



LGA toolbox

Alliance Bioversity International and CIAT,
November 2024.

The Landrace Gap Analysis method

Gaps identification

Passport data

+ Environment
+ Socio-economy

*Potential
distribution*

Potential areas to be collected

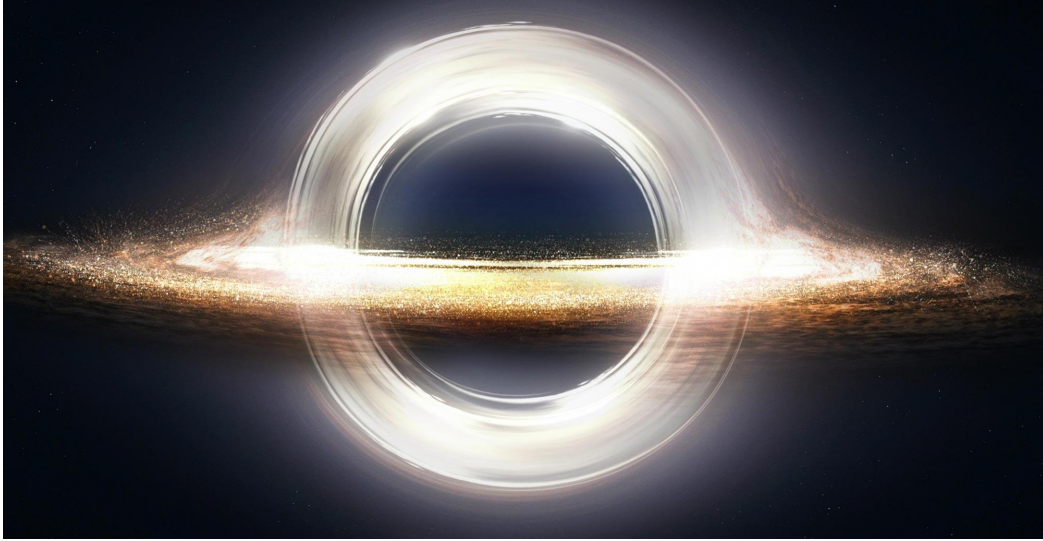
Accessibility
i.e., roads,
railroads, etc.

Environmental
i.e., environmental
uniqueness

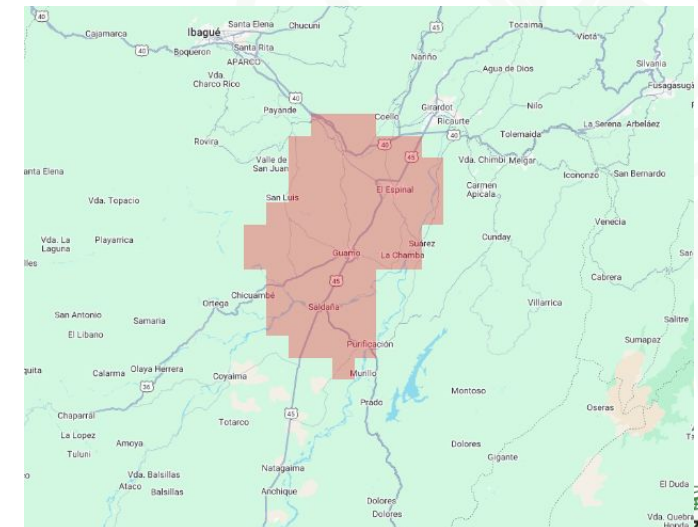
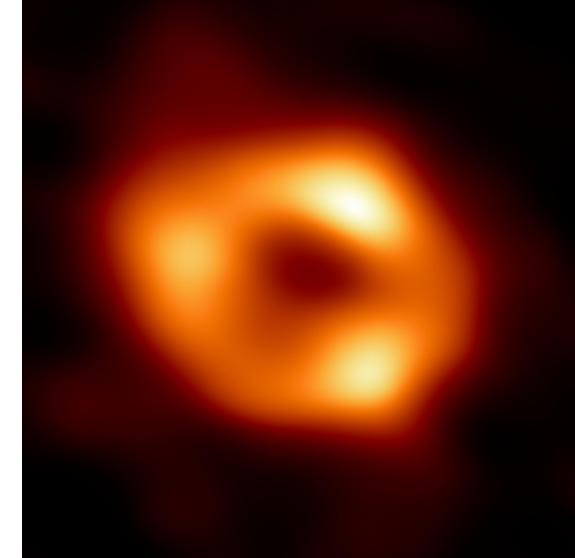
Connectivity
i.e., density of
accessions

What is a model?

