Horizon 2020 Programme

FOCUS-Africa

Research and Innovation Action (RIA)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869575

Start date : 2020-09-01 Duration : 52 Months

FOCUS-AFRICA

A data catalogue providing the complete genotypic, climatic,and phenotypic characterization of the rice and cowpe accessions used in CS3

Authors : Mr. Matteo DELLACQUA (SSSA), Matteo Dell'Acqua, Mercy Macharia, Valentina D'Amico, Marta Solemanegy, Robel Takele Miteku

FOCUS-Africa - Contract Number: 869575

Project officer: Javier BARRIO ALONSO

Document title	A data catalogue providing the complete genotypic, climatic,and phenotypic characterization of the rice and cowpea accessions used in CS3
Author(s)	Mr. Matteo DELLACQUA, Matteo Dell'Acqua, Mercy Macharia, Valentina D'Amico, Marta Solemanegy, Robel Takele Miteku
Number of pages	30
Document type	Deliverable
Work Package	WP5
Document number	D5.5
Issued by	SSSA
Date of completion	2024-12-04 14:48:17
Dissemination level	Public

Summary

D5.5 document

Approval

Date	By
2024-12-05 01:08:24	Dr. Alberto TROCCOLI (WEMC)
2024-12-05 03:46:46	Mrs. Roberta BOSCOLO (WMO)

FOCUS-AFRICA

Full-value chain Optimised Climate User-centric Services for Southern Africa

A data catalogue providing the complete genotypic, climatic, and phenotypic characterization of the rice and cowpea accessions used in CS3

Deliverable D5.5

Lead Beneficiary: SSSA

November/2024

[1]Macharia Mercy, [1, 2]Solemanegy Marta, [1]Caproni Leonardo, [1]Takele Robel, [2]Munisse Paulino, [2]Amane Manuel ,[1] Buizza Roberto, [1]D'Amico Valentina, [3]Mutolo Raquelina, [3]Carlevaro Beatrice, [1]Pè Mario Enrico, [1]Dell'Acqua Matteo

1 Institute of Plant Sciences, Scuola Superiore Sant'Anna Pisa, Italy
2 Instituto de Investigação Agraria de Moçambique, Mozambique
3 PLAN International, Mozambique

www.focus-africaproject.eu

This project has received funding from the European Commission's Horizon 2020 Research and Innovation Programme. The content in this presentation reflects only the author(s)'s views. The European Commission is not responsible for any use that may be made of the information it contains.

Document Information

Grant Agreement:	869575
Project Title:	Full-value chain Optimised Climate User-centric Services for Southern Africa
Project Acronym:	Focus-Africa
Project Start Date:	1 September 2020
Related work package:	WP 5: Develop end-user tailored climate services prototypes
Related task(s):	Task 5.3: Development and delivery of prototype climate services
Lead Organisation:	Scuola Superiore Sant'Anna, Pisa, Italy
Submission date:	30 November 2024
Dissemination Level:	1

History

Date	Submitted by	Reviewed by	Version (Notes)
08/11/2024	Matteo Dell'Acqua; Mercy Macharia	Alberto Troccoli	Returned for edits 19/11
23/11/2024	Matteo Dell'Acqua; Mercy Macharia	Alberto Troccoli	Returned for edits 27/11
29/11/2024	Matteo Dell'Acqua; Mercy Macharia	Alberto Troccoli	Final version

About FOCUS-Africa

FOCUS-Africa – Full-value chain Optimised Climate User-centric Services for Southern Africa – is developing sustainable tailored climate services in the Southern African Development Community (SADC) region for four sectors: agriculture and food security, water, energy and infrastructure.

It will pilot eight case studies in six countries involving a wide range of end-uses to illustrate how the application of new climate forecasts, projections, resources from Copernicus, GFCS and other relevant products can maximise socio-economic benefits in the Southern Africa region and potentially in the whole of Africa.

Led by WMO, it gathers 14 partners across Africa and Europe jointly committed to addressing the recurring sustainability and exploitation challenge of climate services in Africa over a period of 48 months.

For more information visit: www.focus-africaproject.eu

Coordinator Contact
Roberta Boscolo | Climate & Energy Scientific Officer
Applied Climate Services Division
Services Department
World Meteorological Organization (WMO)
CP 2300, 1211 Geneva SWITZERLAND
email: rboscolo@wmo.int

Table of Content

Tables

Figures

Executive Summary

Case study 3 (CS3) of FOCUS Africa was tasked with supporting the development and tailoring of climate services for food security in Mozambique, a country whose economy is largely based on subsistence farming. Agriculture is a vital sector in Mozambique, employing around 70% of the population and contributing to rural livelihoods. Most farming is small-scale and rainfed, with a focus on staple crops including rice and cowpea. The sector faces challenges, including limited access to modern farming inputs, infrastructure, and market access. Climate variability, especially drought and floods, further impacts agricultural productivity. In this framework, the strategy of CS3 has been the development of two interconnected climate services for smallholder farming communities: (i) identifying and characterizing the most promising rice and cowpea varieties for local cultivation with a focus on traits that enhance resilience to climate variability, and (ii) developing a seasonal forecast, along with a forecasting tool, to predict the onset of the rainy season, guiding crop sowing schedules.

Co-producing climate services is an “iterative and interactive” process involving collaboration between climate service providers and users, with success measured by how useful and usable the resulting information product is for end-users. Within the Mozambican case study, the team used data-driven participatory approaches to develop and test the draft trial climate services, aiming to ensure that the final product remains relevant and sustainable beyond the project's lifespan. A key component of our approach included a socioeconomic focus, utilizing surveys, participatory variety selection activities, and focus group discussions. The climate services produced by CS3 have been previously reported.

In D2.4 we provided an in-depth analysis of the state of the art in cowpea and rice plant genetic resources, along with the strategy, methods, and preliminary analyses of CS3 results in the identification of most promising crop genotypes. We outlined our efforts to assemble a comprehensive portfolio of genetic resources, sourced from both local and global gene banks, and characterized them in terms of agronomic, climatic, and genetic diversity. Our preliminary findings from the previous report uncovered existing genetic diversity in the rice and cowpea core collections, identified genomic loci with potential for local climate adaptation, and demonstrated the ability of existing germplasm to adapt under projected climate scenarios.

In D5.2 we described in detail the forecasting tool that we have developed to predict the onset of the rainy season. We described the AquaBEHER R package that we developed and made freely available, and how it integrates daily evapotranspiration data into a soil-water balance model, enabling estimates and visualizations of the wet season calendar based on climate data from weather stations or spatial datasets. We further developed a graphical user interface (GUI) named AquaBEHERgui to allow broader usage beyond command-line expertise. A prototype seasonal forecast was developed using ECMWF data and demonstrated to stakeholders using AquaBEEHERgui, including met services and agricultural experts. The forecast used a tercile-based system, categorizing early, normal, or late onset probabilities.

In this deliverable we focused on the harmonization and release of a comprehensive data catalogue that represents all the main data generated in the project by CS3. This data catalogue includes multidisciplinary data spanning from climate information, to genomics, to socioeconomic enquiries, reflecting the complexity of the work carried out in CS3. The catalogue is online at this doi: 10.6084/m9.figshare.27628041 and is set to become fully accessible upon publication of the scientific results. While most of the data is in digital format, some of the results achieved by CS3 are more tangible in Mozambique, reflecting practical outcomes that are observable and directly impactful in the local

context; the catalogue also features a set of cowpea and rice seeds prioritized for climate adaptation based on this transdisciplinary characterization. The data collected can be analysed either as individual datasets, which refer to distinct data types such as genomic, phenotypic, or climatic information (e.g., to characterize the genomic diversity of crops) or as a combination of datasets from different domains which involves integrating data from various fields such as genetics, climate science, and agronomy to understand the genetics of adaptation. This deliverable is divided in two sections: the first part of the document will focus on the individual datasets composing the data catalogue, providing a brief description of their content and selected highlights about the information they contain. The second part of the document will focus on use cases of the data developed by CS3, showcasing its potential application with a focus on adaptation strategies that may enhance resilience food security.

Keywords

Data catalogue, food security, agroclimate, genomics, participatory research, smallholder farming systems, rice, cowpea, Mozambique

1 Data Catalogue

Here we provide a description of the datasets generated by CS3. The list is not comprehensive and focuses on the most important and large-scale data generated in the project. The datasets are presented independently, without implying a specific chronology or interconnectedness. All datasets can be accessed at this doi: [10.6084/m9.figshare.27628041].

