**Landscape genomics reveals genetic signals of environmental adaptation of African wild eggplant**

Emmanuel O. Omondi1, Chen-Yu Lin2, Shu-Mei Huang2, Cheng-An Liao4, Ya-Ping Lin2, Ricardo Oliva3, Maarten van Zonneveld1

1Genetic Resources and Seed Unit, World Vegetable Center, Headquarters, 60 Yi-Min Liao, Shanhua, Tainan 74151, Taiwan

2Biotechnology, World Vegetable Center, Headquarters, 60 Yi-Min Liao, Shanhua, Tainan 74151, Taiwan

3Plant Pathology, World Vegetable Center, Headquarters, 60 Yi-Min Liao, Shanhua, Tainan 74151, Taiwan

4Department of Horticulture, National Taiwan University, Taipei, Taiwan

**Correspondence email (****[maarten.vanzonneveld@worldveg.org](mailto:maarten.vanzonneveld@worldveg.org) ;** [**emmanuel.omondi@worldveg.org**](mailto:emmanuel.omondi@worldveg.org) **)**

**Abstract**

Crop wild relatives provide a valuable resource for improving crops. They possess desirable traits that confer resilience to various environmental stresses. To fully utilize crop wild relatives in breeding and conservation programs, it is important to understand the genetic basis of their adaptation. Landscape genomics associates environments with genomic variation and allows for examining the genetic basis of adaptation.

In this study, we applied landscape genomics to examine the differences in allele frequency of 15,416 Single Nucleotide Polymorphisms (SNPs) among 153 accessions of wild eggplant relatives from Africa, the principal hotspot of these wild relatives. Further, we explored the correlation between these variations and the bio-climatic and soil conditions at their collection sites.

Our results showed that the environment has a greater impact on the genetic variation in the eggplant wild relative populations compared to the geographical distances between collection sites while controlling for population structure. These findings indicate the relevance of the environment in shaping genetic variation in eggplant relatives over time. We detected 396 candidate SNPs associated with ten environmental factors by applying four genotype-environment association methods. Some of these SNPs signal genes involved in pathways that help with adaptation to environmental stresses such as drought, heat, cold, salinity, pests, and diseases. These candidate SNPs will be useful for marker-assisted improvement and characterizing the germplasm of this crop for developing climate-resilient eggplant varieties. The study provides a model for how we can apply landscape genomics to the wild relatives of other crops.

**Keywords**

Climate-resilient, Crop wild relatives, Environmental stress, Genotype-environment association, Marker-assisted improvement.

**Introduction**

Crop wild relatives possess traits of interest for breeding climate-resilient varieties because many are adapted to marginal environments (Kapazoglou et al., 2023). However, often it is not clear what specific adaptive traits they possess and to which abiotic stresses they are adapted (Rellstab et al., 2015), and linkage drag with undesirable traits makes it difficult to detect them (Huang et al., 2023; Chitwood-Brown et al., 2021). Landscape genomics is an emerging research discipline with a high potential to speed up the detection of valuable traits supporting breeding programs with information about a wide range of associations between specific genome locations and specific environmental factors. Landscape genomics integrates spatial statistics, population genomics, and landscape ecology to rapidly discover various adaptive markers associated with a wide range of environmental factors (Haupt & Schmid, 2022; Manel et al., 2010). It has been successfully applied to detect genes associated with the environmental adaptation of wild plants (Chang et al., 2022; Lei et al., 2019; Morente-López et al., 2018; Lasky et al., 2015).

Eggplants, including, brinjal eggplant (*Solanum melongena* L.) and African eggplants (*S. aethiopicum* L., *S. anguivi* Lam., and *S. macrocarpon* L.), are important vegetables globally and regionally, belonging to the Solanaceae family. Despite its significance in food production worldwide, eggplant has trailed in the development and use of genomic tools compared to other Solanaceae crops such as potato and tomato (Gramazio et al., 2018). This has changed over recent years due to the development of new genomic resources, including a high-quality *de novo* assembled eggplant genome (Barchi et al., 2021; Gramazio et al., 2019). These genomic resources allow us to start screening eggplant genebank accessions for genes associated with environmental adaptation. Crop wild relatives are especially interesting to screen because they possess large untapped genetic diversity and traits of environmental adaptation that disappeared from eggplant varieties during domestication and breeding.

The African wild eggplant populations are particularly interesting because sub-Saharan Africa is a hotspot of wild relatives of all domesticated eggplants including brinjal eggplant (Syfert et al., 2016; Aubriot et al., 2018). African eggplant wild relatives have shown significant morphological variations and thrive in a myriad of ecological habitats spanning from the equatorial savanna to almost barren desert landscapes (Weese & Bohs, 2010). These *Solanum* species are poorly studied for breeding purposes and they are underrepresented in seed banks (Syfert et al., 2016). However, it is highly probable that each population has developed specific adaptations to suit their respective local environmental conditions, resulting in a multitude of variations.

So far, landscape genomics has not been widely applied to detect genes associated with adaptive traits in crop wild relatives. Traditionally, landscape genomics focuses on single species (Richardson et al., 2016). However, for crop wild relatives, a limited number of records are often available for individual species, and breeders are often interested in screening multiple species in the same crop gene pool for traits of interest (Engels & Thormann, 2020). Therefore, it is common that crop genomic studies cover gene pools with multiple species (Lin et al., 2022; Tripodi et al., 2021; Barchi et al., 2021). This means that by associating genetic signals of multiple crop wild relatives across the environmental gradient of a landscape, we can gain a broad insight into eco-evolutionary patterns across crop gene pools and identify various options for breeding.

In this study, we apply landscape genomics to screen African eggplant wild relatives for SNPs associated with environmental adaptation. Our objectives were to: *i*) evaluate the population structure of eggplant wild relatives from diverse environments, taxa, and geographies in West and Eastern Africa; *ii*) estimate the contribution of environmental, population structure, and geographic factors in shaping the genomic variation across eggplant wild relatives genepools; *iii*) identify candidate SNPs and their association with the environmental factors. The genotype-environment association was also applied to predict the adaptive landscape; *iv*) investigate the potential role of genes associated with candidate SNPs in enabling local adaptation.

**Materials and methods**

**Plant material**

We genotyped 153 accessions of 17 eggplant species including 15 wild species and two cultivated species (*S. macrocarpon* and *S. aethiopicum*) collected from wild or feral populations. Royal Botanical Garden, Kew provided the taxonomic classification, and we further checked in World Flora Online (WFO, 2023). The collections comprised several species representing collections from different West and East African countries collected during the Global Crop Wild Relatives project (<http://www.cwrdiversity.org/>; (Müller et al., 2021; Dempewolf et al., 2014) and other initiatives (**Figure 1**). The accession collection points represent different Köppen climate zones (**Figure 5; Table S8**) (Beck et al., 2018). All accessions are available at the World Vegetable Center (WorldVeg) genebank (<https://genebank.worldveg.org/>).

**Genotyping**

According to the manufacturer's instructions, we isolated the genomic DNA from fresh leaves of five seedlings per accession using the FavorPrep Plant Genomic DNA Extraction Mini Kit (FAVORGEN). We then constructed the sequencing library following the approach of Elshire *et al*. (2011). Genomic DNA was quantified by Qubit and normalized to 100ng in 96-well plates. We digested the DNA samples using the restriction enzyme *Ape*KI and ligated them with two adapters for sequencing, followed by the polymerase chain reaction to amplify the target DNA fragments to complete the sequencing library preparation. A service provider did sequencing with the Illumina HiseqX platform in a pair-end 150bp run.

For the SNP calling, we followed mainly the manual of Stacks software (Catchen et al., 2013). In short, we filtered the raw reads by quality and demultiplexed using the process radtags program. We then mapped the retained reads to the eggplant reference genome (Eggplant\_V4.1.fa) ([Barchi et al., 2021](https://doi.org/10.1111/tpj.15313)) using the Burrows-Wheeler Aligner (BWA) version 0.7.17 (Li & Durbin, 2009). We sorted and indexed the reads using Samtools version 1.15.1 (Li et al., 2009), after which we performed the variant calling using the gstacks and population programs in Stacks software. We further filtered the SNPs and the accessions with less than 20% missing data and a Minor Allele Frequency (MAF) > 0.05, giving the final high-quality SNP dataset comprising 15,146 SNPs used for the analysis.

**Environmental data**

We selected climate and soil data from two open-source databases for our models (**Table S2**). We downloaded the grids for 19 bioclimatic variables, solar radiation, wind speed, and vapor pressure derived from WorlClim 2.1 (Fick & Hijmans, 2017) at a resolution of 2.5 minutes. The 19 bioclimatic variables were each downloaded as annual data averages between 1970 and 2000. We averaged the monthly solar radiation, wind, and vapor pressure rasters to obtain annual value rasters from this period. Soil variables included nitrogen, soil organic carbon, organic carbon density, organic carbon stock, cation exchange capacity, pH, clay sand, and silt content. We downloaded the soil data from the SoilGrids database released in 2016 (<https://soilgrids.org/>) through ISRIC—WDC Soils (Hengl et al., 2017) at 250-meter resolution and at a depth of 15-30 cm, approximately the depth at which the eggplant roots can grow. We aggregated the resolution of the soil dataset to match that of the climate data, ensuring they are consistent in both resolution and extent. We averaged the aggregated soil values using the *resample* and *extent* functions of the *raster* package in R (Hijmans, 2023).

For each accession, we extracted the data of the environmental variables with the *extract* function of the R *raster* package (Hijman, 2023) using the GIS coordinates at sampling points (**Table S1**) to obtain a full data set of all the climate and soil variables (**Table S3**). For the modeling, we selected the environmental variables based on Variance Inﬂation Factors (VIFs). VIFs less than five are considered low correlation (James et al., 2017).