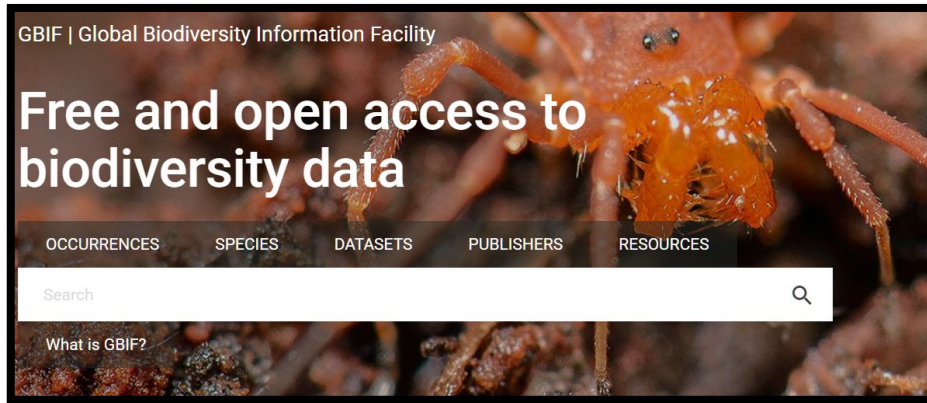
What we really want



What we get



Occurrence data



WIEWS - World Information and Early Warning System on Plant Genetic Resources for Food and Agriculture

Home Background Data Resources Glossary

Ex situ search

Accession-level information of plant genetic resources secured in genebanks (*ex situ*) under medium and long term storage can be retrieved through the search below.

Year
2022

Country
Country

Holding institute
Holding institute

Current selected institute(s)
Selection currently empty

Buttons: x, +, x

Genesys Accession data > Directory > Resources > My List 0 Login

Accessions v Search Genesys...

USDA United States Department of Agriculture
Agricultural Research Service

GRIN-Global

USDA National Plant Germplasm System

Version: 2.3.11.2

Accessions Descriptors Reports GRIN Taxonomy GRIN Help Contact Us Your Profile

Cart

Welcome!

Effective Jan. 1, 2025, the U.S. National Plant Germplasm System will modify its SMTA acceptance policy to distinguish between propagative and non-propagative germplasm. [Click here for more information.](#)

Log in New User

The Datatable

CROPNAME	DECLATITUDE	DECLONGITUDE	status	source_db	database_id	issue_txt_desc
brassica_rapa	25.012526	102.700013	H	GBIF	gbif_1	Difference in ELEVATION and STRM (elevation) greather than 150 mts or missing; Missing Country admon level 1 in COLLSITE; Missing Country admon level 2 in COLLSITE
brassica_rapa	25.275367	103.003972	H	GBIF	gbif_2	Difference in ELEVATION and STRM (elevation) greather than 150 mts or missing; Missing Country admon level 1 in COLLSITE; Missing Country admon level 2 in COLLSITE
brassica_rapa	25.275367	103.003972	H	GBIF	gbif_3	Difference in ELEVATION and STRM (elevation) greather than 150 mts or missing; Missing Country admon level 1 in COLLSITE; Missing Country admon level 2 in COLLSITE
brassica_rapa	25.012526	102.700013	H	GBIF	gbif_4	Difference in ELEVATION and STRM (elevation) greather than 150 mts or missing; Missing Country admon level 1 in COLLSITE; Missing Country admon level 2 in COLLSITE
brassica_rapa	24.385833	102.806111	H	GBIF	gbif_8	Missing Country admon level 2 in COLLSITE

G: Germplasm accession

H: Herbarium, Human observation, botanical garden, other sources of occurrences.

Spatial predictors

ENVIREM

ENVIRONMENTAL RASTERS FOR ECOLOGICAL MODELING

WorldClim

Maps, graphs, tables, and data of the global climate



LGA toolbox

The screenshot displays the LGA Toolbox application interface. On the left is a dark sidebar with three menu items: 'Wizards' (marked with a red circle and the number 1), 'Gap scores' (marked with a red circle and the number 2), and 'Coord quality'. The top of the application has a green header bar with the text 'LGA Toolbox', a hamburger menu icon, and user/notifications icons. Below the header is a light gray navigation bar with tabs: a settings gear icon, 'Working directory' (the active tab), 'Define study area', 'Covariates', and 'Passport data'. The main content area is divided into two panels. The left panel, titled 'Working directory set up', contains a 'select folder' section with a text input field showing 'No folder path selected', a 'Browse' button, and a blue 'Select' button. The right panel, titled 'Working directory set up', contains a descriptive paragraph and a numbered list of three steps for setting up the working directory.