1.1 Assembly of cowpea and rice germplasm collections

In D2.4, we provided a detailed description of the cowpea and rice core collections used in CS3. This included reporting about the multiplication and purification of seeds acquired from national and international gene banks. Briefly, cowpea (*Vigna unguiculata*) accessions were sourced from the Agricultural Research Institute of Mozambique (IIAM), Eduardo Mondlane University in Mozambique, and the International Institute of Tropical Agriculture (IITA) in Nigeria. Rice (*Oryza sativa*) accessions were sourced from the IIAM gene bank and the International Rice Research Institute (IRRI) in the Philippines. The collections, primarily composed of landraces, were restricted to accessions with a historical origin in Mozambique and neighboring southern African countries, including South Africa, Zimbabwe, Malawi, and Tanzania. The final collection is composed of 417 cowpea accessions and 324 rice accessions**.** The CS3 team curated the passport data for all the accessions, including essential information about each sample, such as original sampling coordinates, facilitating the documentation and tracking of these genetic resources **(Dataset 1 & Dataset 2)**. The curated information includes an IIAM-generated gene bank accession number, collection location, geographic coordinates (where available), and type of material (improved or landrace). All accessions are stored at the local IIAM gene bank in Mozambique, are linked with characterization data (see below), and can be requested by researchers and breeders for further study and use.

1.2 Genomic characterization of cowpea and rice collections

Detailed descriptions of the DNA extraction, genotyping strategies, data processing methods and results for both cowpea and rice collections can be found in D2.4. Briefly, DNA extraction was followed by genotyping, a process looking at an organism's DNA to identify specific genetic differences or variations. For cowpea, genotyping resulted in 3,245,002,228 data points, which is sufficient for high-resolution genetic analyses. The raw DNA sequence produced is available in a public, freely accessible database at the National Center for Biotechnology Information (NCBI) under the project code **PRJNA951214**. For analysis, the raw reads were aligned to the cowpea reference genome, to identify Single Nucleotide Polymorphisms (SNPs). SNPs are small DNA variations within an organism’s genome. The alignment yielded 202,946 genome-wide (SNPs) (**Dataset 3**), which were further filtered to an operational, high-quality set of 18,969 SNPs (**Dataset 4**). For climate analyses, the SNP set was further pruned to 2,242 markers to minimize redundancy and statistical errors (**Dataset 5**). The target number of SNPs varies by study objectives, with larger sets suited for genetic diversity analysis and smaller sets preferred for climate-related studies to ensure efficiency and reliability. Climate SNPs are typically fewer because they

are selected for their relevance to climate-related traits or environmental adaptation, rather than broad genetic diversity. While we did not have a fixed target, the selected numbers were deemed sufficient based on several parameters including the genetic variation captured, SNP quality, and computational efficiency. The genetic analysis of the cowpea collection revealed three main groups, with one distinct group made up mostly of Mozambican varieties (**Figure 1a**). The rest of the accessions, from various southern African countries, did not form clear geographical groups. The Mozambican cluster showed low genetic diversity, meaning the plants were genetically similar, while the other two groups, consisting of accessions from various countries and improved varieties, had more variation. This could indicate that Mozambican plants have had less exposure to outside genetic material, possibly due to local conditions or limited seed exchange. It may also suggest that the current ex situ collection doesn't fully represent the genetic diversity found in natural Mozambican cowpea populations. This underscores the need for a revised sampling strategy to better capture the full range of genetic diversity in Mozambique. For the rice core collection, genotyping resulted in 331 million paired-end reads, which is considered a large volume of data based on its high sequencing depth, enabling accurate representation of genetic variation. This substantial amount of data allows for high-resolution genotyping within the rice core collection supporting robust downstream analyses for genetic and climate adaptation studies. These reads are available on the NCBI database under project number **PRJNA952322**. Aligning the raw reads to the *Oryza sativa* reference genome, R498v3, resulted in more than 1 million SNPs which were further filtered resulting in 18,1330 high quality SNP markers to be utilized for downstream analysis (**Dataset 6**) and 1,296 for climate analysis (**Dataset 7**). In rice (*Oryzae sativa*), generating over 1 million SNPs is not unusual, especially with high-density genotyping methods like the double digest restriction-site associated DNA (ddRAD) method used in this study. The initial large dataset is then refined to focus on high-quality, informative SNPs that are relevant for specific analyses. As with the cowpea analysis, the 18,130 SNPs for downstream analysis and 1,296 SNPs for climate analysis in rice were deemed sufficient based on the captured genetic variation, the quality of the SNPs, and their relevance to the specific analysis objectives. Improved rice varieties generally formed one group. Most accessions from Zimbabwe were also grouped together, as were those from Tanzania, suggesting some level of geographical structuring. In contrast, landrace accessions from Mozambique were spread evenly across the different clusters (**Figure 1b**), indicating greater genetic diversity within the Mozambican population.

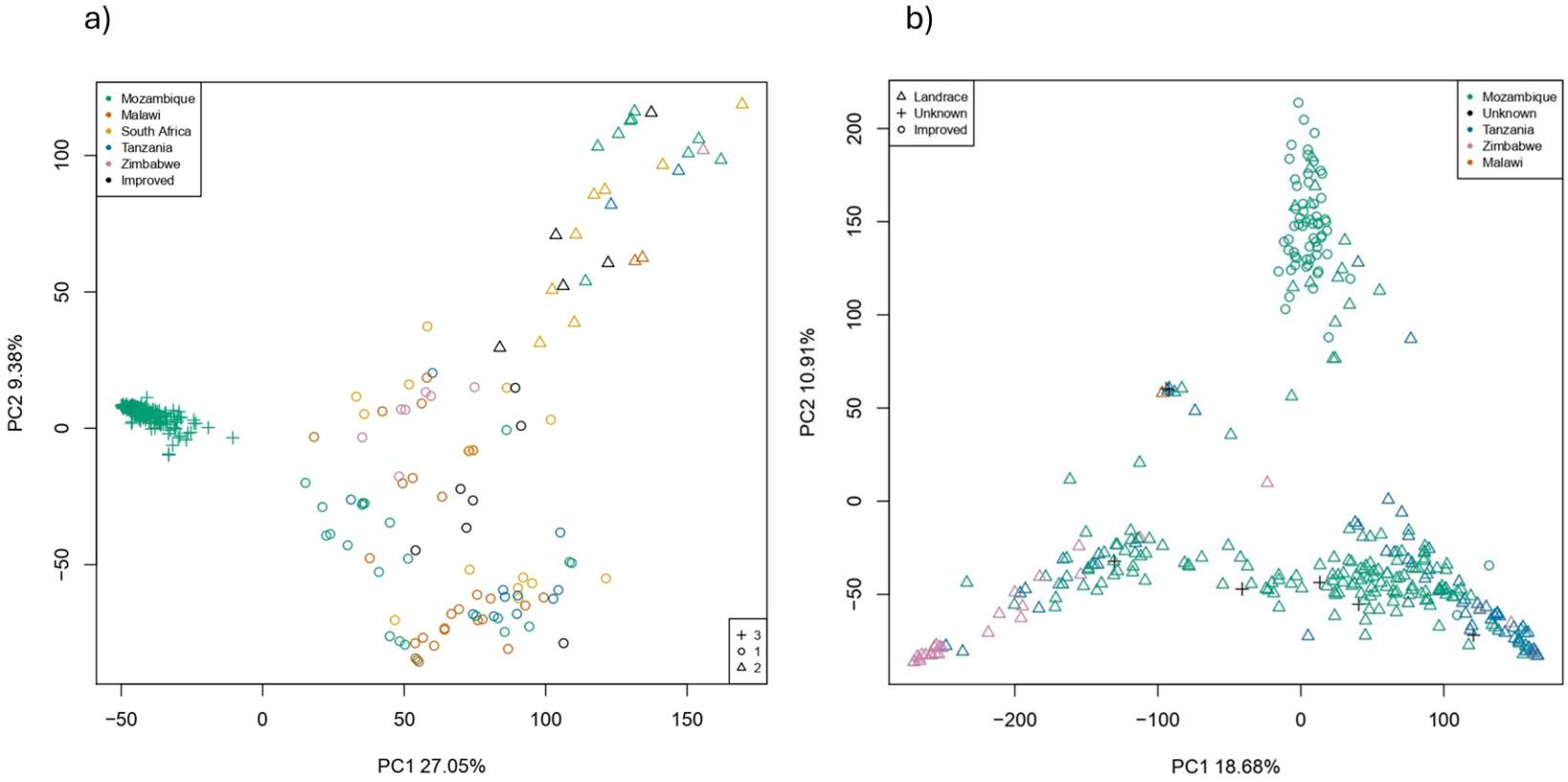

Figure 1. The population structure of 389 cowpea landrace accessions and 28 improved varieties. Principal component analysis (PCA) derived from SNPs data. Different colours indicate region of origin as shown in the legend with symbols Discriminant Analysis of Principal Components (DAPC) clusters. b) The population structure of 262 rice landrace accessions and 62 improved varieties. Principal component analysis (PCA) derived from SNPs data. Different colours indicate region of origin as shown in the legend with symbols showing type of material.