**Population structure and differentiation analysis**

We used the program STRUCTURE ver.2.3.5 (Pritchard *et al*., 2000b) to investigate the population structure. We assigned the 17 species as our populations in the STRUCTURE analysis. These species represented seven eggplant clades (the Melongena, Anguivi, Arundo, Coagulans, Giganteum, Acanthophora, and Aculeastrum clades) (Aubriot et al., 2016; Syfert et al., 2016; <http://www.solanaceaesource.org>). Considering this phylogenetic structure, we ran sub-populations varying from K = 2 to K = 10, each with ten independent runs and a burn-in of 10,000 iterations for each run. To determine the optimal K, we generated Delta (𝚫) K for each K with a web-based program, STRUCTURE HARVESTER ver0.6.94 (Earl & Vonholdt, 2012). We then aligned and plotted Delta K-s against K-s with the CLUMPAK server (Kopelman et al., 2015). The admixture coefficients (**Table S4**) for the optimal K value for each individual were plotted as pie charts using the *map* function in the *maps* R package (Becker et al., 2022) onto the map showing the sampling site. In addition to STRUCTURE analysis, a phylogenetic tree was constructed based on the SNP markers using the Unweighted Pair-Group Method with Arithmetic Averaging (UPGMA) in TASSEL v5.2.89 (Bradbury et al., 2007) and plotted using FigTree v1.4.4 (Rambaut, 2010).

**Partitioning of the genomic variation and identification of candidate SNPs**

We used four genome scan methods to make a complimentary selection of candidate SNPs among the different methods;

First, we used simple Redundancy Analysis (RDA) to associate the obtained SNPs with the selected environmental factors without controlling for population structure and geographical distances. RDA is a multivariate method for assessing a linear relationship between two or more factors (Legendre & Legendre, 2012). RDA was performed with the *rda* function of the R package *vegan* (Oksanen et al., 2019). We carried out 5,000 permutations to test the significance of explanatory variables with the R function *anova.cca*.

Second, we used partial RDA, which allows the partitioning of genomic variation into components explained by different factors. To partition the genomic variation into these components, we conducted a full model RDA with the selected environmental factors, spatial autocorrelation, and population structure, and then did a partial RDA conditioned on covariates to estimate the proportion of SNP variation explained by the factors that were included in the model. We did the partial RDA with the *rda* function and the number of permutations as the simple RDA explained above. To account for the effect of spatial autocorrelation on SNP variation in the partial RDA analysis, we applied distance-based Moran's eigenvector maps (dbMEMs) in RDA (Legendre & Legendre, 2012; Dray et al., 2006). This involved first building a neighborhood connection network of 153 collection points. With this network, we constructed a spatial weighting matrix of inverse geographical distances (km-1) following the method by Forester et al. (2018). The spatial weighting matrix was then decomposed to generate dbMEMs. Subsequently, we performed forward selection with the forward*.sel* function (Dray et al., 2022) to identify dbMEMs that associate signiﬁcantly with spatial genetic structure. We then applied the selected dbMEMs in RDA to capture comprehensive spatial autocorrelation (**Table S5**). To account for the effect of population structure on SNP variation in the partial RDA analysis, we used the ancestry coefﬁcients estimated by the STRUCTURE program with the optimal K (K = 8) as covariates. In RDA, SNP outliers were defined as the SNPs having loadings along the first three RDA axis +/- 3 SDs from the mean for each axis following Capblancq et al. (2018).

In the preliminary analysis, we noticed that population structure clustering was largely related to eco-geographical habitats. To separate the connections of population structure to environment and geography, we calculated the proportion of population structure attributed to environmental factors and spatial autocorrelation. We substituted the SNPs with the ancestry coefficient from the STRUCTURE analysis as our new response variable in the RDA models. Our new response variables in the RDA models were the ancestry coefficients from the STRUCTURE analysis with the optimal K.

Third, we performed Latent Factor Mixed Model (LFFM), a univariate method to associate the obtained SNPs with the selected environmental factors. LFFM allows control for population structure using the R package *lffm* (Caye et al., 2019). In the LFMM analysis, we controlled the population structure using the optimal K, which we initially determined using the STRUCTURE program and q-values computed. We considered SNPs with a False Discovery Rate (FDR) < 0.05 to be candidate SNPs.

Fourth, we used PCAdapt as an outlier differentiation method. The R package *PCAdapt* uses a PCA-based approach to simultaneously infer population structure and identify outlier loci related to this structure. In contrast to the three previous methods, it does not return associations between obtained SNPs and the selected environmental factors (Luu et al., 2016). With this different approach, we expected to capture other candidate SNPs not yet captured by the three gene-environment association methods explained above. We adjusted the *p* values using the Bonferroni method in the *p.adjust* function of the R *stats* package (R Development core team). After that, we applied an FDR <= 0.05 as the signiﬁcance level for detecting the outlier loci.

**Prediction of adaptive landscapes with a selected set of candidate SNP markers**

After identifying the candidate SNP markers and their associated environmental factors from all four methods mentioned, it is possible to predict the level of candidate SNP markers in a specific environment. The model shows the geographic patterns in environmental adaptation on a grid map of a so-called adaptive landscape (Capblancq & Forester, 2021). We carried out a simple RDA to model the adaptive landscape with the set of candidate SNPs - from the four methods explained above - and a set of environmental variables most strongly correlated with the putative adaptive variation. In this case, we used the candidate SNPs as the multivariate response in the simple RDA using the selected environmental variables as the explanatory variables. After that, we calculated an adaptive index for each geographic grid cell of the adaptive landscape based on the gene-environment associations following the procedure outlined by Steane et al. (2014). The adaptive index provides an estimate of the adaptive similarity or difference of all the grid cells in the landscape as a function of the environmental predictor values of each grid cell (Capblancq & Forester, 2021). We geographically mapped the indices for RDA axes 1 and 2 using the R package *ggplot2* (Wickham, 2016). Visualizing the adaptive landscape enabled us to observe the geographic distribution of the adaptive alleles across the population ranges of the crop wild relatives involved. A higher positive or negative adaptive index score is associated with changes in allele frequency of candidate SNPs across environmental gradients. The modeled landscape was limited to the geographic areas of 153 populations of eggplant wild relatives using the *ConvexHull* function of the R package *dismo* (Hijman et al., 2022). We extended the convex hull with 10% of the longest geographic distance using the *gBuffer* function.

**Gene annotation**

We identified genes linked to the candidate SNPs using the Sol Genomics Network data file transfer protocol (FTP) database for the eggplant genome consortium version 4.1 ([Barchi et al.,2021](https://doi.org/10.1111/tpj.15313))(https://www.solgenomics.net/ftp/genomes/Solanum\_melongena\_V4\_Pangenome/Annotation\_V4/). The candidate gene search and Gene Ontology (GO) terms were assigned using the *Arabidopsis* information resource (tair) (<https://www.arabidopsis.org/>) and Uniprot (https://www.uniprot.org/) databases. We characterized the genes and their functions, particularly those associated with abiotic stress and relevant to the environmental adaptation of the eggplant wild relatives.

**Results**

**Population structure**

The STRUCTURE harvester calculated an optimal number of eight groups (K =8) (**Figure S1)**. The ancestry coefficients in the clustering with K=8 were used in the genome scan analysis of RDA and LFFM. From K=2 to K=10, the cultivated *S. macrocarpon* and its wild ancestor *S. dasyphyllum and S. dasyanthum* clearly separated from the other 14 species **(Figure S2)**. This group belongs to the Anguivi clade, and the accessions are mainly from the West African eggplant populations.

The UPGMA phylogenetic tree base on the SNPs showed five genetic groups (**A**). While the groups represent seven *Solanum* clades, they largely reflect the geographic regions where the populations were sampled (**Figure 2B**).

**Genomic Environmental Association analysis**

**Simple RDA**

A simple RDA showed genetic differentiation within countries of origin on the first axis (Figure 3A). The second axis showed genetic differentiation among the populations with accessions from Nigeria, Ghana, and Uganda, forming a central group between accessions from Kenya and Sudan. These results suggest a large environmental effect on genetic differentiation. This observation is also consistent with the STRUCTURE analysis results that showed clustering largely due to environmental characteristics. The first RDA axis strongly correlated with the precipitation of the coldest quarters (r = -0.83: **Table S10**). Other environmental variables significantly correlated to the first axis included soil water pH (r = 0.69) and solar radiation (r = 0.60). The second RDA axis was highly correlated with the mean temperature of the wettest quarter (r = -0.79), soil organic carbon density (r = 0.60), soil nitrogen (r = 0.60), and precipitation of the warmest quarter (r = 0.55). The simple RDA identified one hundred and thirty candidate SNPs on the first three RDA axes that cumulatively explained 54% of the SNP variation. Fifty-one candidate SNPs were detected on the first axis, while forty-three and thirty-six candidate SNPs were detected on the second and third axes. Most of the SNPs were associated with solar radiation, mean temperature of the wettest quarter, and soil nitrogen (30, 28, and 24, respectively; **Figure 3B**; **Table S6**).

**Partial Redundancy analysis models and the genomic variation partitioning**

Conditioning the RDA on population structure and geographical distance significantly reduced the environmental variables' effects compared to the simple RDA (**Figure 4A; Table S10**), indicating a high correlation between the environmental factors and other factors (geographic distances and population structure). The first RDA axis was strongly correlated with precipitation of the coldest quarter (r = -0.30), nitrogen (r = -0.29), and mean temperature of the wettest quarter (r = 0.23), while the second axis was also strongly correlated to nitrogen (r = -0.29) and solar radiation (r = 0.23) (**Figure 4A; Table S10**). One hundred thirty-two candidate SNPs were identified on the first three RDA axes (**Figure 4B; Table S6**). Most of these SNPs were associated with soil nitrogen (74) content and twenty-four with solar radiation. We did not observe a clear grouping pattern of the accessions from different countries in the RDA.