LGA Toolbox

Wizards

Gap scores

Coord quality

Working directory

Define study area

Covariates

Passport data

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...

Study area

Working directory

Define study area

Covariates

Passport data

Geographic area

Create

Import

Select

Select one region:

World

Write file name

Create

Geographic area selector wizard

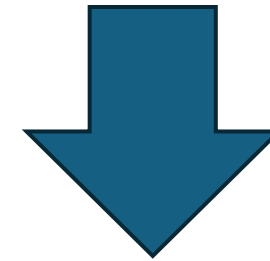
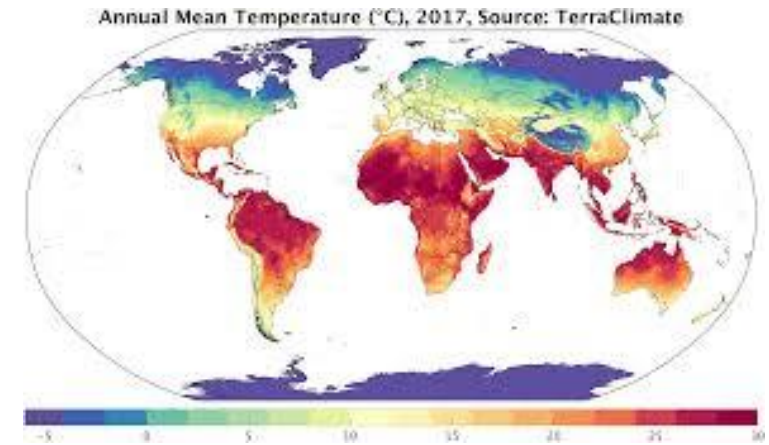
the wizard will support you to define the strategic areas/regions to analyze.

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.

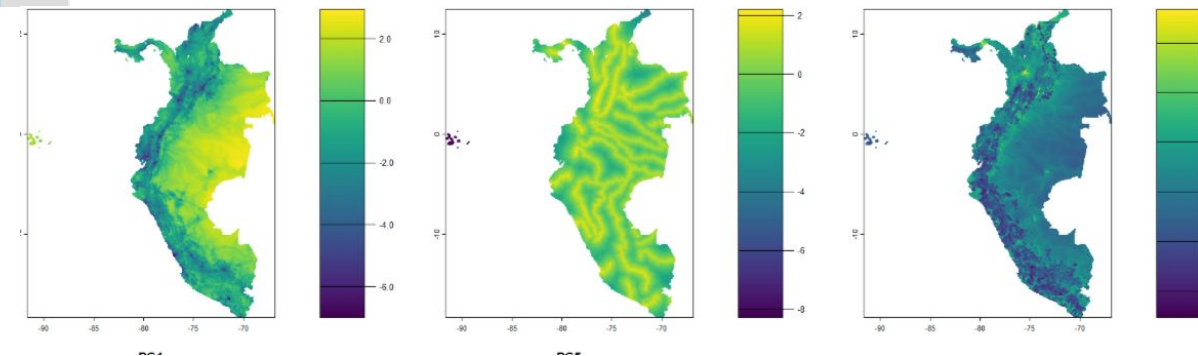
file output: input_data/“mask_name”.tif

+

–



Which GIS process are we applying ?



Occurrence data setting up

Working directory

Define study area

Covariates

Passport data

Data base set up

1. Select .csv database

brassica_rapa_to_process.csv

Upload complete

2. Write response column number

1

☐ Predict missing groups

Set up database

DescriptionParametersPreview dataResults

Show 5 entries

	crop_name	Latitude	Longitude	status	source_db	databa
1	brassica_rapa	35.35	75.63333333	G	PAK253	921
2	brassica_rapa	35.46666667	75.46666667	G	PAK253	922
3	brassica_rapa	35.35	75.5	G	PAK253	923
4	brassica_rapa	35.35	75.5	G	PAK253	924
5	brassica_rapa	36.36666667	74.83333333	G	PAK253	925