1.3 Phenotypic characterization of cowpea and rice collections

Cowpea field trials were conducted at two locations, Nampula and Chokwe, where 351 genotypes were evaluated with two replications at each site. The experimental designs and phenotypic evaluation guidelines are outlined in the previous report (D2.4). Collected phenotypic traits include Growth Habit (GRH), Growth Pattern (GRP), Plant Pigmentation (PPG), Hundred Seed Weight (HSW), Grain Yield (GY), and Pod Weight (PW) **(Dataset 8).** The rice field experiment was conducted exclusively at the Chokwe station with a total of 220 rice genotypes selected for the trial. The traits evaluated include Basal leaf sheath color (BLS), Leaf sheath anthocyanin staining (LSA), Leaf lamina (presence of anthocyanin) (LLA), stem resistance (SRE), stem strength (SST), Hundred seed weight (HSW), Panicle length (PL) and Grain yield (GY) **(Dataset 9).**

1.4 Socioeconomic characterization of Nampula farmers

Throughout the project, socio-economic data was collected to ensure the implementation of activities according to the participatory framework adopted, which includes four phases: co-exploration, co-design, co-development, and co-evaluation. The co-exploration of users' backgrounds and needs was conducted through two Focus Group Discussions (FGDs) in October 2021 in the target communities of Namachepa and Rieque, involving 48 farmers, 24 men and 24 women (**Datasets 10 and 11; Table 1**). Additionally, a survey was carried out with 248 households in the Mogovolas District between September and October 2022 (**Dataset 12; Table 1**). The co-design and co-development phases were informed by a FGD with 24 participants (12 men and 12 women) in Rieque in May 2023 (**Dataset 13**). Finally, the co-evaluation of the six identified genotypes with potential for climate adaptation—an essential part of the participatory framework and detailed in subsequent section of the document—was conducted through two FGDs` in September and October 2024 in both Namachepa and Rieque, involving 24 men and 24 women **(Datasets 14 and 15).**

This project has received funding from the European Commission's Horizon 2020 Research and Innovation Programme. The content in this presentation reflects only the author(s)'s views. The European Commission is not responsible for any use that may be made of the information it contains.

Table 1. A synthesis of the FGDs conducted and the data collected

STUDY AREA	DATES	PARTICIPANTS	TOPICS
Namachepa and Rieque Communities	September 2021	20 (10 men and 10 women in each community)	• Socio-economic context • Cowpea use and cultivation • Cowpea diversity • Climate change perception and its impact on agricultural production • Availability and use of climate information
Rieque community	May 2023	24 (12 men and 12 women)	• Use of weather/climate information for cropping activities • Building cow pea seasonal calendars for the past years • Scenario simulation. Building cow pea seasonal calendars/crop management plan based on the 3 provided weather bulletins
Namachepa and Rieque Communities	September-October 2024	48 (12 men and 12 women in each community)	• Varietal traits appreciated by farmers • Assessment of cowpea genotypes with potential for adaptation to future climates • Scenario Simulations: Understanding How Farmers Combine Varieties from the Available Portfolio Under Different Climate and Weather Conditions

1.5 Participatory Variety Evaluation of cowpea and rice collections

In a previous report (D2.4), we outlined the methodology for conducting Participatory Variety Selection (PVS) on 220 rice genotypes at Chókwè in May 2023. Concurrently, PVS was conducted on 351 cowpea lines at the Chókwè and Nampula field stations. Nampula is in the northern part of Mozambique, while Chokwe is situated in the southern part. This multi-site evaluation for cowpea captured preferences that account for variations in climate, soil type and farming practices across these two regions. The PVS approach engaged both male and female farmers to evaluate cowpea varieties based on their preferences, aiming to identify gender-specific preferences and understand how these preferences influence farming practices and crop adoption. This process is linked with genetic diversity by correlating farmers' preferred traits with the genetic characteristics of the evaluated varieties, helping to identify genotypes that align with the needs of different farmer groups. Farmers in Nampula and Chokwe ranked each cowpea accession separately, with rankings conducted by men and women at each site (**Dataset 16; Dataset 17; Dataset 18 and Dataset 19).** Evaluations covered five traits: growth habit (GRH), terminal leaf shape (TL), branching (BR), pods per plant (PPP), and length of the growth cycle (CLY). Each trait was rated on a scale of 1 to 5, with 5 being the highest possible score. The ranking of rice accessions by male and female farmers involved evaluating key traits including Plant Height (PH), Tillering (TL), Lodging Resistance (LR), Grain Size (GS), Grain Color (GC), Aroma (AR) and Threshability (TBY) (**Dataset 20 and Dataset 21).** In Chokwe, gender sensitive focus group discussions (FGD) were conducted after PVS, to assess farmers' perceptions and rationale behind their preferences in cowpea and rice varieties. Farmers participating in FGDs highlighted preferred traits for cowpea (**Table 2**) and rice cultivation (**Table 3**). The gender-specific preferences outlined in this study highlight the necessity for inclusive agricultural development strategies that recognize and address the priorities of both women and men. By integrating these preferences into

breeding programs, training initiatives, and policy frameworks, stakeholders can enhance food security, improve community resilience, and foster sustainable agricultural practices. Acknowledging and leveraging these differences will be crucial for the effective implementation of agricultural innovations that are both acceptable and beneficial to the entire community.

Table 2.Farmer-Preferred Cowpea Traits Discussed by Farmers During Focus Group Discussions in Chokwe

Cowpea Trait	Women	Men
Growth Habit	Both groups preferred prostrate varieties due to their ability to produce for an extended period. These varieties, unlike determinate varieties, continue to generate leaves and pods, ensuring multiple harvests.	
Grain Size:	Both men and women emphasized the importance of large-grain varieties for their market appeal and ease of cooking.	
Leaf Characteristics	Farmers in both groups preferred cowpea varieties with tender, less fibrous leaves. These are easier to cook and are consumed for longer periods during the crop's growth cycle.	
Market Preference	Both groups indicated that large-grain varieties are more popular in the market, driving their preference for larger-seeded varieties that can be sold at a higher price.	
Crop Cycle	Long-cycle varieties for drought resistance and food security.	Short-cycle varieties for quicker harvest and crop rotation.
Use of Biomass	Focus on food production (leaves and pods) over a long period.	Preference for highly branched varieties for more biomass and weed control.
Pod Characteristics	Preferred varieties with more pods and large grains for ease of cooking.	Fewer pods with larger grains for traditional dish "Nkululo."
Preservation Methods	Traditional preservation (boiling and sun-drying leaves).	Preservation using refrigeration and sun-drying.
Cooking Preferences	Broad, tender leaves that regrow after harvesting.	Lanceolate (Sacana) leaves that cook well with fewer ingredients.
Climate Resilience	Long-cycle, drought-resistant varieties with deep roots.	Short-cycle varieties that allow faster crop rotation and resilience.

Table 3. Farmer-Preferred Rice Traits Discussed by Farmers During Focus Group Discussions in Chokwe

Rice Trait	Men	Women
Yield	High priority: yield considered most important	
Tillering & Panicle Fullness	Appreciated for its impact on yield and profitability	
Grain Size & Weight	Prefer long, heavy grains for better market sales	Prefer heavy grains, weight more important than size
Grain Colour	Mentioned but not emphasized as a critical factor	Strong preference for white grains (market and home use)
Grain Aroma	Valued for market sales and family consumption	Highly valued, with strong emphasis on aroma for home and market
Resistance to Lodging	Critical for reducing losses and ease of harvesting	Critical due to labour concerns and potential grain loss
Crop Cycle Length	Short cycle preferred to enable crop rotation and prevent pest issues	Short cycle critical to facilitate crop rotation and lower labour costs
Pest Resistance	Focused on rice damage from birds and rats	Concerned about pest impact on both rice and other crops
Plant Height	Tall plants prone to lodging and wind damage, but not a major focus	Dislike tall plants, citing risk of lodging and physical harm
Market Orientation	Market-focused, but yield and resilience are top concerns	Strong market orientation, with emphasis on aesthetics, aroma, and market demand

1.6 Bioclimatic characterization of cowpea and rice collections

We analyzed local climate conditions using 30 years of historical data (1970–2000) from WorldClim Version 2.1, a high-resolution global climate dataset widely used in agricultural studies for its consistency, global coverage, and relevance to crop suitability and climate impacts. WorldClim provides high-resolution (up to 1 km²) historical climate data derived from weather station observations, making it ideal for fine-scale landscape genomics and ecological studies. It offers long-term data and a broad range of bioclimatic for understanding the relationship between genetic diversity and environmental factors. Compared to reanalysis datasets like ERA5, which have coarser spatial resolution (~30 km) and require additional processing for ecological modeling, WorldClim is better suited for capturing localized climatic variations influenced by topography. The bioclimatic variables from WorldClim include temperature and precipitation-derived data such as precipitation derived variables bio1 (annual mean temperature), bio2 (mean diurnal range),bio3 (isothermality), bio4 (temperature seasonality), bio5 (max temperature of warmest month),bio6 (min temperature of coldest month),bio7 (temperature annual range),bio8 (mean temperature of wettest quarter),bio9 (mean temperature of driest quarter),bio10 (mean temperature of warmest quarter),bio11 (mean temperature of coldest quarter),bio12 (annual precipitation),bio13 (precipitation of wettest month),bio14 (precipitation of driest month),bio15 (precipitation seasonality),bio16 (precipitation of wettest quarter),bio17 (precipitation of driest quarter),bio18 (precipitation of warmest quarter) and bio19 (precipitation of coldest quarter). Our analysis focused on selected, non-collinear bioclimatic variables, specifically: bio2 (mean diurnal range), bio3 (isothermality), bio5 (maximum temperature of

the warmest month), bio12 (annual precipitation), bio14 (precipitation of the driest month), bio15 (precipitation seasonality), and bio18 (precipitation of the warmest quarter).