The partitioning in genetic variance by environmental, geographic distance, and population structure explained 14% of the SNP variation (**Table 1**). The environmental variables accounted for 8% of the total variation, while geographical distances and the population structure explained 3% and 2% of the SNP variation, respectively.

Since genetic clusters largely corresponded to geographical areas, we partitioned the population structure to determine the proportions that environmental and spatial factors can explain. Environmental factors and spatial autocorrelations significantly explained 19% of the population structure. The environmental factors alone accounted for 11% of the population structure, while spatial autocorrelation only accounted for 5%. These results suggested a more significant effect of environmental factors than geographical distances on the population differentiation of the wild eggplants.

**Candidate SNPs**

In summary, we detected 396 candidate SNPs using the four methods (Figure 5; Table S10). These SNPs were distributed across all twelve chromosomes. Most SNPs were associated with soil nitrogen content, solar radiation, mean temperature of the wettest quarter, and precipitation of the driest month. The simple RDA, partial RDA, LFMM, and PCAdapt identified 130, 132, 51, and 114 candidate SNPs, respectively. The candidate SNPs identified by the four methods had minimal overlap between them. In total, 29 SNPs were identified by two methods. No SNPs were identified by three or more methods. Simple RDA and partial RDA identified 14 common SNPs. Partial RDA and LFMM identified ten common SNPs. LFMM and simple RDA identified two common SNPs, and simple RDA and PCAdapt identified three. A Manhattan plot showing the SNP positions for each of the four methods is shown in **Figures S3; S4**.

Approximately fifty-seven of the total three hundred and ninety-six candidate SNPs (**Table S7**) were directly associated with candidate genes that contribute to adaptation to different abiotic stressors such as heat, cold, drought, and salinity in the Uniprot and TAIR databases. For example, SNP Chr7:81532866 (**Table 2**), located in chromosome 7, is associated with nitrogen content in the soil and is linked to a protein FIP1 (characterized in *Arabidopsis thaliana*). This protein is known to be involved in response to water deprivation and salt stress. Temperature signals also stimulate it (Luhua et al., 2013). Seventy-four of the detected candidate SNPs were mapped within genes for proteins of unknown functions.

We illustrated the allele frequency distribution of the SNP Chr7:81532866 (**Figure 6**). We observed a clear difference in the allele frequency distribution along the environmental gradient associated with different climates and soil nitrogen content. The allele distribution patterns are distinct between Sudan's dry semi-arid, hot regions, where the minor alleles are observed, and the tropical monsoon and dry winter savannas of Nigeria and Ghana, where the major alleles dominate.

**Adaptive landscape**

The modeled environmental space enriched with candidate SNPs is illustrated in **Figure 7.** The allele frequencies along the first RDA axis (69.2%) were primarily associated with solar radiation, precipitation of the warmest quarter, and soil variables (soil water pH, clay and silt content, and organic carbon density). In comparison, the second RDA axis (19.4%) was mainly associated with the mean temperature of the wettest quarter, organic carbon density, and soil nitrogen content (**Figure 7A**). When our adaptive landscape model was projected geographically, the first RDA axis showed a contrast between the dry semi-arid hot climates experiencing higher radiations in East Africa and the tropical and temperate climates of West Africa (**Figure 7B**). Our landscape models also showed that the adaptive scores in the second RDA axis contrast areas with high and low values for the mean temperature of the wettest quarter and soil nitrogen content (**Figure 7B**).

The species average and range of the adaptive scores are shown in **Table S11**. On the first RDA Axis, *Solanum coagulans* and *S. incanum s*howed, on average, the highest positive adaptive scores, which are mainly related to an environment with high solar radiation and high soil water pH. *Solanum anomalum* and *S. macrocarpon* showed the most negative adaptive scores on the first RDA axis, mainly related to an environment with low solar radiation and soil water pH. Some populations of *S. cerasiferum* and *S. dasyphyllum* also showed low adaptive scores on the first RDA axis, while other populations of these species returned moderate scores. These two species also show a higher range for the adaptive scores in the first and second RDA axes due to the environmental differences of their sampling sites within each species (**Table S9**). On the second RDA axis, *S. coagulans* and *S. incanum* also showed the highest positive adaptive scores compared to other species, mainly related to an environment with low soil nitrogen content and high temperature in the wettest quarter. *Solanum aculeatissimum* and *S. mauense* showed the most negative adaptive scores on the second RDA axis, mainly related to an environment with high soil nitrogen content and low temperatures in the wettest quarter. On the second RDA axis, some populations of *S. cerasiferum* also had high adaptive scores; on average, populations of this species returned moderate scores.

**Discussions**

**The genetic variation across eggplant crop wild relatives significantly corresponds to environmental adaptation.**

Our findings suggest that environment accounts for a higher proportion of explainable genetic variation of the wild eggplant relatives in our study than population structure and spatial autocorrelation. These findings provide us with insights into to which extent populations of wild eggplant relatives have been shaped by environmental adaptation. They highlight the environment's role in speciation in the eggplant genepools, which is in line with previous reports on eggplant species divergence (Weese & Boh, 2010).

The RDA analysis showed that climate and soil are important environmental drivers of genomic variations in our materials. Including a range of environmental data in landscape genomics studies will help effectively evaluate the complex factors contributing to local adaptation in nature (Dauphin et al., 2023). So far, few landscape genomic studies have integrated soil factors such as pH and nitrogen in the analysis. Even though the climate may influence soil development (Joswig et al., 2022), our findings highlight the importance of including soil factors in landscape genomics as drivers of environmental adaptation. This is also in line with other studies, for example, a study on the adaptation of two grasshopper species in the Australian Alps revealed significant GEAs with soil pH, among other factors (Yadav et al., 2021). Arabidopsis demes were also found to be locally adapted in their native habitat to soils with moderately high carbonate (Terés et al., 2019).

Our study observed a large environmental effect compared to the geographic effect in line with other studies for the plant species *Boechera stricta* genotypes from populations experiencing different climatic conditions (Lee & Mitchell-Olds, 2011; Lin et al., 2021). In other studies, geographical distance and environmental factors explained comparable proportions of the genetic variation (Gibson & Moyle, 2020; Lasky et al., 2015; Lasky et al., 2012). Thus, the proportional effects of environment and geography depend heavily on the species, its environment, and the species’ history in that environment.

Although the proportion of the explainable SNP variation due to environmental factors was higher than geographic or population structure, it still accounted for only 8%. This is in line with other landscape studies that report a low percentage of variation explained by the environment (Dauphin et al., 2023). This is possibly due to the lack of comprehensive environmental data or other evolutionary forces, for instance, pests and diseases unrelated to the factors used in our study. Also, while our environmental data provides information at a larger geographic scale, they may not necessarily capture all local environmental heterogeneity. Another reason for low explainable SNP variation may arise because RDA and LFMM can only model linear associations between the environment/space and SNPs and, therefore, will fail to capture non-linear associations that might exist (Borcard et al., 2011). Nevertheless, RDA models remain effective because they can effectively account for covariation among environmental variables and genetic markers, as is often the case in nature (Capblancq *et al*., 2021).

Our findings show that the eggplant's West and East African populations exhibited distinctive genetic responses due to climatic and soil factors. The first adaptive component was associated with solar radiation and soil water pH. This component contrasts the West African sampling areas from the North and East African sampling areas, mainly along solar radiation, soil water pH, and temperature gradients. The positive adaptive score along the RDA1 for clay, silt content, and organic carbon density in the adaptive space further suggests that abiotic soil properties are an important selective agent. The climatic factors, however, showed negative scores on the adaptive landscape. This indicates that the covariation between soil and climate is an oversimplification, so it would be important to investigate the effect of the two in driving selection. Identifying species with interesting traits is paramount to effectively using CWR diversity. The description of the adaptive landscape enables the identification of putatively adapted genotypes returning high adaptive scores. In this study, *S. incanum* and *S. coagulans* showed the highest adaptive score of 4.7 and 4.2, respectively. The high adaptive score for these species affirms records that they are known to grow in desert conditions and, especially *S. incanum*, is considered a powerful source of phenolics and tolerant to abiotic stress such as drought (Plazas et al., 2022: Gramazio et al. 2017; Knapp et al., 2013; Meyer et al., 2012).

We observed admixture for some accessions of *S. incanum, S. campylacanthum, S. nigriviolaceum, S. coagulans, and S. dasyphyllum* species from our structure analysis, suggesting gene flow and interspecific hybridization occur regularly among eggplant wild relatives. The interfertile nature of the eggplant species can explain the admixtures and clustering of the species from different *Solanum* clades. Interspecific hybridization between eggplant species has been shown mostly in crossing experiments (Bukenya & Carasco, 1995; Plazas et al., 2016) and between domesticated eggplants and their wild ancestors (Meyer et al., 2012). However, we have not found reports so far on the interspecific hybridization among eggplant wild relatives in their natural environment. This provides evolutionary insights, has implications for *in situ* conservation, and requires further investigation to understand these natural patterns of interspecific hybridization. These findings also confirm the relevance of carrying out landscape genomics for crop wild relatives at the genepool level rather than individual species as the genepools of eggplant and many other crops, such as those of pumpkin and amaranth have significant levels of interspecific fertility (Sanjur et al., 2002; Ya-ping et al., 2021).

One limitation of our study was the seemingly biased sampling of species in the different regions involved in our study. This is mainly due to the rare record of some of the species. Furthermore, as much as crop wild relatives are known to thrive in diverse marginal environments, some species may be limited to particular environments (Renzi et al., 2022). Expanding this study with new collections will provide further information about environmental adaptation. However, our current study provides significant findings to inform breeders and conservation managers.