Showing 1 to 5 of 20 entries

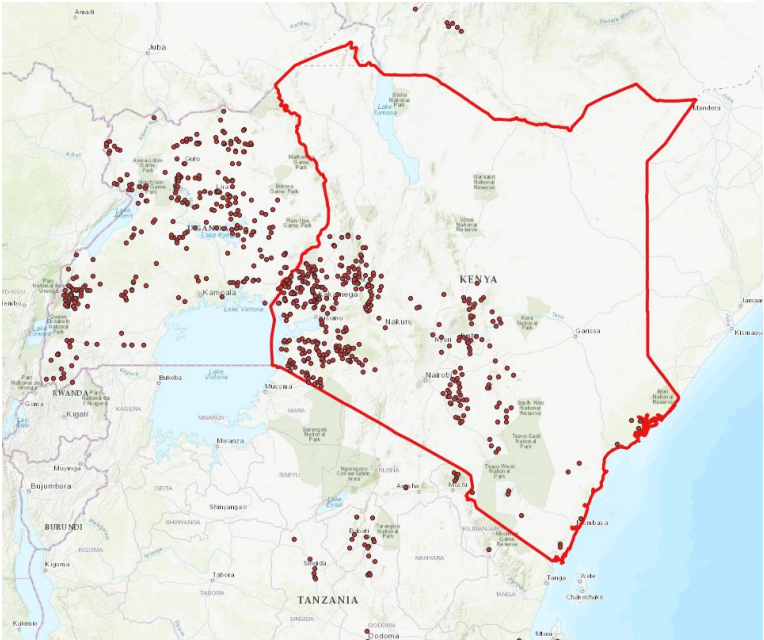
Previous1234Next

787Rows

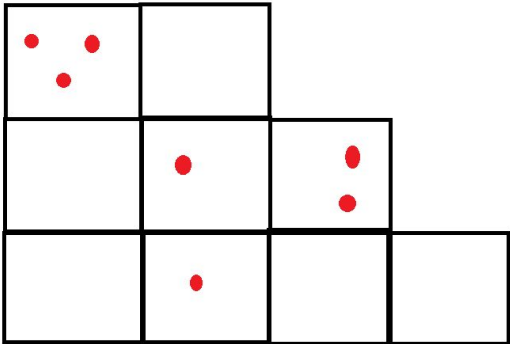
1Groups

0%Missing Values

Remove coordinates outside study region



Remove duplicated coordinates by pixel



Gap scores

Wizards

Gap scores

Coord quality

SDM modelling

Geo-Scores

Geo score assessment

Landrace spatial distribution

Select Group/Class/Race to process:

none

Model settings

Number of cross validation folds:

510

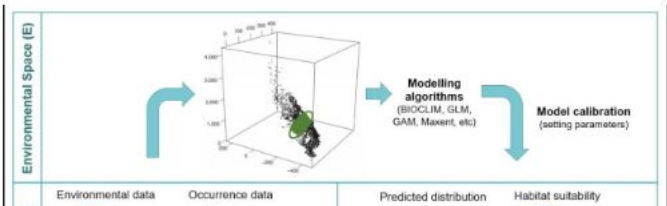
Run model

DescriptionResults

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.



Geo score assessment

SDM modelling Geo-Scores Geo score assessment

Simulation:

Number of simulations:

5 10

5 6 7 8 9 10

Buffer size (in Km):

50Km 100Km 200Km

50 65 80 95 110 125 140 155 170 185 200

Progress:

0%

Calculate

Description Results

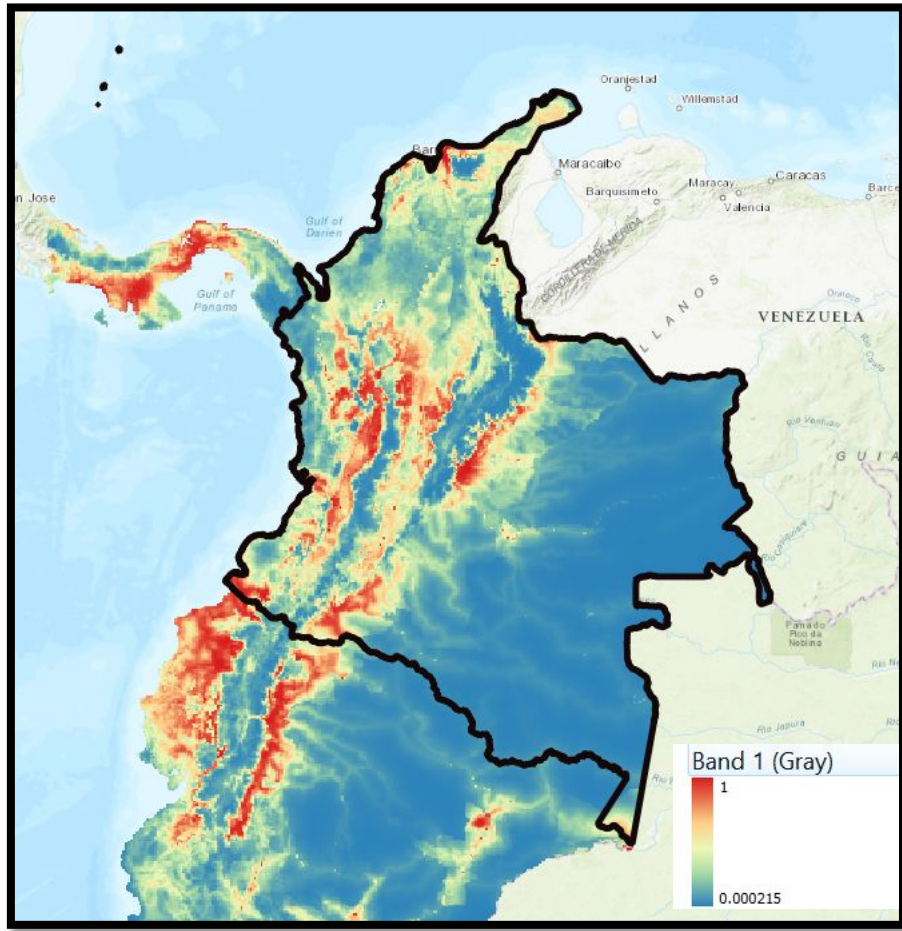
Mapping ex situ conservation gaps

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).

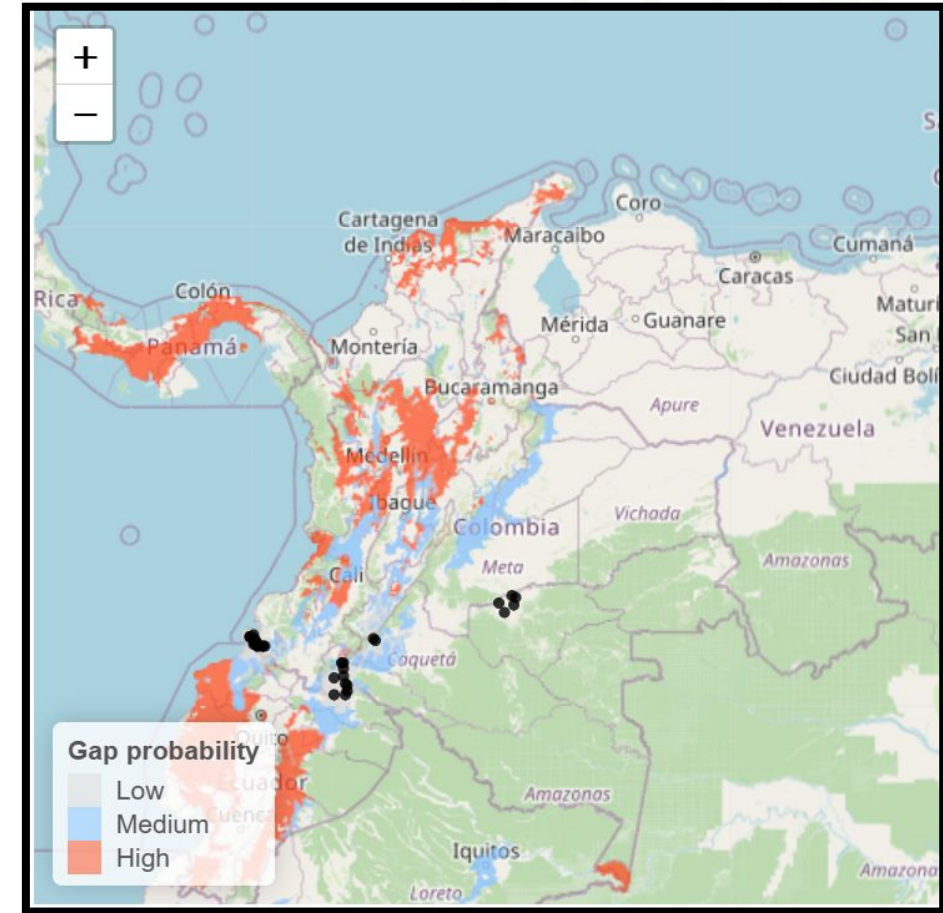
Aspects to consider:

1. Low number of G accession may make it impossible to calculate any of the gap scores
2. Low number of G accessions make it impossible to calculate any of the gap scores

How to prioritize areas for collecting ?