We observed significant climate variability across southern African regions where cowpea and rice are traditionally grown—an essential factor for identifying alleles potentially adapted to local environmental conditions (**Figure 2**). Detailed climate data is available for cowpea and rice sampling points (**Dataset 22 and Dataset 23**) according to the passport data collected in **Dataset 1 and Dataset 2**.

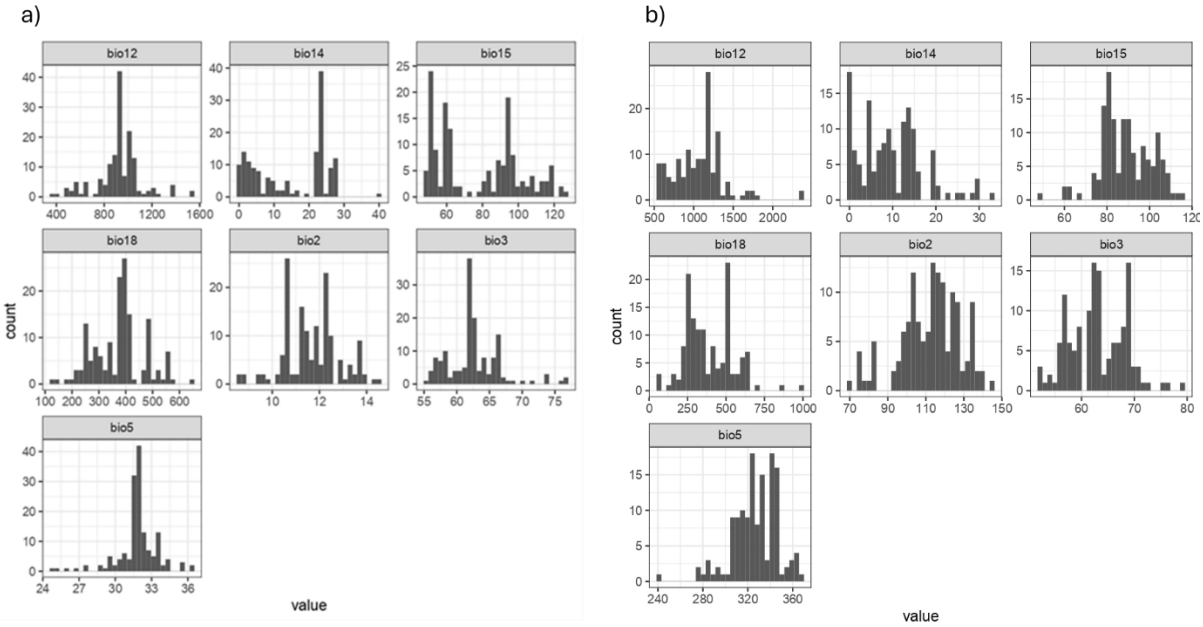

Figure 2. Bioclimatic variables distribution in a) cowpea and b) rice growing areas of southern Africa. Histograms showing frequency of selected bioclimatic variables including bio2 (mean diurnal range): °C (degrees Celsius), bio3 (isothermality, the ratio of the mean diurnal temperature range to the mean annual temperature, reflecting the stability of daily temperature fluctuations): Unitless (ratio, expressed as a percentage), bio5 (maximum temperature of warmest month): °C (degrees Celsius), bio12 (annual precipitation): mm/year (millimeters per year), bio14 (precipitation of driest month): mm/month (millimeters per month), bio15 (precipitation seasonality): Unitless (coefficient of variation), bio18 (precipitation of warmest quarter): mm/quarter (millimeters per quarter). The x-axis represents the range of values for each bioclimatic variable, while the y-axis represents the count, or the number of occurrences, within each value range.

1.7 Integrated selection of genotypes with adaptation traits for resilience

This section outlines the selection framework for prioritising a subset of resilient cowpea genotypes, combining farmers' participatory rankings with phenotypic, genotypic and climatic assessments. The first criterion of selection was based on farmers' evaluations of genotypes through field observation**.** PVS rankings of genotypes from men and women were combined for each location. To qualify, a genotype had to achieve an average score of 4 out of 5 and rank within the top 50 accessions in each gender's scoring. This combined analysis resulted in the identification of 35 genotypes in Nampula (**Dataset 24**) and 27 genotypes in Chokwe (**Dataset 25)**. The second step involved filtering these genotypes based on the phenotypic data and insights from focus group discussions, thus , cowpea accessions with the following trait were selected; a prostrate growth habit (GRH) for adaptability, Indeterminate growth pattern (GRP), and high-yield potential for traits including high pod weight (PSW), hundred-seed weight (HSW) for grain quality, and overall grain yield (GY) to meet market and productivity needs . Finally, for the genotypes with climatic data, we specifically selected those representing a diverse range of bioclimatic variability, ensuring adaptability to different environmental conditions. This resulted in a total of 32 selected genotypes, including 25 landrace genotypes from Mozambique, 2 improved varieties from IIAM and IITA genebanks, 1 landrace from Tanzania, and 4 landraces from Malawi.

A field trial was conducted to further evaluate these 31 accessions. The on-station trial was conducted in two replications at the IIAM fields in Chokwe during the growing season 2023/2024. The comprehensive phenotypic trials included assessments of the following traits: Growth Habit (GRH), Growth Pattern (GRP), Leaf Color Intensity (LCI), Leaf Area (TL and TW), Leaf Shape (TLS), Days to Flowering (DF), Days to First Maturity (DFM), Flower Color (FC), Immature Pod Pigmentation (IPP), Pod Length (PL), Number of Locules per Pod (NLPP), Seed Weight (SW), Total Seed Weight (TSW), and 100 Seed Weight (HSW) (**Dataset 26**). Clustering analysis was conducted using the phenotypic traits collected for 31 cowpea genotypes to identify distinct groups based on their phenotypic trait similarity. This resulted in 3 groups with, cluster 3 having only one genotype (**Figure 3; Dataset 26**).