Because the adaptive landscape is associated with environmental factors, conservation managers may also apply our findings in effectively managing populations experiencing different ecological conditions and facilitate future studies on the response of eggplant populations to future environments. This tool can be included in conservation assessments to target areas for *in situ* conservation and germplasm collecting.

**Landscape genomics detected candidate SNP markers for both climate and soil factors.**

Our outlier detection approaches identified 396 candidate SNP markers - less than one percent of the total number of SNPs of the initial selection. Percentages similar to ours have also been observed in other studies (Chang et al., 2022: Mdladla et al., 2018). We attribute this to several reasons. Likely, more loci are under selection, but the stringency applied in our analysis did not allow their detection. The control for population structure in GEA methods can sometimes be overly conservative (Forester et al., 2018). Therefore, we applied multiple methods in our study to capture more candidate SNPs. We attribute their complementarity to the different assumptions, strengths, and limitations of the four methods. At the same time, the conservative selection reduces the number of false positives, and a large proportion of the final set of candidate SNPs could be associated with candidate adaptive genes that have been reported. Secondly, adaptation occurs mostly due to minor and linked allele frequency modifications across multiple genetic loci that show weak selection and may not be detected through the genome scan (Stetter et al., 2018). Finally, the GBS method that we applied in generating our SNPs only allows the examination of low coverage of the genomes, thereby limiting the total number of SNPs used in our analysis and the ability to detect a considerable number of adaptive candidate genes. Future whole-genome studies will allow the detection of more candidate SNP markers.

Among the most notable adaptive candidate genes, we detected a candidate protein *FIP1,* a regulator of seed dormancy and germinationthrough the abscisic acid (ABA) pathway (Li et al., 2023). Seed dormancy is an important adaptive trait to prevent germination from occurring at an inappropriate time. This protein, correlated to soil nitrogen content, shows a directional change in allele frequency to soil nitrogen content. The minor allele frequency was observed predominantly in the Sudan population experiencing hot and dry climates where nitrogen availability is limited (Peri et al., 2019). Under nitrogen deficiency, roots show root-related differential expression of proteins contributing to enhanced root growth (Qin et al., 2019). Other candidate genes associated with soil nitrogen contents and genes regulating root development detected in our study include protein *AKT1*, an important potassium ion channel in *Arabidopsis* roots that is involved in potassium ion intake and improves plant response to water and salt stress (Ren et al., 2013; Nieves-Cordones et al., 2012); and protein *XLG3* that plays a role in the root morphogenesis by regulation of cell proliferation, root waving and root skewing (Pandey et al., 2008). These findings offer valuable insights for investigating eggplant root response to nitrogen deficiency and the development of cultivars with high nitrogen use efficiency through genetic improvements.

The multiple associations of soil nitrogen content with candidate SNPs from gene functions related to root nodulation and bacterial and fungal infections could also suggest that soil microbiomes have an important role in the environmental adaptation of eggplant crop wild relatives. Recent research is starting to unravel the role of these microbiomes in plant adaptation to environmental stress (AL-surhanee et al., 2021; Pasbani et al., 2020). The inclusion of microbiome diversity in landscape genomics could provide further insights into the environmental adaptation of plant species.

Our study also detected candidate genes, *CPN60B4, ATX2*, *CDF2,* and *GSH1,* regulating the transition from vegetative to reproductive and general flower development. From the adaptation perspective, suppressing or accelerating the transition from vegetative to reproductive development in plants shows clear fitness consequences in harsh environments (Shafiq et al., 2014; Li et al., 2010). Flowering too early or too late can increase the chances of floral damage and the risk of incomplete seed development due to adverse weather conditions (Inouye, 2008).

Other candidate genes that are promising candidates for environmental adaptation include Phosphate transporter PHO1 homolog 10 (*PHO1\_H10*) induced by dehydration, salt, and cold treatment (Wang et al., 2008); *ANN3* up-regulated by cold and dehydration stresses while down-regulated by heat shock stress(Cantero et al., 2006); *SUD1* expression in drought conditions (Lü et al., 2012); *AKT1* expressed in response to water deprivation and salt stress (Nieves-Cordones et al., 2012); *GSH1* and *RH38* expressed in response to heat and cold stress (Gong et al., 2002).

While our study detected numerous candidate genes conferring environmental adaptation, we imagine these genes could be linked to other genes in gene clusters that drive traits involved in environmental adaptation. Therefore, these genes can be useful in gene co-expression network analysis to map other genes that could contribute to adaptation traits.

This is the first study identifying the candidate SNPs to detect adaptive genes in eggplant wild relatives. Our set of candidate SNP markers provides a new genomic tool for eggplant breeding while our study provides an example of how landscape genomics can be applied to crop wild relatives. Our analysis thus provides a toolbox for breeders and researchers to establish new phenotyping experiments to test specific relations between genes and the environment. Associating the candidate SNP markers to environmental factors is a primary step in uncovering the adaptive process. However, these correlations do not necessarily confirm the occurrence of environmental adaptation in nature. Therefore, the next step is to validate these genes with functional analysis in transcriptome analysis of gene expression.

**Conclusions**

Overall, this study provides valuable insights into the environmental adaptation of eggplant CWR with clear applications for breeders and conservation managers. We showed that environmental selection is a critical factor in the genetic variation within a crop's wild relatives that are both widespread and diverse for both climate and soil factors.

**Acknowledgments**

We thank the WorldVeg biotechnology team, especially Shu-hui Chu, for DNA extraction and library preparations; and the WorldVeg genebank team for seeds and nursery preparations. We thank Royal Botanical Garden’s, Kew, especially Chris Cockel, for sharing seeds of the 153 accessions of eggplant wild relatives. Funding for WorldVeg’s general research activities is provided by core donors: Taiwan, UK aid from the UK government, the United States Agency for International Development (USAID), the Australian Centre for International Agricultural Research (ACIAR), the Federal Ministry for Economic Cooperation and Development of Germany, Thailand, Philippines, Korea, and Japan. The Taiwan Africa Vegetable Initiative has provided specific funding for this research. M.v.Z. received financial support from BMZ commissioned by the Deutsche Gesellschaft für Internationale Zusamme through the Fund International Agricultural Research, grant number: 81275070.

**References**

AL-surhanee, A. A., Soliman, M. H., & Ouf, S. A. (2021). The role of soil microbes in the plant adaptation to stresses: Current scenario and future perspective. *Frontiers in Plant-Soil Interaction*, 237-258. https://doi.org/10.1016/B978-0-323-90943-3.00006-7

Aubriot, X., Singh, P., & Knapp, S. (2016). Tropical Asian species show that the Old World clade of ‘spiny solanums’ (Solanum subgenus Leptostemonum pro parte: Solanaceae) is not monophyletic. *Botanical Journal of the Linnean Society*, *181*(2), 199-223. https://doi.org/10.1111/boj.12412

Aubriot, X., Knapp, S., Syfert, M. M., Poczai, P., & Buerki, S. (2018). Shedding new light on the origin and spread of the brinjal eggplant (Solanum melongena L.) and its wild relatives. *American Journal of Botany*, *105*(7), 1175-1187. https://doi.org/10.1002/ajb2.1133

Barchi, L., Rabanus-Wallace, M.T., Prohens, J., et al. (2021) Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding. *The Plant Journal*, 107, 579–596. https://doi.org/10.1111/tpj.15313

Beck, H., Zimmermann, N., McVicar, T., *et al.* (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data,* **5**, 180214. <https://doi.org/10.1038/sdata.2018.214>

Becker OScbRA, Minka ARWRvbRBEbTP, Deckmyn. A (2022). \_maps: Draw Geographical Maps\_. R package version 3.4.1. [https://CRAN.R-project.org/package=maps](https://cran.r-project.org/package=maps).

Borcard, D., Gillet, F., Legendre, P. (2011). Numerical Ecology with R. Springer International Publishing AG. https://doi.org/10.1007/978-3-319-71404-2

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633-2635. https://doi.org/10.1093/bioinformatics/btm308

Bukenya, Z. R., & Carasco, J. F. (1995). Crossability and cytological studies in Solanum macrocarpon and Solanum linnaeanum (Solanaceae). Euphytica, 86, 5-1. <https://doi.org/10.1007/BF00035933>.

Cantero, A., Barthakur, S., Bushart, T., Chou, S., Morgan, R., Fernandez, M., Clark, G., & Roux, S. (2006). Expression profiling of the Arabidopsis annexin gene family during germination, de-etiolation, and abiotic stress. *Plant Physiology and Biochemistry*, *44*(1), 13-24. https://doi.org/10.1016/j.plaphy.2006.02.002

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22(11):3124-40. https://doi: 10.1111/mec.12354.

