Mining valuable genetic variants for climate change adaptation from the world's largest cassava collection

Monica Carvajal-Yepes[1*], Jorge Franco[2*], Jessica A. Ospina[1], Monica Velez[1], Kai Sonder[3] and Peter Wenzl[1]

(1)Alliance Bioversity &CIAT, Cali, Valle del Cauca, Colombia, (2) CIAT,CIMMYT, CGIAR, Paysandu, Uruguay, (3) International Maize and Wheat Improvement Center, Carretera México-Veracruz, Km. 45, El Batán, 56237 Texcoco, México

(*) These authors contributed equally to this work

Background

Climate change is challenging crop production, threatening food security and the livelihoods of billions of people. A total of 13,518 cassava landraces are conserved by nine genebanks globally (***Figure 1A***). Among these, the International Center for Tropical Agriculture (CIAT) in Colombia, conserves 4,947 landraces from 27 different countries (***Figure 1B***). These landraces, selected over thousands of years, are a premier source for novel alleles conferring abiotic stress tolerance. To support a fast and cost-effective deployment of climate-adaptive alleles from the CIAT collection, 3,050 cassava landraces with geographic coordinates were used to extract variables characterizing the climate at their collection sites for an environmental Genome-Wide Association Study (EnvGWAS) (***Figure 1C***).

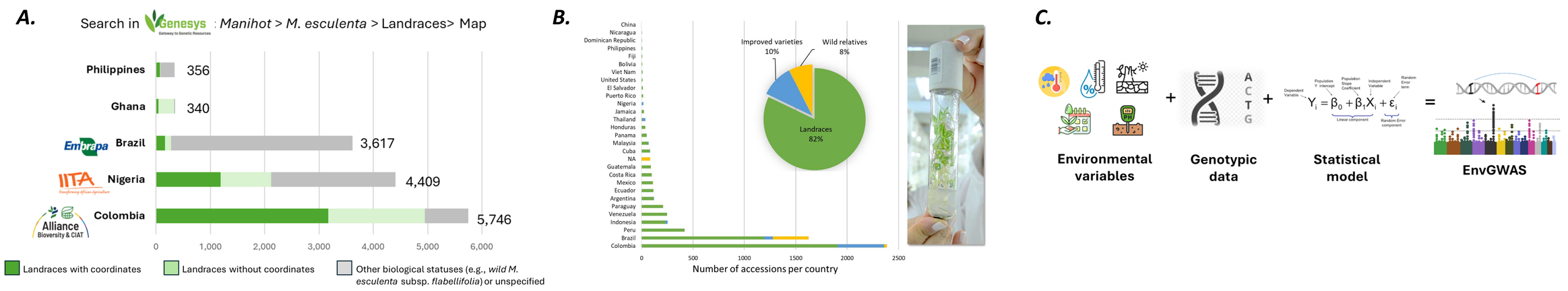

Figure 1. (**A**) The top five countries conserving cassava diversity ex situ, highlighting the proportion of landrace accessions with collection site coordinates. (**B**) The composition of the largest cassava collection (totaling 5,957 accessions) conserved at the Alliance Bioversity and CIAT in Colombia. (**C**) Key components for the EnvGWAS.

Methods and Results

The analysis incorporated 96 environmental variables (**Table 1**), 17,114 SNPs, and 38,170 SilicoDArT markers across 3,050 cassava landrace accessions. The **maximum** temperature reached **36.2°C**. Monthly **precipitation**, **ranged from 0 to 736 mm**. The **Aridity index** spanned from **0 to 6.1**, the relative humidity ranged between **33.9% and 96.5%,** the VPD varied from **0.16 to 2.4 kPa**, and the length of the growing period extended from **0 to 366 days**. Population structure for SNP and SilicoDArT data was assessed through ADMIXTURE and hierarchical clustering based on mrd or Jaccard distances (**Figure 2A**). A total of 14 ancestral populations from admixture were used to correct for population structure in the EnvGWAS analyses.