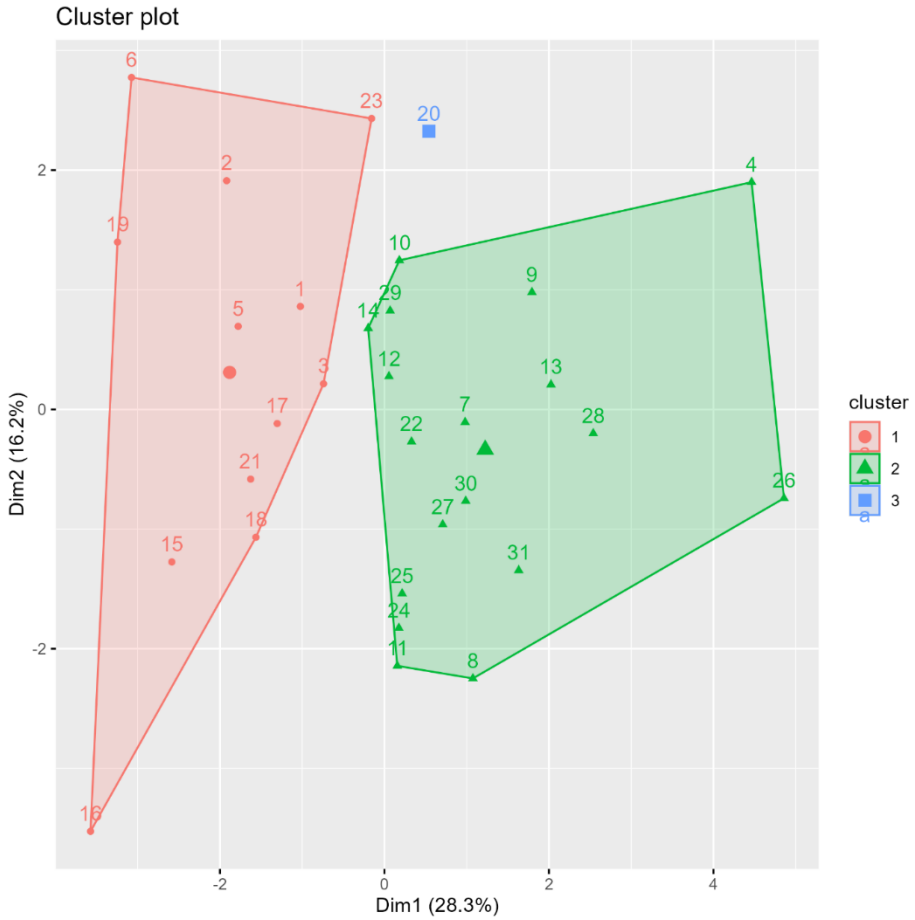

Figure 3. Cluster plot of 31 cowpea accessions based on their Phenotypic traits

2 Examples of Data Usage

This section provides real-world examples of the use of the generated data within CS3. It is not meant to be a comprehensive list or even an endpoint, as much work can still be done to distill information from the collected data. However, the four examples outlined here may illustrate the application of the data generated in various domains. First, molecular markers are used in combination with climate data to assess the performance of rice genotypes under predicted future climate data from climate models, often based on Shared Socioeconomic Pathways (SSPs)to identify genotypes resilient to future environmental changes. Second, data from genomic sequencing and phenotypic characterization of plant materials is used in a Genome-Wide Association Study (GWAS) to identify genetic markers linked to important seed traits in cowpea. Third, agroclimatic indices are derived that most significantly affect rice and cowpea production in Mozambique, specifically focusing on vapor pressure deficit, longest dry spell, number of heavy rain days, and number of extreme hot days. As a final point, community engagement is emphasized by integrating an analysis of farmers' decision-making in variety selection under climatic pressure such as delayed onset of the rainy season.

2.1 Assessing the potential impact of climate change on the cultivation of rice in southern Africa

To ensure sustainable food security it is crucial to understand the vulnerability of crop cultivation under future climate conditions. The CS3 team collected historical bioclimatic data from rice sampling points within southern Africa (**Dataset 23**). We also generated genomic data from these rice accessions (**Dataset 6**). By applying landscape genomics methods, we linked genomic data to the climatic diversity at the sampling locations. This allowed us to explore the relationship between genetic diversity and both historical and projected climate conditions. In short, we employed a Gradient Forest (GF) machine learning algorithm to assess how well the current rice genetic diversity aligns with the requirements for future climates and identify critical areas of genetic diversity that may contribute to resilience under future conditions.

The GF was modelled utilizing past climate data and MEMs (Moran's Eigenvector Maps) as predictors, with 1296 SNPs as response variables (**Dataset 7**). This approach measures the allelic turnover, which is the change in allele (gene variants) frequencies across populations, locations, or environments. It helps determine how genetic diversity shifts across different regions or climates, showing possible adaptations. In this study, the allelic turnover in the rice collection was best predicted by MEM4 describing spatial distribution (**Figure 4a**). Among the bioclimatic variables, isothermality (bio3) was the one with the highest importance, followed by mean of diurnal range (bio2) (**Figure 4a**).

Next, we tested the performance of the rice core collection in a projected climate. The expected genomic composition was modelled under the SSP 585 (2041-2060) climatic scenario, which represents a significant increase in global temperatures. High-vulnerability areas were identified in Zimbabwe, Malawi, southern Tanzania, and Mozambique. In Mozambique, we observed the highest offset observed in Tete, Niassa, and Zambezia provinces and perhaps because the existing collection does not fully represent these regions (**Dataset 1**). This underscores the need to expand the plant genetic resources collection to enhance diversity and support climate resilience.

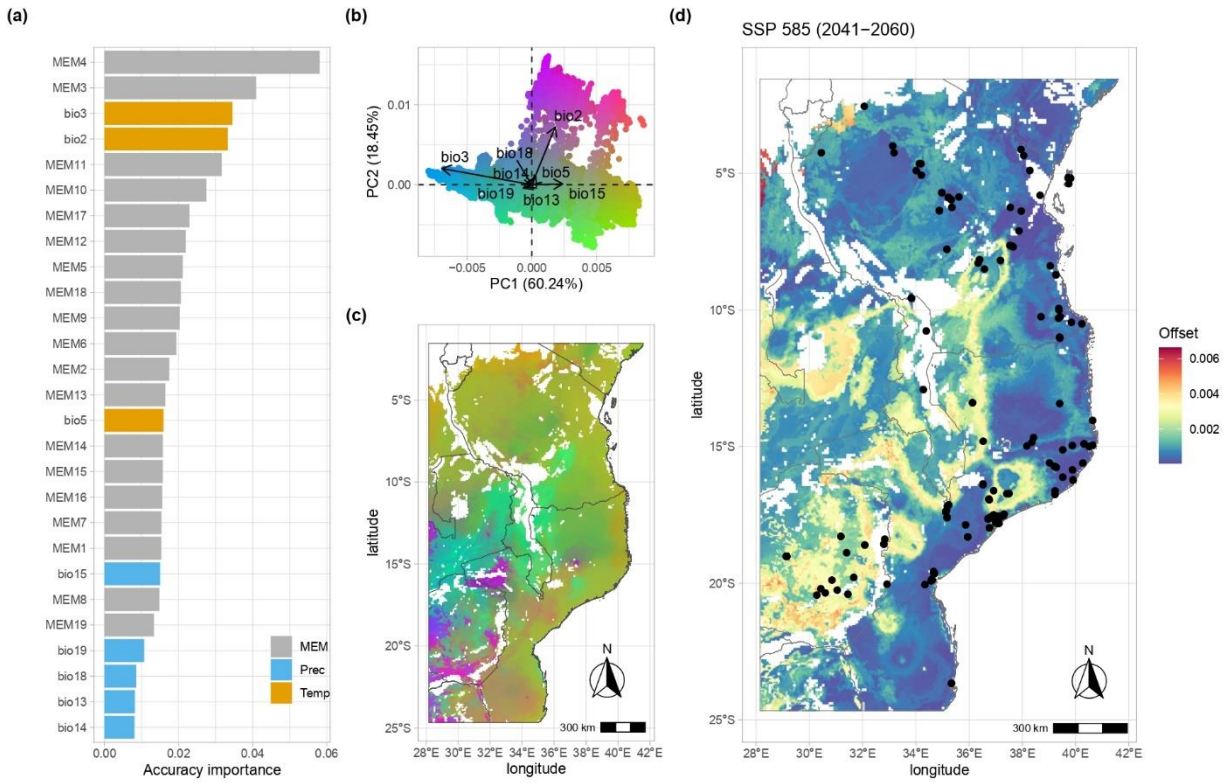

Figure 4. The relationship between bioclimatic and spatial diversity and genomic variation of rice landraces in southern Africa. a) The accuracy importance of bioclimatic and spatial variables is ranked based on the Gradient Forest (GF) model. b) A biplot of the biological space displays Principal Components of the transformed bioclimatic predictors using an RGB colour palette relative to its first three dimensions. c) GF-transformed bioclimatic variables across landscape (southern Africa). Colours correspond to those on bioclimatic-genetic space in 'b'. d) Genomic offset based on Shared Socioeconomic Pathway (SSP) 585 (2041-2060) climate projections. The colour scale illustrates the extent of the discrepancy between present and anticipated shifts in allele frequencies driven by climate change.

2.2 Genome-Wide Association Study of Seed Traits in Cowpea Pre-Breeding Accessions

In crop breeding, Genome-Wide Association Study (GWAS) is used to identify genome regions linked to desirable traits, such as yield, disease resistance, and climate adaptation. By analyzing the genome of a diverse population, GWAS helps breeders select individuals with favorable genes, enhancing the likelihood of improved traits being passed on to offsprings. The prerequisite for a GWAS study is high quality genotypic data and robust phenotypic data. As such, we utilized cowpea genotypic data generated from CS3 (**Dataset 4**). For the phenotypic data, a subset of 106 landrace accessions derived from the cowpea breeding program of Eduardo Mondlane University in Mozambique were selected from the core collection (**Dataset 1**).

We selected healthy seeds from each accession and performed automated phenotyping. This method involved taking high quality standardized photographs and subjecting them to a seed Extractor software that was used to measure traits for seed size and shape including Seed Area (SA), Major Axis Length (MAL), Minor Axis Length (MiAL), Seed Circularity (SC) RGB Colour, and seed colour(RGB) (**Figure 5**). Other traits including Seed shape (SHP), Seed Coat Pattern (SCP), Seed Coat Colour (SCC), Hundred Seed Weight (HSW) were scored manually. The derived seed-size traits were used for a GWAS

analysis. A standard pipeline for GWAS data analysis was followed and is outlined in the D2.4 report. R/corrplot (Taiyun 2017) was used to study the correlation among seed phenotypic traits.