Caye, K., Jumentier, B., Lepeule, J., François, O. (2019). LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. *Molecular Biology and Evolution*, 36 (4) 852–860,<https://doi.org/10.1093/molbev/msz008>

Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods in Ecology and Evolution*, *12*(12), 2298-2309. https://doi.org/10.1111/2041-210X.13722

Capblancq, T, Luu, K, Blum, MGB, Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resour*ces, 18: 1223– 1233. <https://doi.org/10.1111/1755-0998.12906>

Chang, C.W., Fridman, E., Mascher, M. *et al.* (2022). Physical geography, isolation by distance, and environmental variables shape the genomic variation of wild barley (*Hordeum vulgare* L. ssp. *spontaneum*) in the Southern Levant. *Heredity,* **128**, 107–119. https://doi.org/10.1038/s41437-021-00494-x

Chitwood-Brown, J., Vallad, G.E., Lee, T.G., Hutton, S.F. (2021). Characterization and elimination of linkage-drag associated with Fusarium wilt race 3 resistance genes. *Theoretical and Applied Genetics*, 134(7):2129-2140. https://doi:10.1007/s00122-021-03810-5

Daskalova, S. M., Pah, A. R., Baluch, D. P., & Lopez, L. C. (2009). The Arabidopsis thaliana putative sialyltransferase resides in the Golgi apparatus but lacks the ability to transfer sialic acid. *Plant Biology*, *11*(3), 284-299. https://doi.org/10.1111/j.1438-8677.2008.00138.x

Dauphin, B., Rellstab, C., Wüest, R. O., Karger, D. N., Holderegger, R., Gugerli, F., & Manel, S. (2023). Re-thinking the environment in landscape genomics. *Trends in Ecology & Evolution*, *38*(3), 261-274. https://doi.org/10.1016/j.tree.2022.10.010

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C. and Guarino, L. (2017), Past and Future Use of Wild Relatives in Crop Breeding. *Crop Science*, 57: 1070-1082.<https://doi.org/10.2135/cropsci2016.10.0885>

Dempewolf, H., Eastwood, R.J., Guarino, L., Khoury, C.K., Müller, J.V. and Toll. J. (2014) Adapting Agriculture to Climate Change: A Global Initiative to Collect, Conserve, and Use Crop Wild Relatives. *Agroecology and Sustainable Food Systems*, 38:4, 369-377. https://doi: [10.1080/21683565.2013.870629](https://doi.org/10.1080/21683565.2013.870629)

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modeling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*(3-4), 483-493. https://doi.org/10.1016/j.ecolmodel.2006.02.015

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., Wagner, H.H. (2022). adespatial: multivariate multiscale spatial analysis. R package version 0.3-7. https://CRAN.R-project.org/package=adespatial

Earl, D.A. and von Holdt, B.M. (2012). STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method. *Conservation Genetics Resources*, 4, 359-361. http://dx.doi.org/10.1007/s12686-011-9548-7

Engels, J. M., & Thormann, I. (2020). Main Challenges and Actions Needed to Improve Conservation and Sustainable Use of Our Crop Wild Relatives. *Plants*, *9*(8), 968. https://doi.org/10.3390/plants9080968

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., et al. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE,* 6(5):e19379. https://doi:10.1371/journal.pone.0019379.

Fick, S.E. and Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37, 4302–4315. https://doi.org/10.1002/joc.5086.

Forester, B.R., Lasky, J.R., Wagner, H.H., Urban, D.L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology*, 27(9):2215–2233

Gaudet, P., Livstone, M. S., Lewis, S. E., & Thomas, P. D. (2011). Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Briefings in Bioinformatics*, *12*(5), 449-462. https://doi.org/10.1093/bib/bbr042

Gibson, M. J. S., & Moyle, L. C. (2020). Regional differences in the abiotic environment contribute to genomic divergence within a wild tomato species. *Molecular Ecology*, *29*(12), 2204-2217. https://doi.org/10.1111/mec.15477

Gong, Z., Dong, C., Lee, H., Zhu, J., Xiong, L., Gong, D., Stevenson, B., & Zhu, J. (2005). A DEAD Box RNA Helicase Is Essential for mRNA Export and Important for Development and Stress Responses in Arabidopsis. *The Plant Cell*, *17*(1), 256-267. https://doi.org/10.1105/tpc.104.027557

Gong, Z., Lee, H., Xiong, L., Jagendorf, A., Stevenson, B., & Zhu, J. (2002). RNA helicase-like protein as an early regulator of transcription factors for plant chilling and freezing tolerance. *Proceedings of the National Academy of Sciences*, *99*(17), 11507-11512. https://doi.org/10.1073/pnas.172399299

Gramazio, P., Yan, H., Hasing, T., Vilanova, S., Prohens, J., & Bombarely, A. (2019). Whole-Genome Resequencing of Seven Eggplant (Solanum melongena) and One Wild Relative (S. incanum) Accessions Provides New Insights and Breeding Tools for Eggplant Enhancement. *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.01220

Gramazio, P., Prohens, J., Plazas, M., Mangino, G., Herraiz, F. J., Garcia-Fortea, E., & Vilanova, S. (2018). Genomic Tools for the Enhancement of Vegetable Crops: A Case in Eggplant. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, *46*(1), 1–13. <https://doi.org/10.15835/nbha46110936>

Gramazio, P., Prohens, J., Plazas, M., Mangino, G., Herraiz, F. J., & Vilanova, S. (2017). Development and Genetic Characterization of Advanced Backcross Materials and An Introgression Line Population of *Solanum incanum* in a *S. melongena* Background. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.01477

Haupt, M. and Schmid, K. (2022). Using landscape genomics to infer genomic regions involved in environmental adaptation of soybean genebank accessions. bioRxiv 2022.02.18.480989

Hengl, T., Mendes, de Jesus J., Heuvelink, G.B.M., Ruiperez, G. M., Kilibarda, M., et al. (2017) SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, 12(2): e0169748.<https://doi.org/10.1371/journal.pone.0169748>

Hijmans, R. (2023). \_raster: Geographic Data Analysis and Modeling\_. R package version 3.6-14, https://CRAN.R-project.org/package=raster.

Hijman, R.J., Phillips, S., Leathwick, J., Elith, J. (2022). \_dismo: Species Distribution Modelling\_. R package version 1.3-9, https://CRAN.R-project.org/package=dismo

Huang, K., Jahani, M., Gouzy, J., Legendre, A., Carrere, S., Miguel, J., González Segovia, E. G., Todesco, M., Mayjonade, B., Rodde, N., Cauet, S., Dufau, I., Staton, S. E., Pouilly, N., Boniface, M., Tapy, C., Mangin, B., Duhnen, A., Gautier, V., Rieseberg, L. H. (2023). The genomics of linkage drag in inbred lines of sunflower. *Proceedings of the National Academy of Sciences*, *120*(14), e2205783119. <https://doi.org/10.1073/pnas.2205783119>

Inouye, D.W. (2008). Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. *Ecology*, 89(2):353-62. https://doi:10.1890/06-2128.1.

James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). An Introduction to Statistical Learning: With Applications in R. 1st ed. 2013, Corr. 7th printing 2017 edition. Springer; 2013.

Pritchard, J.K., Stephens, M., Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data, *Genetics*, 155(2) 945–959. <https://doi.org/10.1093/genetics/155.2.945>

Joswig, J.S., Wirth, C., Schuman, M.C. *et al.* (2022). Climatic and soil factors explain the two-dimensional spectrum of global plant trait variation. *Nature Ecology and Evolution,* **6**, 36–50. https://doi.org/10.1038/s41559-021-01616-8

Kapazoglou, A., Gerakari, M., Lazaridi, E., Kleftogianni, K., Sarri, E., Tani, E., Bebeli, P.J. (2023). Crop Wild Relatives: A Valuable Source of Tolerance to Various Abiotic Stresses. *Plants*, 12, 328. <https://doi.org/10.3390/plants12020328>

Knapp, S., Vorontsova, M. S., & Prohens, J. (2013). Wild Relatives of the Eggplant (Solanum melongena L.: Solanaceae): New Understanding of Species Names in a Complex Group. *PLOS ONE*, *8*(2), e57039. https://doi.org/10.1371/journal.pone.0057039

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179-1191. <https://doi.org/10.1111/1755-0998.12387>

Lasky, J.R., Upadhyaya, H.D., Ramu, P., Deshpande, S., Hash, C.T., Bonnette, J., Juenger, T.E., Hyma, K., Acharya, C., Mitchell, S.E., Buckler, E.S., Brenton, Z., Kresovich, S., Morris, G.P. (2015). Genome-environment associations in sorghum landraces predict adaptive traits. *Science Advances*, 1(6): e1400218. https://doi: 10.1126/sciadv.1400218.

Lasky, J.R., Des Mmarais, D.L., McKay, J.K., Richards, J.H., Juenger, T.E. and Keitt, T.H. (2012). Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Molecular Ecology*, 21: 5512-5529. <https://doi.org/10.1111/j.1365-294X.2012.05709.x>

Legendre, P. and Legendre, L. (2012). Chapter 11 - Canonical analysis. In P. Legendre and L. Legendre, eds. Developments in Environmental Modelling. Numerical Ecology. Elsevier, pp. 625–710. https://doi.org/10.1016/B978-0-444-53868-0.50011-3

Lee, C.R., Mitchell-Olds, T. (2011). Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology*, 20(22):4631-42. https://doi: 10.1111/j.1365-294X.2011.05310.x.

Lei, L., Poets, A. M., Liu, C., Wyant, S. R., Hoffman, P. J., Carter, C. K., Shaw, B. G., Li, X., Muehlbauer, G. J., Katagiri, F., & Morrell, P. L. (2019). Environmental Association Identifies Candidates for Tolerance to Low Temperature and Drought. *G3 Genes|Genomes|Genetics*, *9*(10), 3423-3438. <https://doi.org/10.1534/g3.119.400401>

Li, Y., Chen, F., Yang, Y., Han, Y., Ren, Z., Li, X., J. Soppe, W. J., Cao, H., & Liu, Y. (2023). The Arabidopsis pre-mRNA 3′ end processing related protein FIP1 promotes seed dormancy via the DOG1 and ABA pathways. *The Plant Journal*. https://doi.org/10.1111/tpj.16239

Li, Y., Huang, Y., Bergelson, J., Nordborg, M., & Borevitz, J. O. (2010). Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, *107*(49), 21199-21204. https://doi.org/10.1073/pnas.1007431107

Li, H., Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14):1754-60. https://doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Data Processing Subgroup, G. P. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079. <https://doi.org/10.1093/bioinformatics/btp352>

Lin, P., Wu, H., Chan, K., and Schafleitner, R. (2022). De novo SNP calling reveals the genetic differentiation and morphological divergence in genus Amaranthus. *The Plant Genome*, *15*(2). <https://doi.org/10.1002/tpg2.20206>