Table 1. Environmental variables extracted using collection site coordinates of 3,050 landraces accessions

Dataset	Datalabel	Unit	Timerange	Resolution used	Source	Number
Rainfall	prec_01 to prec_12	mm	average 1970-2000	1 km	worldclim 2.1	12
Temperature	tmax_01 to tmax_12; tmin_01 to tmin_12	Deg Celsius	average 1970-2000	1 km	worldclim 2.1	24
Bioclimatic variables	bio1- to bio19		average 1970-2000	1 km	worldclim 2.1	19
Aridity index (Ai)	Aridity01 to Aridity12		average 1970-2000	1 km	Global Aridity Index and Potential Evapotranspiration Database - Version 3 (Global-AI_PET_v3)	12
Relative Humidity	Rhumavjan5 to Rhumavdec5	%	1950-1980	9 km	ERA5	12
Vapor pressure deficit (VPD)	Vpdav01 to vpdav12	kPa	1958-1987	4 km	TerraClimate	12
Length of growing period	lgpaez4	Days	1981-2010	2 km	FAO GAEZ v4	1
pH	ph0-5cm1, ph5-15cm	pH		250m	ITC Soil grids 2	2
Sand	sand 0-5cm1, sand 5-15cm	% sand fraction of soil		250m	ITC Soil grids 2	2
					TOTAL	96

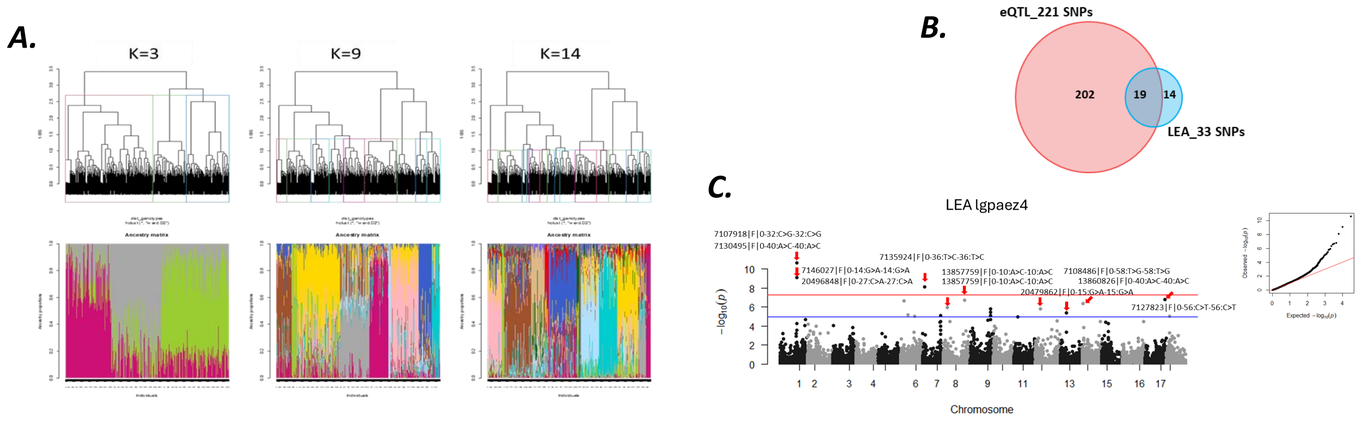

Figure 2. (**A**) Dendrograms and histograms of ancestry coefficients for 3, 9, and 14 populations based on SNP markers.(**B**) Venn diagram illustrating the common SNP markers identified by LFMM2 (blue, labeled as LEA_33 SNPs) and the linear mixed model analysis from the MatrixEQTL R package (pink, labeled as eQTL_221 SNPs).(**C**) Exemplary Manhattan plot and QQ plot for the length of the growing period variable.