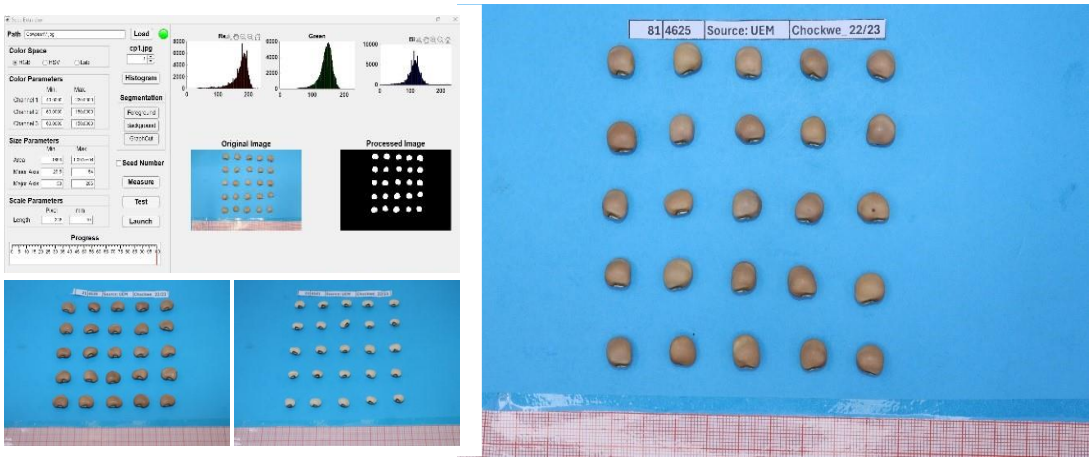

Figure 5. Shows seed photographs and measurement of cowpea seed from using a seed extractor software

The correlation analysis revealed several key relationships between the seed qualitative and quantitative traits (**Figure 6**). We observed a highly positive correlation between Seed Coat Pattern (SCP) and Seed Coat Colour (SCC), this is expected in cowpea as certain colors are often associated with specific patterns. This allows breeders to achieve consistent patterns and colors that meet specific market or farmer preferences. We observed a moderate positive correlation (0.53) between seed area (SA) and hundred seed weight (HSW), indicating that as seed area increases, there tends to be a corresponding increase in the hundred seed weight. However, the correlation is moderate indicating there will be some variability as expected with quantitative traits. We observed a significant negative correlation between Seed circularity (SC) and seed axion ration (SAR) (-0.61), meaning that seeds with higher circularity (SC) (which refers to how closely the shape of the seed approaches that of a perfect circle) are likely to have a lower axion ratio (which typically compares the length and width of the seed). The significance of seed shape includes agronomic performance, marketability/preference and processing efficiency making it an important consideration for both breeders and farmers. As expected, seed size related traits showed high to moderate positive correlation.

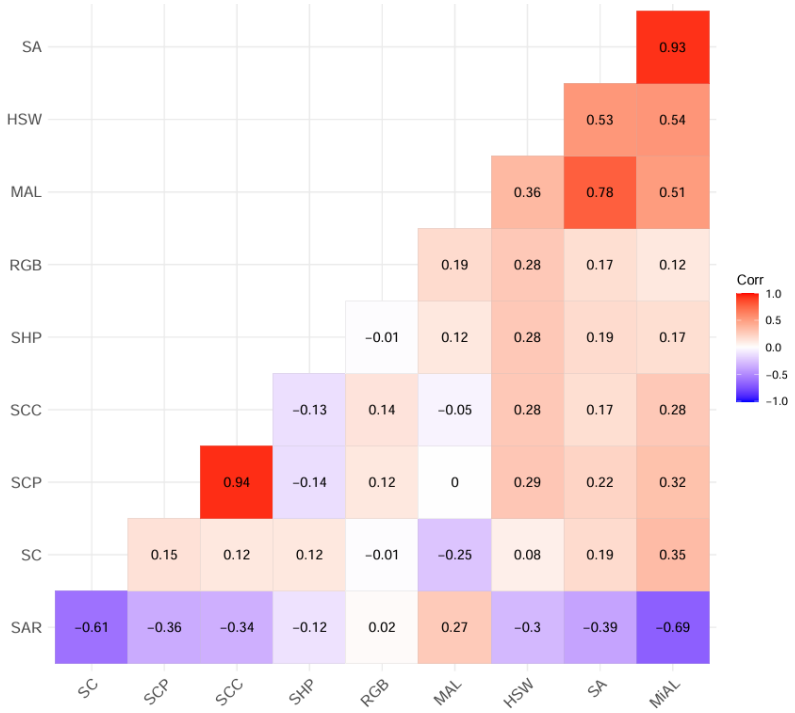

Figure 6. Correlation Analysis between all traits measured on cowpea seed, namely seed shape (SHP), Seed Coat Pattern (SCP), Seed Coat Colour (SCC), Hundred Seed Weight (HSW), Seed Area (SA), Major Axis Length (MAL), Minor Axis Length (MiAL), Seed Circularity (SC), and Seed Axion Ration (SAR). Correlation coefficients, ranging from -1 to +1, are color-coded from blue (negative correlation) to red (positive correlation), with values close to ±1 indicating strong relationships.

The main aim of the GWAS analysis was to identify SNP markers (genetic variations) associated with the seed traits evaluated. These markers are usually found throughout the entire genome and can assist in locating genes or regions of the genome which in turn are useful to breeders for developing robust crop varieties. The SNPs identified are usually referred to as quantitative trait nucleotides (QTNs) as they are potential key genetic variants that may contribute to the observed differences in quantitative traits, between crop accessions. In this study, we identified a total of 11 unique QTNs associated with 4 traits including HSW, SA, SCP and SCC. Trait HSW, had 3 QTNs on chromosomes 7 and 9 (**Figure 7**). Interestingly, highly positively correlated traits SCC and SCP shared three QTNs on chromosome 2 (at locus 14.5Mb), 5 (at locus 3.1Mb) and 10 (at locus 14.8Mb). This is significant as it allows breeders to more efficiently achieve desired combinations of seed coat color and pattern in cowpea by targeting specific points in the genome.

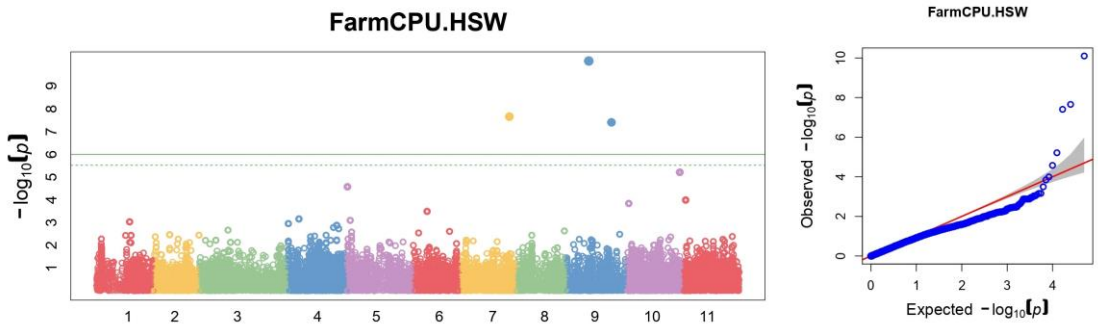

Figure 7. GWAS output for the trait hundred seed weight (HSW). The Manhattan plots report individual SNPs across all chromosomes (x-axis) and -log10 P value of each SNP association (y-axis). The horizontal lines represent a stringent Bonferroni threshold for a nominal p-value of 0.05. points surpassing the threshold represent the position on the genome in which diversity in the collection is strongly associated to the trait. The accompanying QQ plots on the rights demonstrate a good GWAS model fit.

The results generated by GWAS are a step towards unravelling the genetic mechanism behind complex agronomic traits in cowpea, with the promise of accelerating the development of new varieties addressing farmer needs. The seed traits analyzed here have high breeding relevance, as they reflect genetic, physiologic, and ecologic variation that exceed seed morphology. We demonstrate that it is possible to leverage seed phenotyping data from genebanks or breeding programs for GWAS. This approach presents significant advantages as it is cost- and time-efficient, eliminating the need for extensive field trials. Early trait evaluation through seed phenotyping offers insights into characteristics such as seed size, weight, and germination potential, enabling quicker decision-making. In addition, seed phenotyping is automated and allows for high throughput analysis of thousands of uncharacterized plant genetic resources, reducing the logistical challenges associated with field trials.