Lin, Y-P., Mitchell-Olds, T., Lee, C-R. (2021). The ecological, genetic and genomic architecture of local adaptation and population differentiation in Boechera stricta. Proc. R. Soc. B 288: 20202472.<https://doi.org/10.1098/rspb.2020.2472>

Lü, S., Zhao, H., Des Marais, D. L., Parsons, E.P., Wen, X., Xu, X., Bangarusamy, D.K., Wang, G., Rowland, O., Juenger, T., Bressan, R. A., Jenks, M.A. (2012). Arabidopsis *ECERIFERUM9* Involvement in Cuticle Formation and Maintenance of Plant Water Status, *Plant Physiology*, 159(3), 930–944,<https://doi.org/10.1104/pp.112.198697>

Luu, K., Bazin, E., Blum, M.G. (2016). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17(1):67-77. https://doi: 10.1111/1755-0998.12592

Manel, S., Joost, S., Epperson, B.K., Holdereggere, R., Storfer, A., Rosenberg, M.S., Scribner, K.T., Bonin, A. and Fortin, M.J. (2010), Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, 19: 3760-3772.  [https://doi.org/10.1111/j.1365-294X.2010.04717.x](about:blank)

Mdladla, K., Dzomba, E. F., & Muchadeyi, F. C. (2018). Landscape genomics and pathway analysis to understand genetic adaptation of South African indigenous goat populations. *Heredity*, *120*(4), 369-378. https://doi.org/10.1038/s41437-017-0044-z

Meyer, R. S., Karol, K. G., Little, D. P., Nee, M. H., & Litt, A. (2012). Phylogeographic relationships among Asian eggplants and new perspectives on eggplant domestication. *Molecular Phylogenetics and Evolution*, *63*(3), 685-701. <https://doi.org/10.1016/j.ympev.2012.02.006>

Morente-López, J., García, C., Draper, D., & Iriondo, J. M. (2018). Geography and Environment Shape Landscape Genetics of Mediterranean Alpine Species Silene ciliata Poiret. (Caryophyllaceae). *Frontiers in Plant Science*, *9*. <https://doi.org/10.3389/fpls.2018.01698>

Müller, J. V., Cockel, C. P., Gianella, M., & Guzzon, F. (2021). Treasuring crop wild relative diversity: Analysis of success from the seed collecting phase of the ‘Adapting Agriculture to Climate Change’ project. Genetic Resources and Crop Evolution, 68(7), 2749-2756. doi: [10.1007/s10722-021-01229-x](https://doi.org/10.1007/s10722-021-01229-x)

Musa, I., Rafii, M. Y., Ahmad, K., Ramlee, S. I., Md Hatta, M. A., Oladosu, Y., Chukwu, S. C., Mat Sulaiman, N. N., Ayanda, A. F., & Halidu, J. (2020). Effects of Grafting on Morphophysiological and Yield Characteristic of Eggplant (Solanum melongena L.) Grafted onto Wild Relative Rootstocks. *Plants*, *9*(11). <https://doi.org/10.3390/plants9111583>

Nieves-Cordones, M., Caballero, F., Martínez, V., Rubio, F., (2012). Disruption of the *Arabidopsis thaliana* Inward-Rectifier K+ Channel AKT1 Improves Plant Responses to Water Stress, *Plant and Cell Physiology*, 53(2), 423–432,<https://doi.org/10.1093/pcp/pcr194>

Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M., Lahti, L., McGlinn, D., Ouellette, M., Ribeiro, C. E., Smith, T., Stier, A., Ter Braak, C., Weedon, J. (2022). vegan: Community Ecology Package\_. R package version 2.6-4, [https://CRAN.R-project.org/package=vegan](https://cran.r-project.org/package=vegan)

Pandey, S., Monshausen, G. B., Ding, L., & Assmann, S. M. (2008). Regulation of root-wave response by extra large and conventional G proteins in Arabidopsis thaliana. *The Plant Journal*, *55*(2), 311-322. https://doi.org/10.1111/j.1365-313X.2008.03506.x

Pasbani, B., Salimi, A., Aliasgharzad, N., & Hajiboland, R. (2020). Colonization with arbuscular mycorrhizal fungi mitigates cold stress through improvement of antioxidant defense and accumulation of protecting molecules in eggplants. *Scientia Horticulturae*, *272*, 109575. https://doi.org/10.1016/j.scienta.2020.109575

Peri, P.L., Rosas, Y.M., Ladd, B., Toledo, S., Lasagno, R.G., Martínez, P. G. (2019). Modeling Soil Nitrogen Content in South Patagonia across a Climate Gradient, Vegetation Type, and Grazing. *Sustainability*, 11(9):2707. https://doi.org/10.3390/su11092707

Plazas, M., González-Orenga, S., Nguyen, H. T., Morar, I. M., Fita, A., Boscaiu, M., Prohens, J., and Vicente, O. (2022). Growth and antioxidant responses triggered by water stress in wild relatives of eggplant. *Scientia Horticulturae*, 293, 110685. https://doi.org/10.1016/j.scienta.2021.110685

Plazas, M., Vilanova, S., Gramazio, P., Rodríguez-Burruezo, A., Fita, A., Herraiz, F.J., et al. (2016). Interspeciﬁc hybridization between eggplant and wild relatives from diﬀerent genepools. *Journal of the American Society for Horticultural Science*, 141, 34–44

Qin, L., Walk, T.C., Han, P., Chen, L., Zhang, S., Li, Y., Hu, X., Xie, L., Yang, Y., Liu, J., Lu, X., Yu, C., Tian, J., Shaff, J.E., Kochian, L.V., Liao, X., Liao, H. (2019). Adaption of Roots to Nitrogen Deficiency Revealed by 3D Quantification and Proteomic Analysis. *Plant Physiology*, 179(1):329-347. https://doi: 10.1104/pp.18.00716

Rambaut, A. (2010) FigTree v1.3.1. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. http://tree.bio.ed.ac.uk/software/figtree/

Rakha, M., Namisy, A., Chen, R., El-Mahrouk, M. E., Metwally, E., Taha, N., Prohens, J., Plazas, M., & Taher, D. (2020). Development of Interspecific Hybrids between a Cultivated Eggplant Resistant to Bacterial Wilt (Ralstonia solanacearum) and Eggplant Wild Relatives for the Development of Rootstocks. *Plants*, *9*(10). <https://doi.org/10.3390/plants9101405>

Ren, L., Qi, N., Feng, Q., Zhao, S., Zhao, S., Wang, Y., & Wu, H. (2013). Calcineurin B-like protein CBL10 directly interacts with AKT1 and modulates K+ homeostasis in Arabidopsis. *The Plant Journal*, *74*(2), 258-266. https://doi.org/10.1111/tpj.12123

Renzi, J. P., Coyne, C. J., Berger, J., von Wettberg, E., Nelson, M., Ureta, S., Hernández, F., Smýkal, P., & Brus, J. (2022). How Could the Use of Crop Wild Relatives in Breeding Increase the Adaptation of Crops to Marginal Environments? *Frontiers in Plant Science*, *13*. <https://doi.org/10.3389/fpls.2022.886162>

Richardson, J.L., Brady, S.P., Wang, I.J. and Spear, S.F. (2016), Navigating the pitfalls and promise of landscape genetics. *Molecular Ecology*, 25: 849-863. <https://doi.org/10.1111/mec.13527>

Sanjur, O. I., Piperno, D. R., Andres, T. C., & Wessel-Beaver, L. (2002). Phylogenetic relationships among domesticated and wild species of Cucurbita (Cucurbitaceae) inferred from a mitochondrial gene: Implications for crop plant evolution and areas of origin. Proceedings of the National Academy of Sciences, 99(1), 535-540.

Shafiq, S., Berr, A. and Shen, W.-H. (2014). Combinatorial functions of diverse histone methylations in *Arabidopsis thaliana* flowering time regulation. *New Phytologist*, 201: 312-322.<https://doi.org/10.1111/nph.12493>

Rellstab, C., Gugerli, F., Eckert, A.J., Hancock, A.M. and Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, 24: 4348-4370.<https://doi.org/10.1111/mec.13322>

Stetter, M. G., Thornton, K., and Ross-Ibarra, J. (2018). “Genetic architecture and selective sweeps after polygenic adaptation to distant trait optima.” *PLOS Genetics*, 14(11), e1007794.

Steane, D. A., Potts, B. M., McLean, E., Prober, S. M., Stock, W. D., Vaillancourt, R. E., & Byrne, M. (2014). Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Molecular Ecology*, 23, 2500– 2513. <https://doi.org/10.1111/mec.12751>

Syfert, M. M., Castañeda-Álvarez, N. P., Khoury, C. K., Särkinen, T., Sosa, C. C., Achicanoy, H. A., Bernau, V., Prohens, J., Daunay, C., & Knapp, S. (2016). Crop wild relatives of the brinjal eggplant (*Solanum melongena*): Poorly represented in genebanks and many species at risk of extinction. *American Journal of Botany*, *103*(4), 635-651. https://doi.org/10.3732/ajb.1500539

Terés, J., Busoms, S., Martín, L. P., Luís-Villarroya, A., Flis, P., Álvarez-Fernández, A., Tolrà, R., Salt, D. E., & Poschenrieder, C. (2019). Soil carbonate drives local adaptation in Arabidopsis thaliana. *Plant, Cell & Environment*, *42*(8), 2384-2398. https://doi.org/10.1111/pce.13567

Tiwari, L. D., & Grover, A. (2019). Cpn60β4 protein regulates growth and developmental cycling and has bearing on flowering time in Arabidopsis thaliana plants. *Plant Science*, *286*, 78-88. https://doi.org/10.1016/j.plantsci.2019.05.022

Tripodi, P., Rabanus-Wallace, M. T., Barchi, L., Kale, S., Esposito, S., Acquadro, A., and Stein, N. (2021). Global range expansion history of pepper (Capsicum spp.) revealed by over 10,000 genebank accessions. *Proceedings of the National Academy of Sciences*, *118*(34), e2104315118.