Geo score
assessment



Gap Analysis Landrace Results

11-07-2025

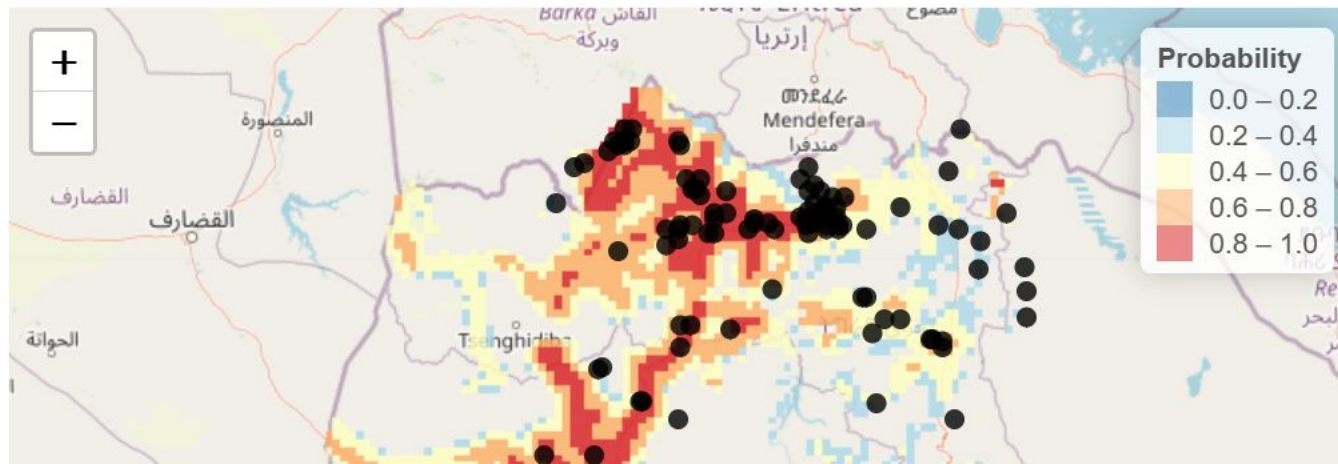
Summary

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).

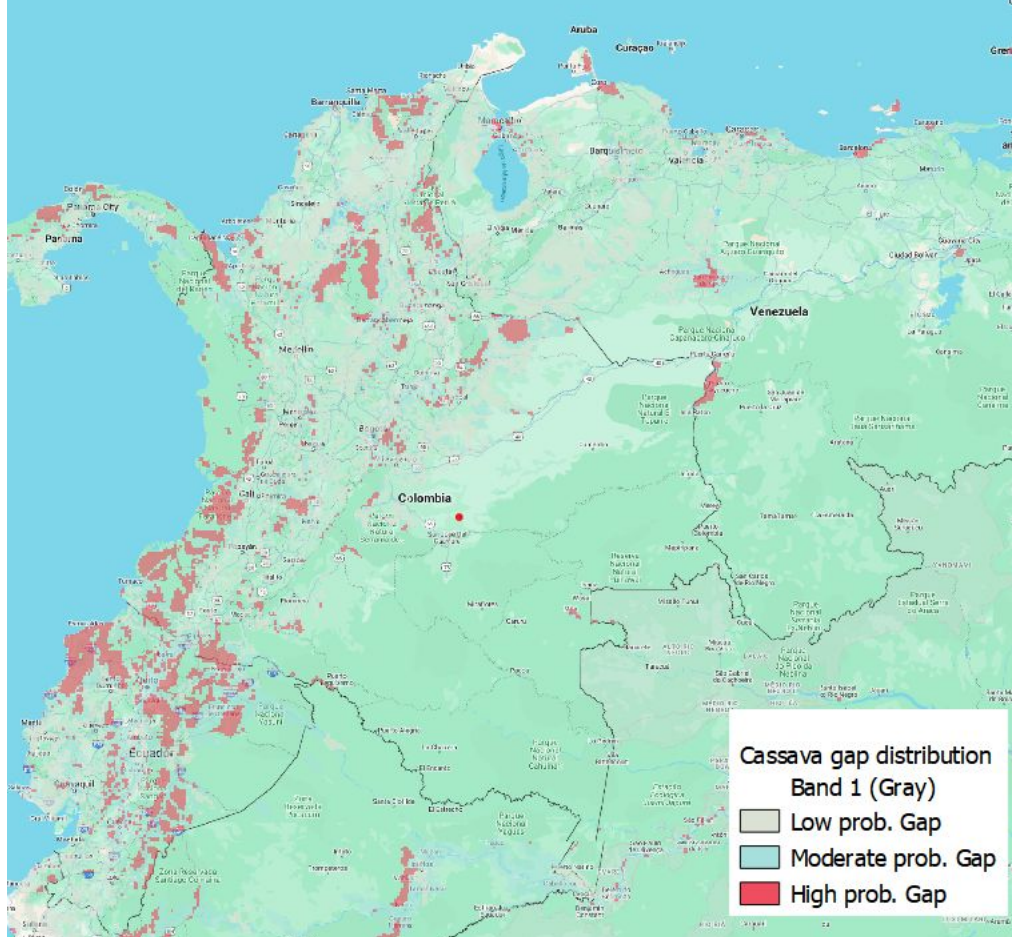
Results for Landrace Group: *eleusine_coracana*

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 **1060** accessions out of a total of **13760**, representing the **8%** of the records. The next graph shows the modelled spatial distribution for the specie. Used accession are shown as darker points.



