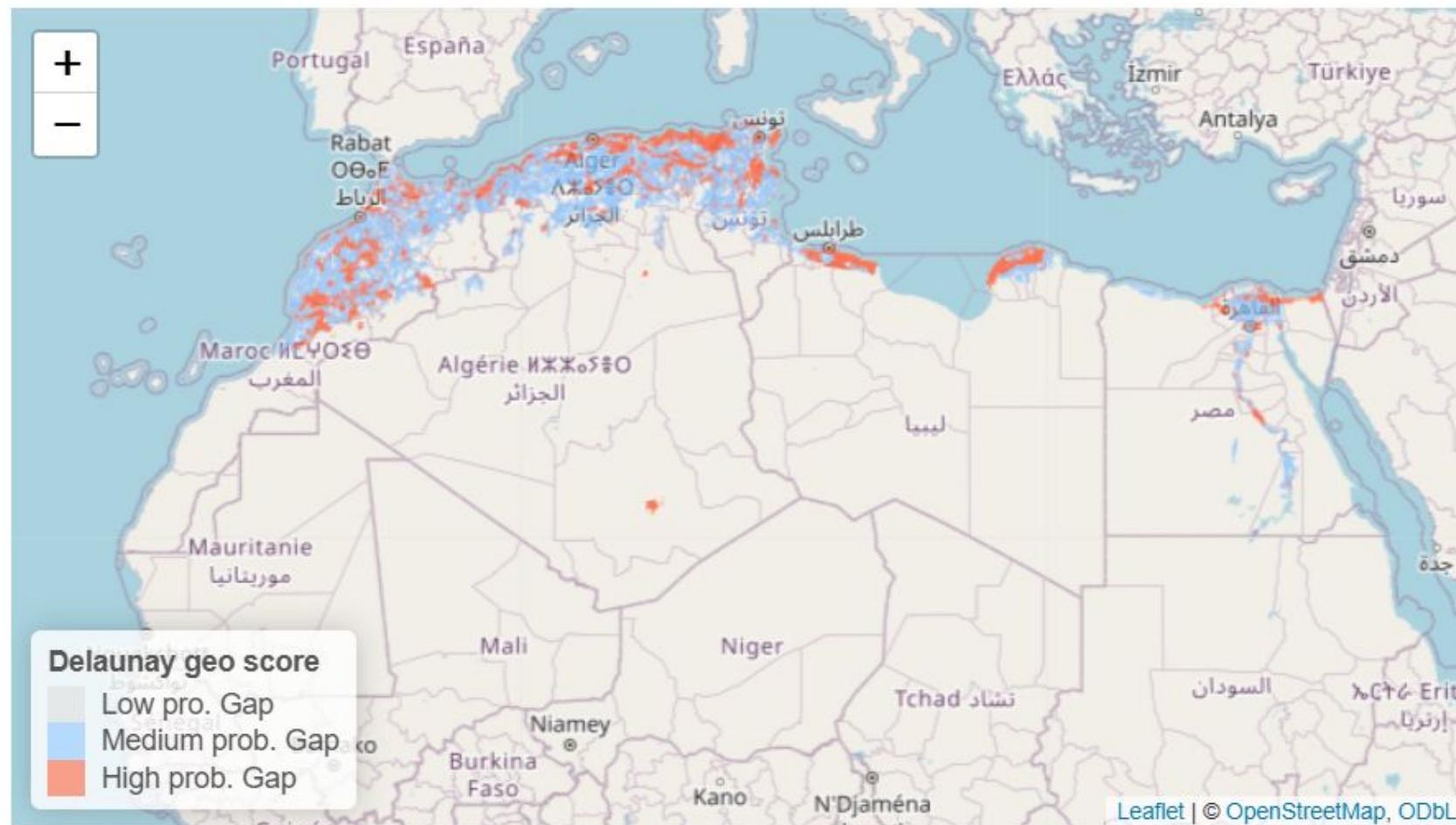
[Summary metrics](#)

[Report](#)