Wang, Y., Ribot, C., Rezzonico, E., Poirier, Y. (2004). Structure and Expression Profile of the Arabidopsis *PHO1* Gene Family Indicates a Broad Role in Inorganic Phosphate Homeostasis, *Plant Physiology*, 135(1), 400–411,<https://doi.org/10.1104/pp.103.037945>

Weese, T. L., and Bohs, L. (2010). Eggplant origins: Out of Africa, into the Orient. *Taxon*, 59:49-56.

Wickham, H. ggplot2: Elegant Graphics for data analysis. Springer-Verlag New York, 2016.

WFO (2023): World Flora Online. Published on the Internet; [http://www.worldfloraonline.org](http://www.worldfloraonline.org/). Accessed on: 16 Jun 2023

Yadav, S., Stow, A. J., & Dudaniec, R. Y. (2021). Microgeographical adaptation corresponds to elevational distributions of congeneric montane grasshoppers. *Molecular Ecology*, *30*(2), 481-498. <https://doi.org/10.1111/mec.15739>

**Data availability statement**

Sample descriptions and environmental variables are provided as supporting materials. The SNP dataset is available on Dryad (<https://datadryad.org/stash/share/QF9UI_3NOWOwwy8S3UINq0qXMwzbVWTBO2uvMWsGvz0>).

R codes are available on github.com (<https://github.com/eomosh/AEPlant/blob/main/epafro154_GEA.R>).

**Author contribution**

Maarten van Zonneveld, Emmanuel Omondi: Conceived the study, wrote the manuscript

Emmanuel O. Omondi: Analyzed the data

Chen-yu Lin, Shu-Mei Huang, Cheng-An Liao: DNA extraction, library preparations, SNP calling

Ya-Ping Lin: Review and editing

Oliva Ricardo: Review and editing

All authors critically revised and approved the manuscript.

**Tables and figures**

**Table 1**: The contribution of the environmental variables (climatic and soil), population structure, and geographical distances to the genomic variation and observed genetic structure in the partial RDA models.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Models | R2 | adj R2 | p (>F) | Proportion of total variance |
| SNP variation | Full: Y ~ env + Struct. + spatial | 0.137 | 0.042 | 0.0012 \*\* | 0.14 |
| Y ~ Env + | (Struct. + Spatial) | 0.076 | 0.013 | 0.045 \* | 0.08 |
| Y ~ Struct. | (Env + Spatial) | 0.022 | 0.010 | 0.008 \*\* | 0.02 |
| Y ~ Spatial | (Env + Struct.) | 0.029 | 0.011 | 0.013\* | 0.03 |
| Confounded Env. /Struct. / Spatial |  |  |  | 0.01 |
| Total unexplained |  |  |  | 0.86 |
| Structure | Struct. ~Env. + Spatial | 0.194 | 0.119 | 0.000 \*\*\* | 0.19 |
| Struct. ~ Env. |(Spatial) | 0.108 | 0.051 | 0.025 \* | 0.11 |
| Struct. ~Spatial |(Env) | 0.045 | 0.029 | 0.023 \* | 0.05 |
| Confounded Env. / Spatial |  |  |  | 0.03 |
| Total unexplained |  |  |  | 0.81 |

Env = Environmental variables; Spatial = Geographical distances (spatial autocorrelations); Struct. = Population structure

**Table 2**: Ten common candidate SNPs between LFMM and partial RDA methods and their gene annotations.

|  |  |  |  |
| --- | --- | --- | --- |
| **SNP** | **Environmental variable** | **Gene/ Protein** | **Function** |
| Chr3:85141788 | Nitrogen | *SIA2* | Required for efficient pollen grain germination and pollen tube elongation **1** |
| Chr3:85141951 | Nitrogen | *SIA2* | Required for efficient pollen grain germination and pollen tube elongation **1** |
| Chr6:723548 | Solar radiation | *RPL19B* | RNA binding **2** |
| Chr7:81532866 | Nitrogen | *FIP1* | Regulation of seed dormancy and germination processes **3** |
| Chr7:81533032 | Nitrogen | *FIP1* | Regulation of seed dormancy and germination processes**3** |
| Chr7:104960610 | Nitrogen | *-* | Protein of unknown function |
| Chr10:73179037 | Nitrogen | *CPN60B4* | Plays an important role(s) in chloroplast development and is a key factor in plant growth, development, and flowering in Arabidopsis **4** |
| Chr10:73179139 | Nitrogen | *CPN60B4* | Plays an important role(s) in chloroplast development. It is a key factor in plant growth, development, and flowering in Arabidopsis **4** |
| Chr10:79565787 | Nitrogen | *RH38* | Essential for mRNA Export and Important for Development and Stress (heat and cold) Responses in Arabidopsis **5, 6** |
| Chr10:79565885 | Nitrogen | *RH38* | Essential for mRNA Export and Important for Development and Stress (heat and cold) Responses in Arabidopsis **5, 6** |

**1** Daskalova et al., 2009;**2** Gaudet et al., 2011; **3** Li et al., 2023; **4** Tiwari and Grover, 2019; **5** Gong et al., 2005; **6** Gong et al., 2002



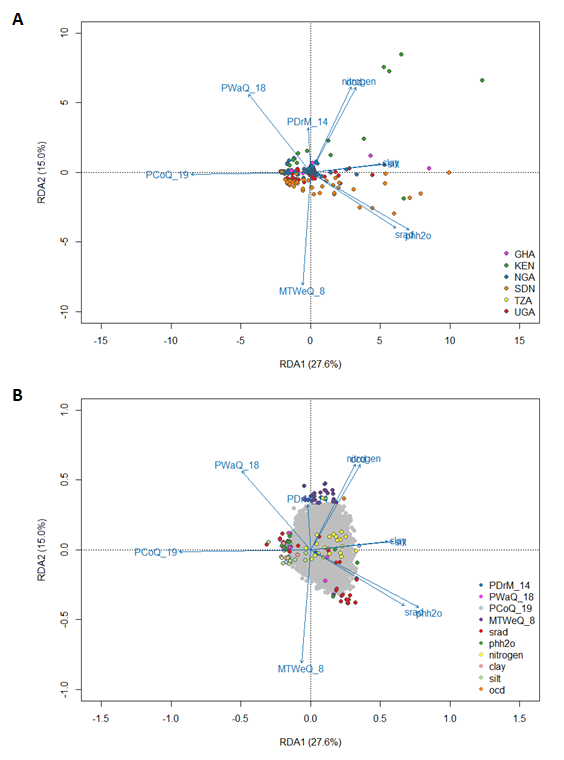
**Figure 1**: Photos of four of the 17 included eggplant species CWR. A: Solanum incanum L.; B: Solanum coagulans Forssk; C: Solanum dasyphyllum Schumach. & Thonn.; and D: Solanum cerasiferum Dunal.