What insights could we obtain from results?



Metrics	Value
Coverage	68%
Total Gap Area	23.000 Km ²
Average Elevation	635 Mts

- **Coverage:** The 68% percent of the specie distribution it is been covered by the germplasm collection (this mean that the 32% remaining it is missing. It's good or not ?).
- The total area highlighted as a gap are **23.000 Km²** (this is good bearing in mind that the whole country has more than 1.142 million km²).
- In average, the gap areas are in an elevation of **635 Mts** above sea level (is this elevation suitable for the crop to grow?)

What additional insights can we gain from this?

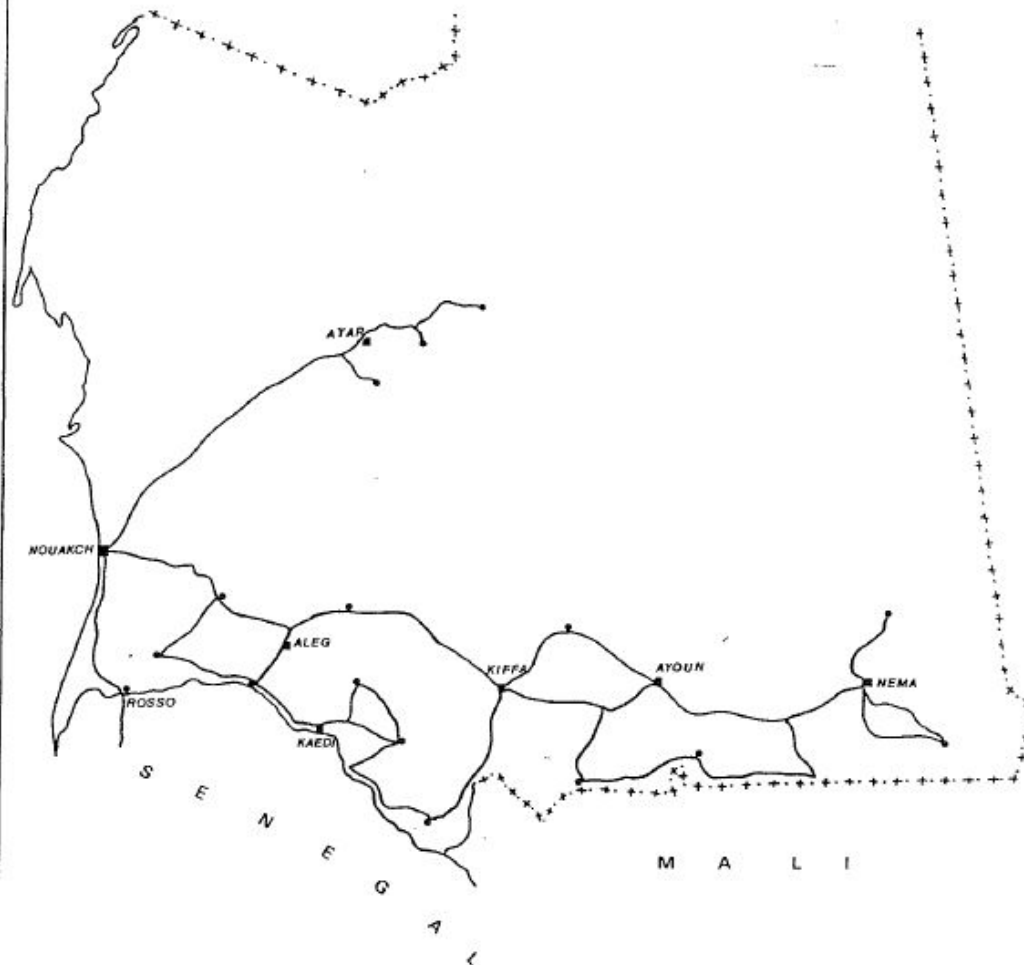
Considerations

- Final gap maps are not filtered by protected areas.
- Final gap maps are not filtered by Land use coverage.
- Spatial model did not use soil properties variables (possible enhancement).
- **Ensure an expert inspection over the results to refine even more the gap maps.**