2.3 Agroclimatic Characterization of areas of cowpea and rice cultivation

The tropical to subtropical climate of Mozambique is favorable to crop production. Rain reaches an average amount of 800 – 1200 mm, with unimodal season which covers from November to April. Average annual temperatures in warmer regions range from 20 to 27°C, and in cooler regions range from 15 to 25°C (USAID, 2018). This climate supports the cultivation of various rainfed crops, such as rice and cowpea. However, crop productivity in Mozambique is far from reaching its potential for several reasons, including climate variability. Crop production in Mozambique is dominated by rainfed smallholder farming which makes it sensitive to climatic stress. CS3 conducted socioeconomic studies (**Dataset 10-13**) to gain insight into climate perception among smallholder farmers, examining how these perceptions influence their agricultural practices. The studies highlighted the impact of altered rainfall patterns, with many farmers reporting a significant reduction in the crop growing period and increased difficulty in predicting the start of the rainy season. Farmers particularly identified the late onset of the rainy season as a major challenge for cowpea cultivation. Building on these findings from surveys and focus discussions, the CS3 team aimed to quantify and characterize key agroclimatic indices impacting rice and cowpea production, tailored to address local farmers' needs and focusing

on adaptation strategies and resilience-building to sustain agricultural productivity under variable and changing climatic conditions.

Crop genotypes have distinct climatic requirements and stress tolerance ranges. The indices help to identify which genotypes may thrive for a specific location under a given agroclimatic condition. Crops such as rice and cowpea can be exposed to various climatic stress, including water deficit and extreme heat, which limit growth and productivity. Effective agroclimatic characterization requires an understanding of the critical climatic determinants that govern crop performance and how these factors are quantified through indices. These indices are time-averaged values or the frequency of climatic events surpassing a specific threshold, either for a single climatic variable or a combination of variables. In CS3 we characterized a set of agroclimatic indices describing climatic determinants and stressors during crop growing period in Mozambique. The agroclimatic indices are derived from the AgERA5 dataset, a gridded daily historical climate data consisted of the daily maximum temperature, minimum temperature and precipitation. Crop specific critical thresholds were subjectively defined aided by expert understanding of the physiology of crops and their interaction with climatic condition. Each index was tailored to capture specific stressors. For instance, extreme heat indices were based on maximum temperature thresholds that correspond to heat tolerance levels in rice and cowpea. **Table 4** describes key agroclimatic indices that impact rice and cowpea production.

Table 4. Description of selected agroclimate indices

Indices	Name	Definition	Unit
VPD	Vapor pressure deficit	the difference between the amount of moisture that's in the air and the amount of moisture that air could hold at saturation	kPa
LDS	Longest dry spell	Maximum number of consecutive dry days (when PR < 2.0 mm)	days
R20mm	Number of heavy rain days	Number of days when PR >= 20 mm	days
TXge35	Number of extreme hot days	Number of days when Tx >= 35 °C	days

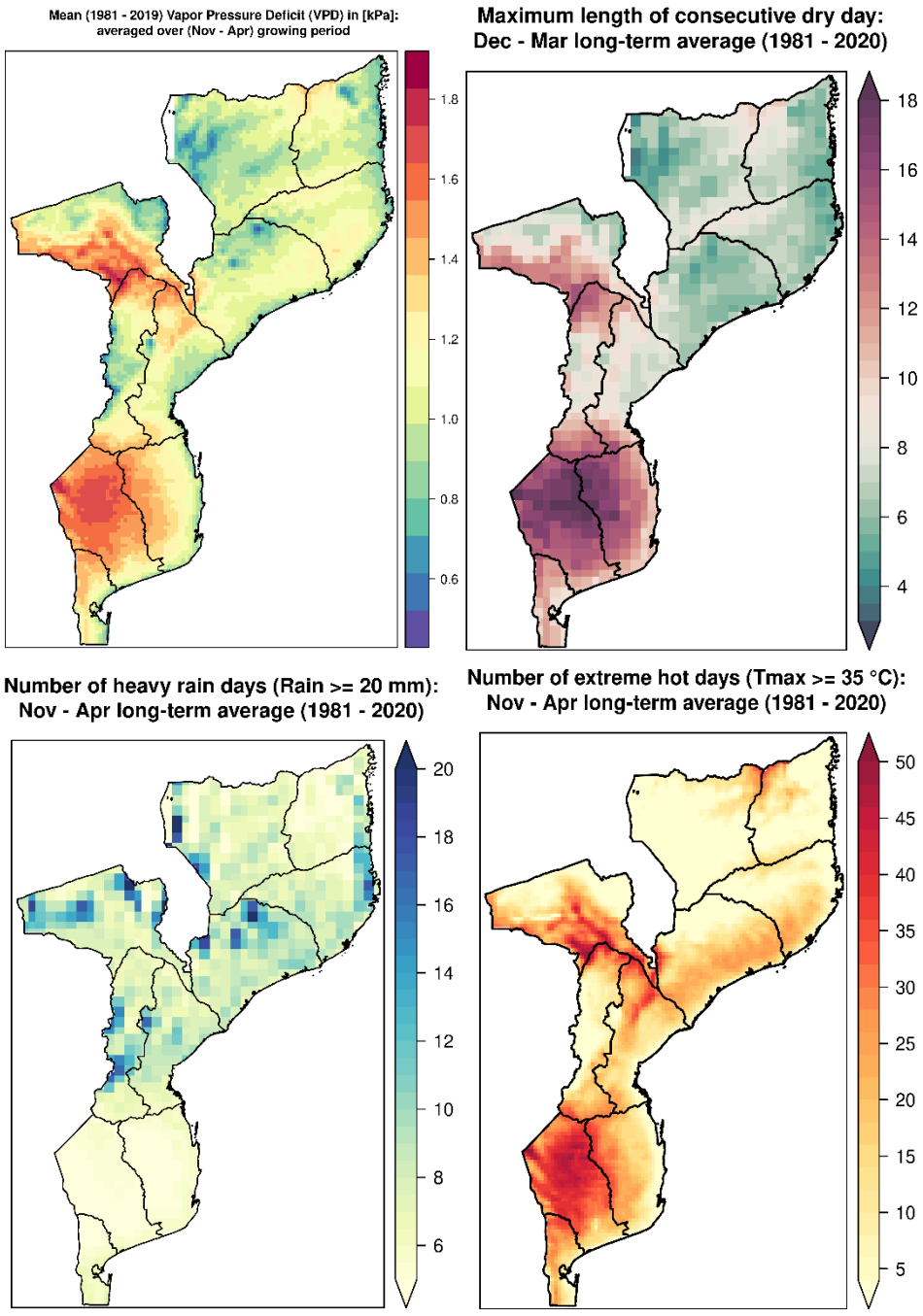

Figure 8. Long-term (1981 - 2020) mean climatology of agroclimate indices

The agroclimatic characterization is intended to support decision making in crop production through tailored climate information. The results are designed to benefit all stakeholders in the agricultural sector, from decision makers to smallholder farmers, who must receive clear and relevant information for their agronomic decisions. However, attention is placed on the analysis of the agroclimatic conditions that most influence Mozambique's rice and cowpea production. While natural climate variability necessitates careful agronomic planning, these indices offer quantitative insights to support

informed decision-making. This information must be constantly updated to give end users the very latest available and accurate information on climate evolution. Feedback from end users, such as local associations and decentralized agricultural technical services, has been a guide to fine tuning of the agroclimatic indices, making them more aligned with local needs and practices.

2.4 Understanding smallholders' decision-making for varietal selection under different climate conditions

This study is grounded in socioeconomic results obtained from CS3 activities. Using data from focus group discussions and surveys (**Datasets 10 to 15**), we reconstructed the operational context of local farming systems, particularly regarding cowpea cultivation. The first step involved identifying smallholder farmers' use of cowpea and their preferences for specific traits. Farmers described their preferred observable traits related to the phenotypic expression of genotypes (**Table 5**). We narrowed our analysis to six key traits, as other important factors, such as aroma and taste, could not be assessed in this study. The six traits identified for further study are: leaf size, growth pattern, pod length, length of the cycle, seed color, and seed size.

Table 5. Traits and characteristics considered by farmers for cowpea selection (FGDs, 2024)

Rieque		Namachepa	
Men	**Women**	**Men**	**Women**
Leaf shape and size	Length of the cycle	Seed size	Seed size
Taste	Leaf size	Leaf	Growth habit
Seed colour	Pods	Seed colour	Yield
Pods size	Market demand	Growth habit and pattern	Pod size
Yield	Seed size	Yield	Leaf size
Growth pattern	Seed colour	Lenght of the cycle	Seed colour
Seed size	Growth habit	Pods	Lenght of the cycle
Length of the cycle	Resistance to diseases	Post-harvest resistance	Aroma
	Resistance to drought	Germination rate	Market price
	Seed availability		Seed availability
			Resistance to diseases

48 farmers (24 men and 24 women) were asked to rank selected traits/characteristics, indicating their importance for varietal selection with a score from 1 (not important) to 5 (very important). The results were heterogeneous, both when comparing men and women, and when comparing the two target locations (**Table 6**).