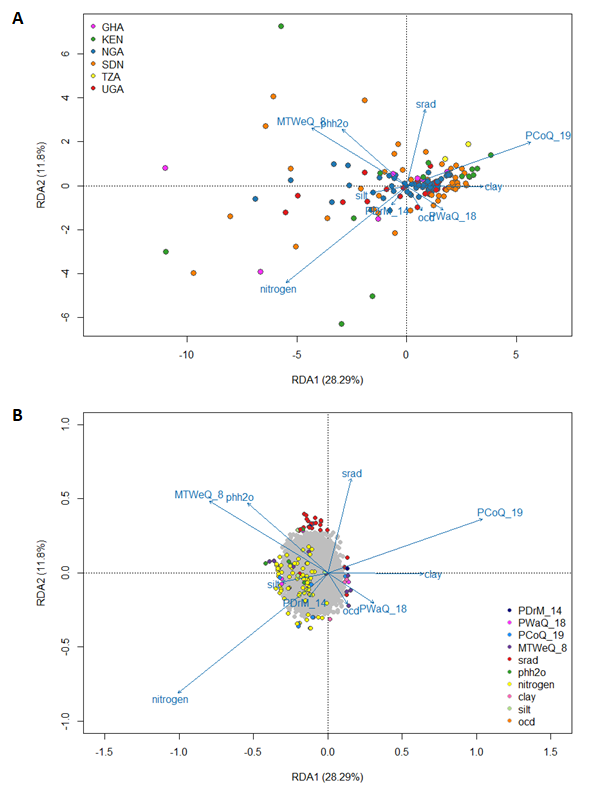




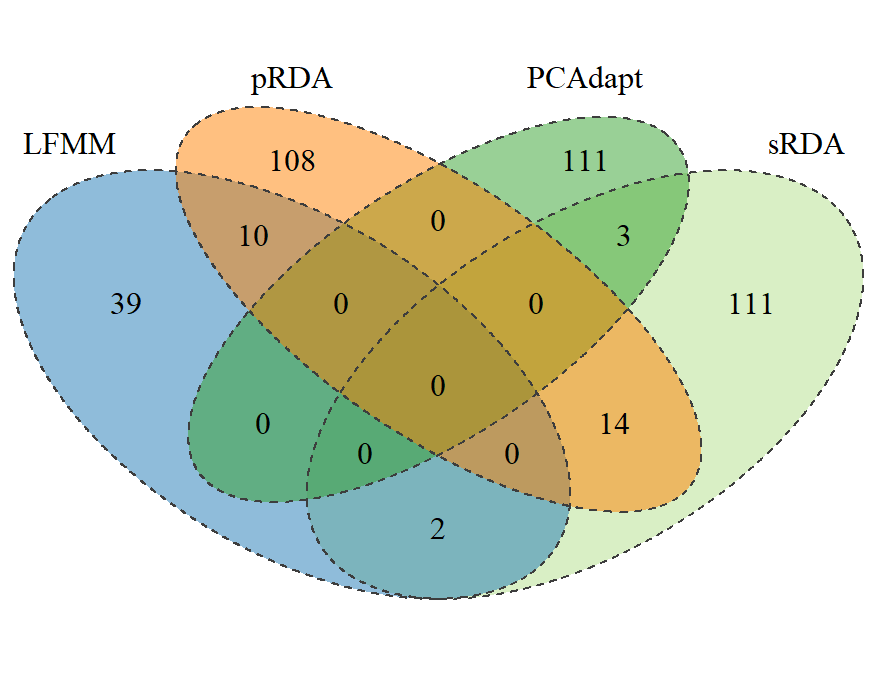
**Figure 2: A** Phylogenetic tree with colors corresponding to the five genetic groups formed in the bar plots of the structure analysis; **B** Map of the sampling areas in Western and Eastern Africa with pie charts showing the admixture proportions from structure analysis (optimal K = 8); **C** The group formations with sorted q values of the ancestry coefficient matrix. The bar plot represents individuals. The groups comprise species from different collection populations and gene pools. The main species for every group included: Groups I – S. anomalum, S. incanum, and S. aethiopicum; Group II - S. macrocarpon, S. dasyphyllum, S. anomalum, and S. incanum; Group III - S. cerasiferum and S. anguivi; Group IV - S. anguivi and S. anomalum; and Group V – S. campylacanthum. Groups I represent the Coagulans, Acanthophora, and Arundo clades; Groups II and V represent the Anguivi and Giganteum clades, while Groups III and IV represent the Melongena and Aculeastrum clades. Admixtures were observed mainly in groups I and III for accessions of S. incanum, S. cerasiferum, S. coagulans, and S. nigriviolaceum species.

****

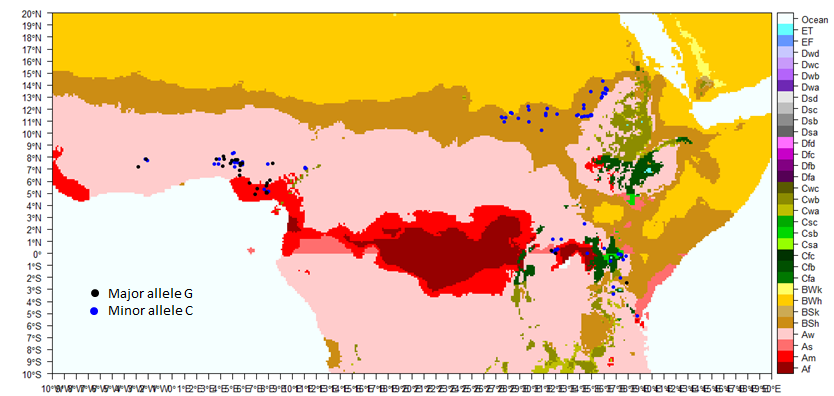
**Figure 3**: Biplot of simple RDA. Blue vectors indicate the direction and values of the environmental variables. **(A)** Colors correspond to the individual accession sampling sites by country. TZA: Tanzania; UGA: Uganda; KEN: Kenya; SDN: Sudan; NGA: Nigeria; GHA: Ghana; **(B)** Outlier SNPs colors correspond to the environmental variable with the strongest association. PDrM\_14: Precipitation of the driest quarter; PWaQ\_18: Precipitation of the warmest quarter; PCoQ\_19: Precipitation of the coldest quarter; MTWeQ\_8: Mean temperature of the wettest quarter; srad: Solar radiation; phh2o: soil pH; nitrogen: Soil nitrogen content; clay: soil clay content; silt: Soil silt content; ocd: organic carbon density.



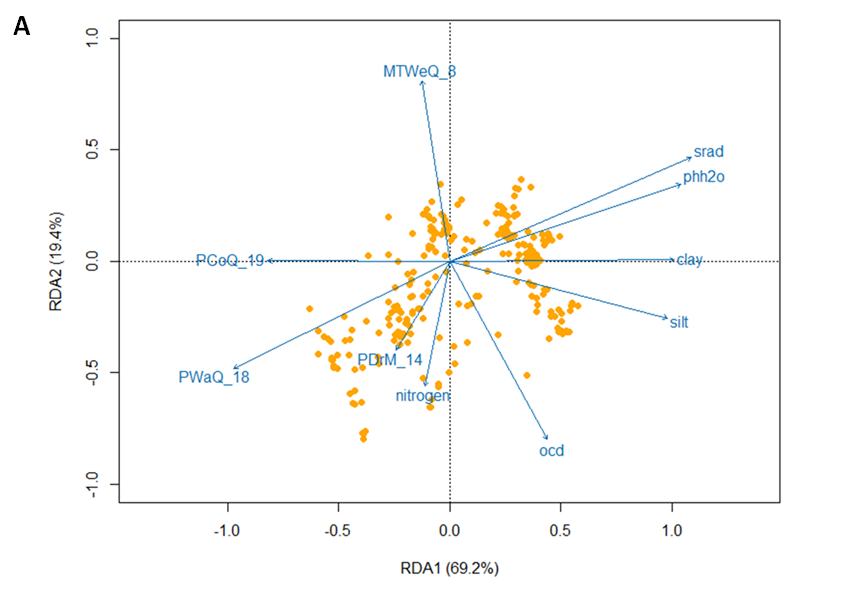
**Figure 4**:Biplot of partial RDA conditioned on geographical distances and population structure. Blue vectors indicate the direction and values of the environmental variables. **(A)** The individual accessions are highlighted in color based on the sampling sites by country. TZA: Tanzania; UGA: Uganda; KEN: Kenya; SDN: Sudan; NGA: Nigeria; GHA: Ghana; **(B)** Outlier SNPs are highlighted in color based on the environmental variable with the strongest correlation. PDrM\_14: Precipitation of the driest quarter; PWaQ\_18: Precipitation of the warmest quarter; PCoQ\_19: Precipitation of the coldest quarter; MTWeQ\_8: Mean temperature of the wettest quarter; srad: Solar radiation; phh2o: soil pH; nitrogen: Soil nitrogen content; clay: soil clay content; silt: Soil silt content; ocd: organic carbon density.

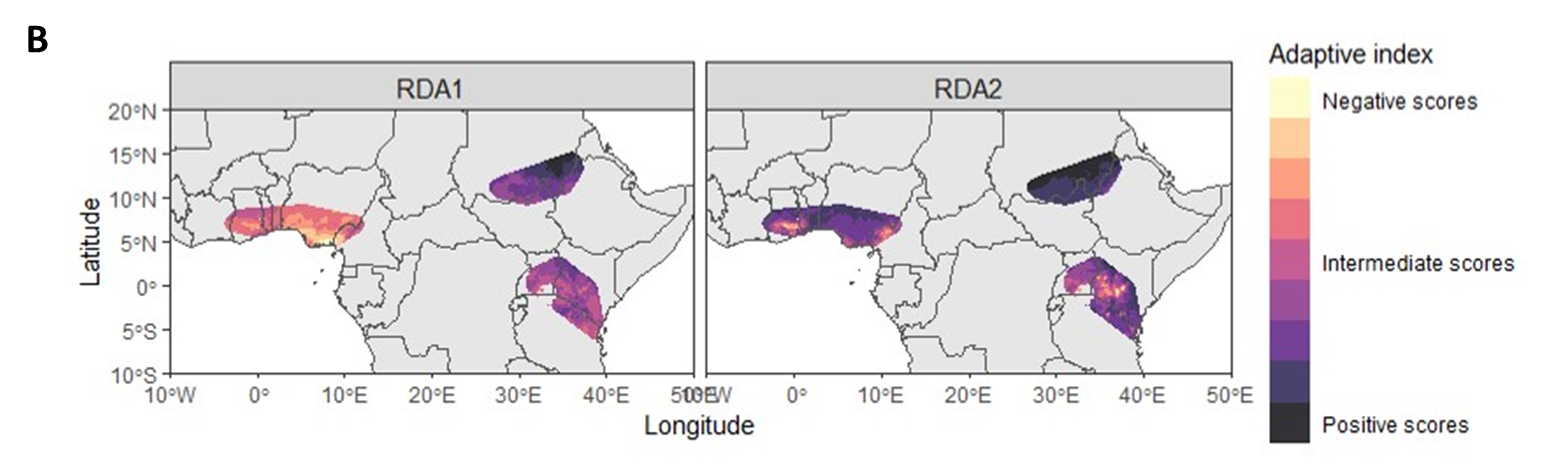


**Figure 5**: A Venn diagram showing the number of significant SNPs detected by four genome scan methods.



**Figure 6**: The sampling points and allele frequency distribution of a candidate SNP Chr7:81532866 detected by partial RDA linked to protein FIP1 expressed in response to water deprivation and salt stress. The colors represent the Köppen climate classification (Detailed description in **Table S8**). The West African sampling climates include the tropical monsoon and tropical savanna with dry winter. East African environments consist of Sudan's dry semi-arid hot climates (BSh) and Kenyan and Ugandan tropical and temperate climates (Af, As, Aw, BSh, and Cfb). The first letters in the climate classification acronyms represent A- tropical climate; B- Arid climate; C- Temperate climate; D- cold continental climate; E- Polar climate.





**Figure 7:** Adaptive landscape with **(A)** the adaptively enriched genetic space showing the association between adaptive loci and climatic drivers of adaptation and **(B)** the spatial projection of the adaptability across the study areas. PDrM\_14: Precipitation of the driest quarter; PWaQ\_18: Precipitation of the warmest quarter; PCoQ\_19: Precipitation of the coldest quarter; MTWeQ\_8: Mean temperature of the wettest quarter; srad: Solar radiation; phh2o: soil pH; nitrogen: Soil nitrogen content; clay: soil clay content; silt: Soil silt content; ocd: organic carbon density.