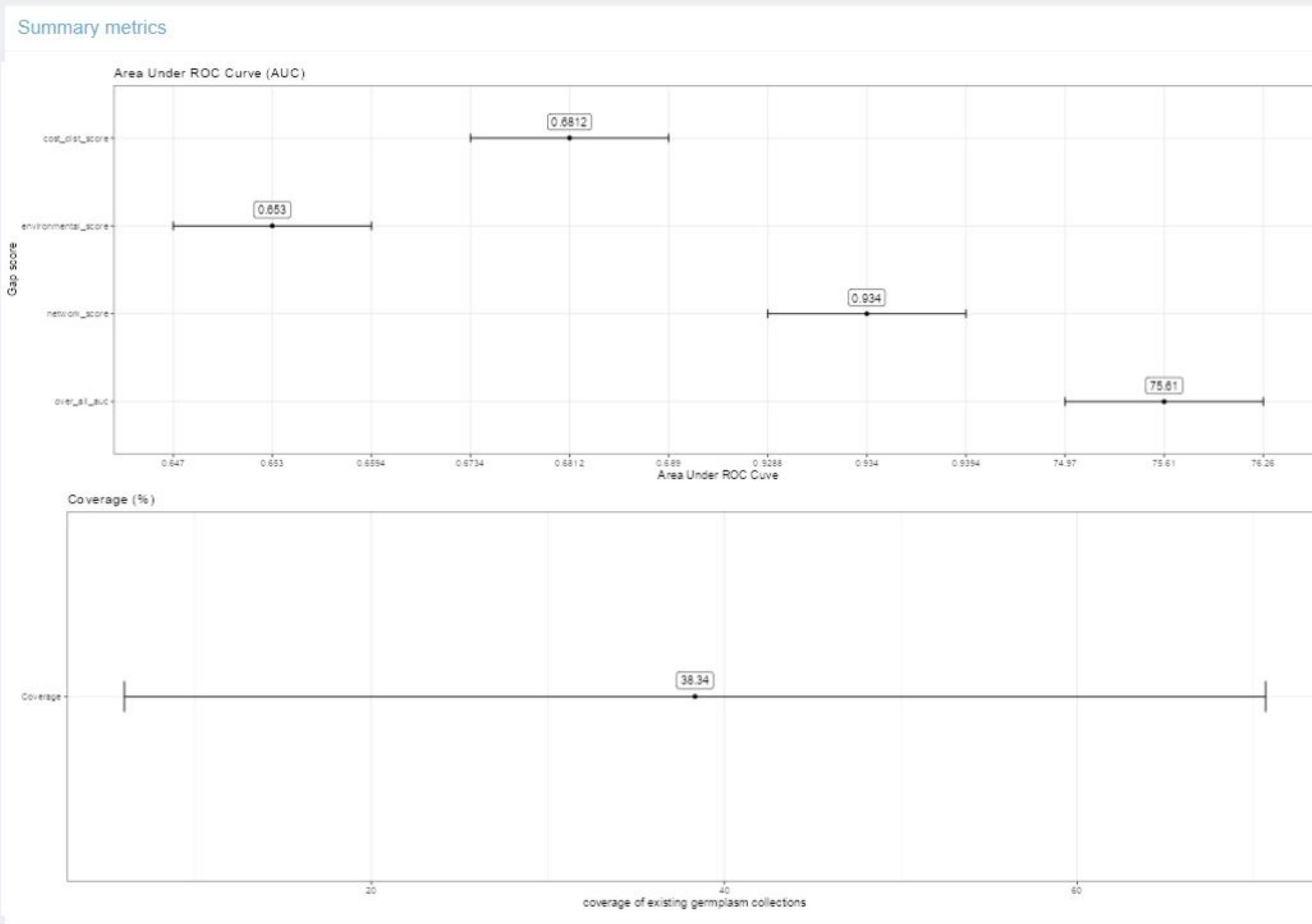
Done

Your collecting áreas are displayed in red

Gap Map



- Environmental score
- Coverage (%) represents how much of your collection is already collected in the species distribution



Report presents the results!