Table 6. Ranking of the selected traits based on the aggregation of the scores assigned by each participant for each trait. (FGDs, 2024)

	Rieque			Namachepa					
	Men	Women	TOT Rieque	Men	Women	TOT Namachepa	Tot. Men	Tot. Women	Tot.
Leaf size	25	54	79	39	50	89	64	104	168
Growth pattern	41	63	104	50	59	109	91	122	213
Pod length	53	58	111	51	64	115	104	122	226
Length of the cycle	46	62	108	40	62	102	86	124	210
Seed colour	23	35	58	54	64	118	77	99	176
Seed size	55	56	120	50	65	115	105	121	226

The traits that received the highest scores are *pod length* and *seed size*, while the least preferred trait is *leaf size*. However, disaggregated data reveals that the least preferred trait in Rieque—*seed color*—is the most appreciated in Namachepa. Looking at gender differences, no significant patterns are observed within either gender. In Rieque, women favored the *growth pattern* (linked to dual use for leaves and grain), while in Namachepa, they preferred larger *seed size*. Men expressed a preference for *seed size* in Rieque and *seed color* in Namachepa, respectively. In general, it is possible to note a greater homogeneity in preferences within each of the two communities (Namachepa and Rieque) compared to those observed across gender dimensions.

The second step involved mapping climate change factors affecting cowpea production. According to the survey conducted in 2022, 93% of interviewed farmers perceived climate changes over the past 10 years, impacting cowpea cultivation. Specifically, 93.1% of farmers observed changes in rainfall patterns. Due to these altered weather patterns, 81.5% of farmers reported a significant reduction in the crop growing period, and an equal percentage found it more challenging to predict the start of the rainy season. Notably, 87.2% of farmers identified the late onset of the rainy season as one of the major problems for cowpea cultivation. With this information, the final FGDs (2024) were thus dedicated to analyzing decision-making related to varietal selection. It included an assessment of preferences for six genotypes identified as potentially adaptable to future climates (through genetic and climatic analyses) but at the same time entailing farmers' preferred phenotypic traits (identified through PVS and previous FGDs). A decision-making simulation for cowpea variety selection under different climatic conditions was also organized to identify the role of varietal selection in adapting to climate change

For the assessment, and considering the possible presence of participants who might not read or write, the traits/characteristics of the genotypes were presented through:

- Sample of each genotype;

An "ID card" created for each genotype. These cards were designed to present farmers with the characteristics of each genotype, which was physically showcased during the FGD using a seed sample. Each card visually displayed the following phenotypic characteristics for each genotype: leaf size, growth pattern, pod length, length of the cycle, seed color, and seed size (see Fig. 9). For example, genotype number 6 is characterized by lanceolate leaves, an indeterminate growth habit, medium pods, an intermediate maturity cycle, cream-colored seeds, and small seed size. This approach enabled farmers to evaluate the genotypes comprehensively, rather than basing their assessment solely on the observable characteristics of the seeds.

Figure 9. The panel on the left shows the ID cards associated with their sample for display. The right panel shows FGD with women (Rieque)

The 6 genotypes were then presented to the farmers, and the first question concerned whether they recognized them. All genotypes were unanimously recognized except for Genotype nr. 6. When it comes to farmers' recognition of these varieties, farmers categorize cowpea varieties into two types/groups: "*Namurua*" (short cycle, produces up to a maximum of two times, smaller seeds and leaves, erect growth habit) and "*Cute*" (longer cycle, produces multiple times and does not dry out easily, bigger seeds and leaves, prostrate growth habit, generally appreciated for its taste). These genotypes were therefore identified as "*Namurua*" or "*Cute*", specifying their characteristics (e.g., Nr. 5: Red Cute (*Cute vermelho*)).

Farmers were then asked to rank the 6 genotypes, indicating their appreciation with a score from 1 (I don't like it at all) to 5 (I like it very much) (see Table 7).

Table 7. Ranking of the 6 genotypes by the 2 target communities (Namachepa and Rieque, September 2024)

	Rieque		Namachepa		
	Men	**Women**	**Men**	**Women**	**Total score**
Genotype nr. 1	64	65	53	65	247
Genotype nr. 2	66	50	31	40	187
Genotype nr. 3	67	54	48	46	215
Genotype nr. 4	17	53	23	32	125
Genotype nr. 5	28	55	55	44	182
Genotype nr. 6	35	62	50	45	192

Looking at the sum of the individual scores, the genotype that obtained the highest score is number 1, being preferred by both the women of Rieque and Namachepa. This genotype is known in the area as “Cute” or “Cute grande”. The reasons given by the participants for this preference are mainly related to its consumption (larger grain, tasty, “fills the pot”, cooks quickly, the pods can also be eaten fresh and are very filling, farmers greatly appreciate its leaves as well as its marketing potential (it is highly appreciated in the market). It is interesting to note the score attributed to genotype number 6, which was not known or used by the participants in the FGDs. The score is, in fact, the third highest among all genotypes, and all participants expressed interest in trying it, especially due to its attractive appearance.

The second part of the 2024 FGDs was dedicated to the varietal decision simulation. It was conducted individually by each participant, considering a plot of land of 1 hectare, where it was possible to select up to 4 cowpea varieties in a single growing season, chosen from a portfolio containing both known/owned varieties and those presented during the assessment. 46 out of 48 participants presented a varietal selection based on the succession of *namurua* (for the first rains) and *cute* (for the second part of the season), or a combination of *namurua* and *cute* in both the first and second parts of the season, validating the decision to start the season with *namurua* to ensure nutrition during periods of food scarcity. One participant stated that they would start the season with another type of bean (*holoco*), while another participant indicated that his/her household would begin only with *cute*.

When asked to reassess their decision based on the provided information of the early or late on-set of the rainy season (i.e., bulletin of the upcoming rainy season), the participants stated that they did not wish to change their varietal combination. Men in Namachepa and Rieque reported that they would only change the month of the first planting, synchronizing it with the rainfall. Women reported that in the case of early rains or considering the risk of flooding, they would not plant *namurua* varieties. Instead, they would plant rice or sweet potatoes, as these are tolerant and resilient crops to flooding. In the case of delayed and short rains, they would change the method and location of planting. They would plant in low-lying areas, in beds where there is still moisture. When asked which varieties among those they know or the six varieties presented they consider potentially more resilient to climatic challenges (early or late onset of the rainy season, potential droughts or flooding), they unanimously responded that the short-cycle varieties that allow for greater flexibility (i.e., *namurua*) are more likely to perform better in these conditions.

Conclusion

CS3 aimed to enhance food security and resilience among smallholder farmers by identifying crop varieties capable of thriving under changing climate conditions, ultimately supporting sustainable agricultural practices in the region. The preliminary analysis of the data collected highlights a complex relationship between varietal choices and climate change. Firstly, it is possible to note both the existence of patterns and heterogeneity in the expressed preferences. The key takeaway message is that future breeding initiatives should first identify different categories of farmers and accommodate their varying preferences (both transversal and gender/local-specific) through the identification of a diverse portfolio of varieties.

However, the varietal decision-making analysis shows that choices are mostly rigid, and strongly influenced by the length of the production cycle (*namurua* short cycle varieties are always cultivated first), despite farmers' consumption preferences or the specific conditions of the upcoming rainy season. This remains true even in the presence of information about the characteristics of the upcoming rainy season, provided during the simulations, for instance, through bulletins. When it comes to the adaptation strategies to weather variability and climate change, therefore, we can cautiously suggest that future project initiatives should not exclusively rely on the identification of the most resilient and preferred varieties. It is essential to preliminarily identify different categories of farmers according to their "*climate smart readiness*." For the less ready, namely the most exposed to food insecurity, it is essential to identify complementary actions, potentially aimed at reducing risk aversion to encourage the uptake of innovation (such as new resilient varieties and weather information). Specifically, two lines of actions were identified: i. actions directly focused on increasing food security for less "*climate smart ready*" farmers, such as including nutritional components and yield results (as a way to generate additional income) when experimenting with the new varieties and ii. field-based activities, such as demonstration fields, which can, for instance, show the potential of the identified varieties in-situ and reduce the perceived risk associated with experimenting innovations.

We believe that what we have outlined above provides a roadmap for implementing transdisciplinary strategies aimed at strengthening climate resilience in smallholder communities, not only in Mozambique but also in other regions facing similar challenges. A key achievement of CS3 has been the funnel-based approach to data integration, which began with the acquisition of core collections from gene banks, progressed through the extraction of genomic, climatic, and agronomic data, and culminated in the evaluation of these varieties by farmers. This systematic process has generated a robust set of findings that deepen our understanding of the Mozambique system and offer valuable insights with broader applicability. The data catalogue developed through CS3, integrating genomic, climatic, and agronomic data, can support further research into climate resilience and crop adaptation. While not an endpoint, this catalogue provides a valuable resource for exploring new avenues in breeding programs and may assist policymakers in designing more effective initiatives for enhancing agricultural sustainability and food security in smallholder farming systems in Mozambique and beyond.