Two models were employed for SNPs: the latent factor mixed model (lfmm2) from the LEA R package, and the linear mixed model analysis from the matrixEQTL R package, applied to all variables (**Figure 2B**). For SilicoDArT markers, the analysis was performed using the linear mixed model from the matrixEQTL R package. The study identified 55 candidate associations (19 SNPs and 36 SilicoDArTs) across multiple chromosomes (**Figure 2C**).

The distribution of environmental variables for each of the 55 significant markers was reviewed, and favorable alleles/genotypes for drought and heat tolerance were selected (**Figure 3A**). Using these 55 significant markers, we screened the 3,050 accessions to identify individuals harboring the maximum number of favorable alleles. We propose 52 accessions to be potentially heat and/or drought tolerant originate from sites with dry climates in Ecuador (23), Peru (19), Brazil (5), Costa Rica (2), Colombia (2), and Indonesia (1) (**Figure 3B**).

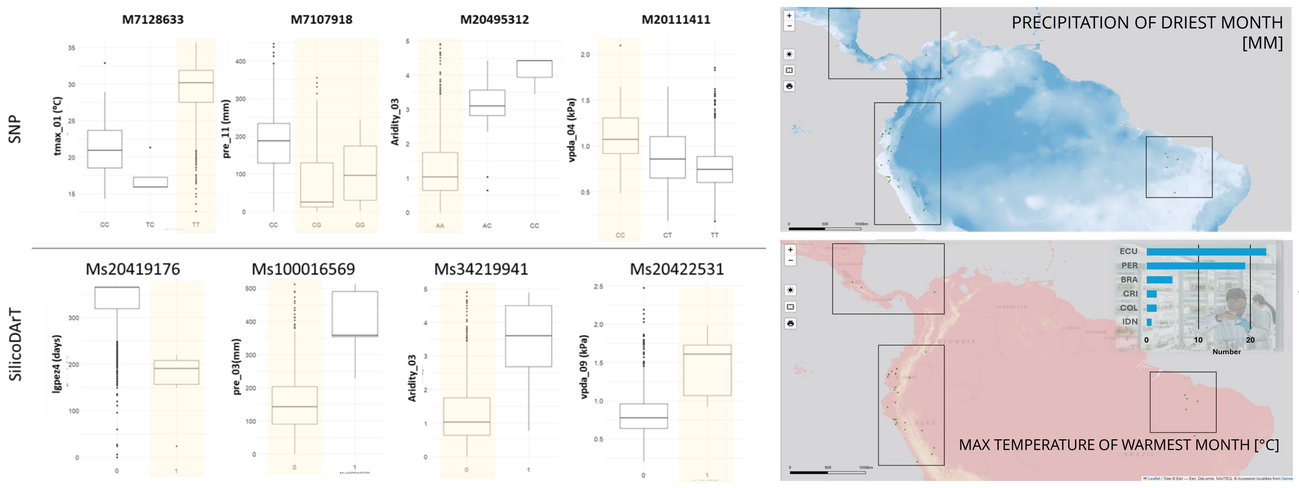

Figure 3. (**A**) Boxplots of the distribution of four environmental variables of significant markers highlighting (yellow box) the selected positive genotypes. (**B**) Maps display 52 accessions with probabilities of heat and/or drought tolerance based on significant variants identified through EnvGWAS.

Conclusions

- ✓ A list of genetic variants associated to drought and heat tolerance, as well as a list of accessions harboring most of genetic variants were facilitated to CIAT cassava breeding program.
- ✓ Our results are being compared with those from the International Institute of Tropical Agriculture (IITA), another CGIAR center dedicated to conserving African cassava diversity.
- ✓ This study provides best-bet candidate accessions for cassava breeding programs to develop climate-smart cassava varieties.

Acknoledgements: This work has been supported by Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), the CGIAR Genebank Initiative and the Allele Mining Initiative funded by the BMGF. The Allele Mining Initiative is lead by CIMMYT.