Report

Gap Analysis Landrace Results

11-15-2025

Summary

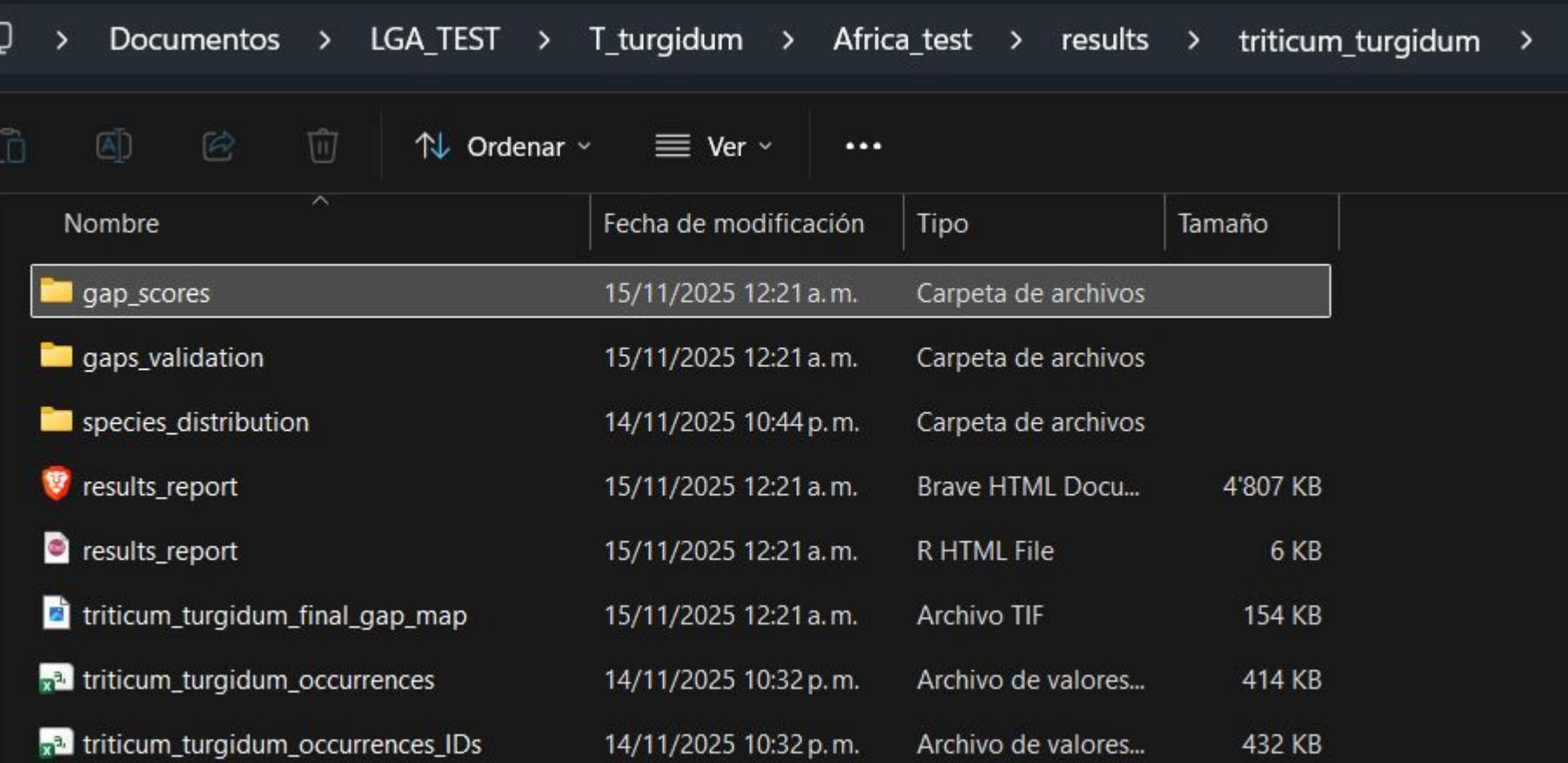
Crop landraces have been defined as "dynamic population(s) of a cultivated plant that has historical origin, distinct identity and lacks formal crop improvement, as well as often being genetically diverse, locally adapted and associated with traditional farming systems". A landrace can be further classified as autochthonous when grown in the original location where it developed its unique genetic and socioeconomic characteristics through grower selection and allochthonous when introduced from another region and then locally adapted (Ramirez-Villegas, et al., 2020).

Results for Landrace Group: *triticum_turgidum*

1. Spatial distribution:

A data cleaning process was applied, during which accessions outside the continental mask and those with duplicated coordinates were removed. The cleaned passport data used for the gap analysis comprises a total of 716 accessions out of a total of 23933, representing the 3% of the records. The next graph shows the modelled spatial distribution for the species. Used accession are shown as darker points.

The results were stored in the folder provided at the beginning



A screenshot of a Windows File Explorer window. The path in the address bar is: Documentos > LGA_TEST > T_turgidum > Africa_test > results > triticum_turgidum >. The window shows a list of files and folders:

Nombre	Fecha de modificación	Tipo	Tamaño
gap_scores	15/11/2025 12:21 a.m.	Carpeta de archivos	
gaps_validation	15/11/2025 12:21 a.m.	Carpeta de archivos	
species_distribution	14/11/2025 10:44 p.m.	Carpeta de archivos	
results_report	15/11/2025 12:21 a.m.	Brave HTML Docu...	4'807 KB
results_report	15/11/2025 12:21 a.m.	R HTML File	6 KB
triticum_turgidum_final_gap_map	15/11/2025 12:21 a.m.	Archivo TIF	154 KB
triticum_turgidum_occurrences	14/11/2025 10:32 p.m.	Archivo de valores...	414 KB
triticum_turgidum_occurrences_IDs	14/11/2025 10:32 p.m.	Archivo de valores...	432 KB