MAURITANIE

itineraires de prospection

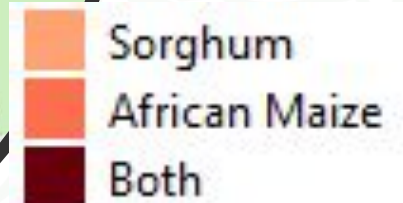
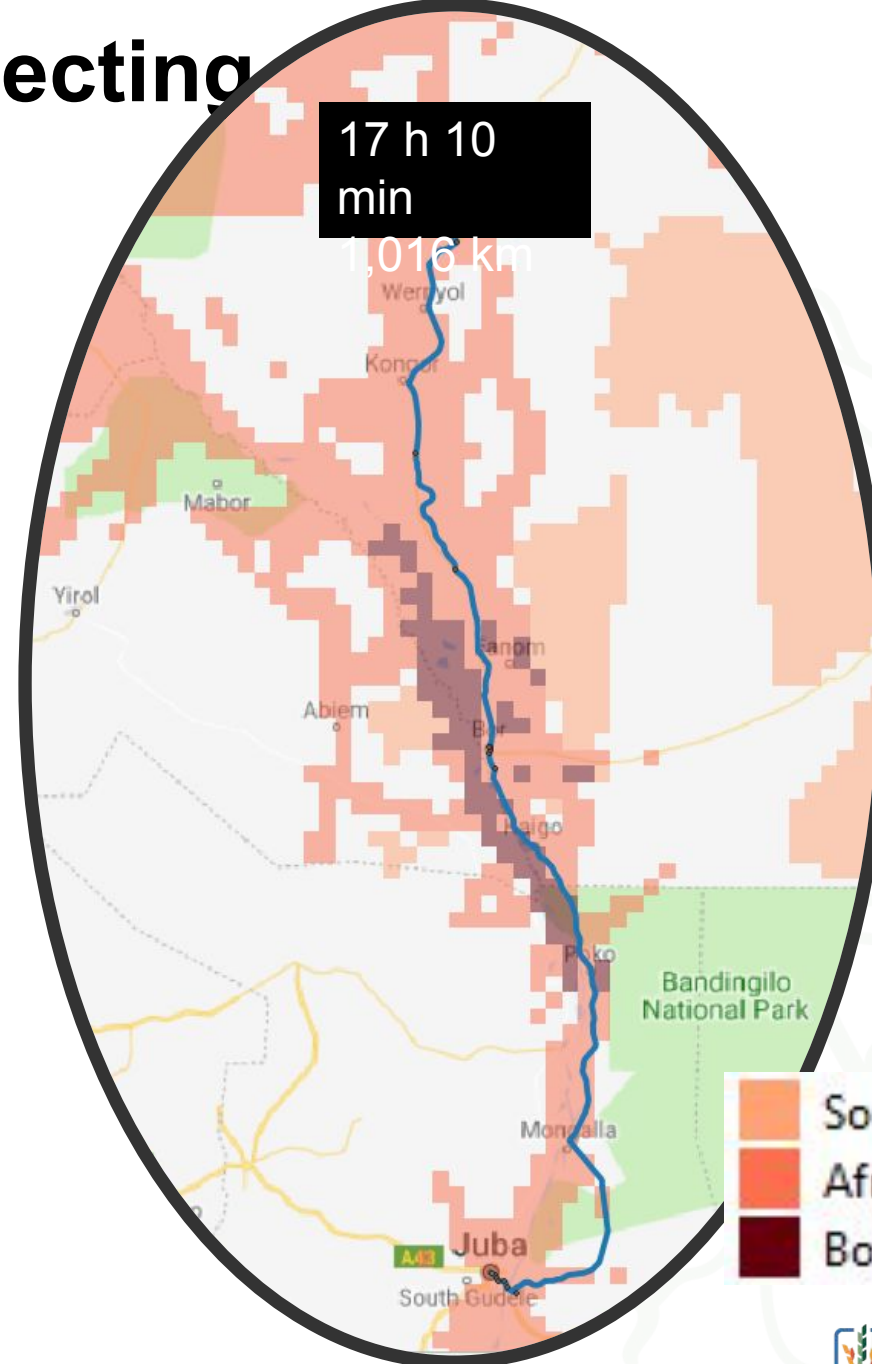
oct-nov 1984





Google Maps

Gap analysis can support collecting



Takeaways

- Something is better than nothing
- Our approach can give hints of current landrace distribution
- Comprehensive and reliable methodology
- Completely availability of gaps maps for crops
- Useful for planning future collecting mission



Thank you!