REVISTA BIO CIENCIAS
http://revistabiociencias.uan.edu.mx
https://doi.org/10.15741/revbio.12.e1747
ISSN 2007-3380

*Original article / Artículo original*

# Diversity and genetic structure in populations of *Moringa oleifera* Lam. cultivated in mexico through SNP markers

# Diversidad y estructura genética en poblaciones de *Moringa oleifera* Lam. cultivadas en México a través de marcadores SNP

*Ruiz Hernández, R.* [1], *Pérez Vázquez, A.* [1], *Hernández Rodríguez, M.* [2*], *Rayas Amor A. A.* [3].

[1]*Campus Veracruz. Colegio de Postgraduados. Carretera Xalapa Veracruz km 88.5, Tepetates. 91690, Manlio Fabio Altamirano, Veracruz, México.*
[2]*Campus Montecillo. Colegio de Postgraduados. Calle, km 36.5, Montecillo. 56230, Texcoco de Mora, Estado de México, México.*
[3]*Unidad Lerma. Universidad Autónoma Metropolitana. Av. De las Garzas No. 10, Colonia El Panteón, C.P. 52005. Lerma de Villada, Estado de México, México.*

**Please cite this article as/Como citar este artículo:** Ruiz-Hernández, R., Pérez-Vázquez, A., Hernández- Rodríguez, M., Rayas-Amor A. A. (2025). Diversity and genetic structure in populations of Moringa oleífera cultivated in mexico through SNP markers. *Revista Bio Ciencias*, 12, e1747. https://doi.org/10.15741/revbio.12.e1747

**Article Info/Información del artículo**
Received/Recibido: Septiembre 7th 2024.
Accepted/Aceptado: June 19h 2025.
Available on line/Publicado: July 08th 2025.

## ABSTRACT

*Moringa oleifera* is a widely distributed species due to its adaptive capacity and nutritional properties. In Mexico, several reintroductions of this species have taken place, and although morphological diversity has been observed, its genetic diversity at the DNA level remains unknown. This study aimed to determine the genetic diversity and structure of 14 cultivated populations of *M. oleifera* in Mexico using SNP markers. Seeds were germinated in a greenhouse, and DNA was extracted from young leaves and stems using a CTAB-based protocol. This DNA was then used for genotyping each population with DArTseq™ technology. A total of 9,862 SNPs were obtained, of which 0.64 % were polymorphic. Within populations, moderate levels of diversity were observed (Ae = 1.52, He = 0.33, Ho = 0.44, Shannon index = 0.44). The genetic background of the populations revealed two groups through cluster analysis and PCoA. Population genetic structure analysis using STRUCTURE indicated ancestry from two original populations (K = 2). The 14 moringa populations exhibit genetic diversity that can be leveraged for sustainable management and conservation.

**KEY WORDS:** Polymorphism, genotyping, genetic resources.

***Corresponding Author:**

**Martha Hernández-Rodríguez**. Posgrado en Recursos Genéticos y Productividad – Genética. Campus Montecillo. Colegio de Postgraduados. km 36.5 carretera Federal México-Texcoco, 56264, Montecillo, Texcoco, Estado de México, México. Teléfono (595) 952 0265. E-mail: hernandez.martha@colpos.mx

Revista Bio ciencias
ISSN 2007-3380

**RESUMEN**

*Moringa oleifera* es una especie con amplia distribución mundial debido a su capacidad adaptativa y a sus propiedades nutricionales. En México, han ocurrido varias reintroducciones de esta especie y aunque se ha observado diversidad morfológica, se desconoce su diversidad a nivel de DNA. El objetivo fue determinar la diversidad y estructura genética de 14 poblaciones de *M. oleifera* cultivadas en México a través de SNP. Las semillas se germinaron en invernadero y el DNA se extrajo de hojas y tallos jóvenes mediante un protocolo basado en CTAB. Este DNA se utilizó para realizar el genotipado de cada población con la tecnología DArTseq™. Se obtuvieron 9,862 SNP, de ellos el 0.64 % fueron polimórficos. Dentro de las poblaciones se observaron niveles moderados de diversidad (Ae=1.52, He=0.33, Ho= 0.44, índice de Shannon = 0.44). El trasfondo genético de las poblaciones permitió visualizar dos grupos a través del análisis de conglomerados y PCoA. La estructura genética de las poblaciones mediante STRUCTURE indicó que la ascendencia proviene de dos poblaciones originales (K = 2). Las 14 poblaciones de moringa presentan diversidad genética que puede ser empleada para un manejo sostenible y su conservación.

**PALABRAS CLAVE:** Polimorfismo, genotipificación, recursos genéticos.

## Introduction

To address challenges such as hunger, malnutrition, low food productivity, and climate change, the agricultural system is compelled to identify and utilize new species. Additionally, the limited variety of consumed foods can lead to an imbalance in the diversity and quantity of micronutrients required for proper physiological function (Calicioglu *et al.,* 2019). Consequently, new crops must not only align with dietary trends and preferences but also exhibit high yield, improved nutritional content, greater tolerance to various stress factors, minimal environmental impact, and enhanced efficiency in the use of natural resources (Hendre *et al.,* 2019).

*Moringa oleifera* Lam. is a tree native to India that can grow at altitudes of up to approximately 1,800 masl. It is tolerant to high temperatures and can thrive in soils with low moisture and nutrient levels (Muhl *et al.,* 2011). Due to its adaptive traits, moringa can grow in diverse climates and is currently cultivable across 37 % of the Earth's surface. Its leaves contain significant amounts of proteins, essential amino acids, lipids, carbohydrates, fiber, and secondary metabolites (Ruiz-Hernández *et al.,* 2022a). Given its nutritional and dietary properties, moringa is included in the

Revista Bio ciencias
ISSN 2007-3380

list of African orphan crops and is considered a promising alternative for future food security (Hendre & Van Deynze, 2015). However, agricultural, technological, and scientific investment in this crop has been minimal, resulting in low productivity and limited agronomic management reports (Hendre *et al.,* 2019), highlighting the importance of generating new information about this species.

Moringa is a diploid species (2n = 2X = 28) with an allogamous reproductive system, and its genome has been sequenced to 91.78 %. The genome size is estimated at 281.92 Mb, with an estimated 19,465 genes (Tian *et al.,* 2015), though a later resequencing identified 32,062 putative protein-coding genes (Shyamli *et al.,* 2021). In Mexico, morphological differences have been observed among cultivated populations (Ruiz-Hernández *et al.,* 2021). The genetic structure of these populations, based on eight microsatellites, identified three original populations (Ruiz-Hernández *et al.,* 2024).

However, more molecular markers could provide a more detailed characterization of the genetic diversity and structure of moringa populations in Mexico. This variation among populations results from both their genetic makeup and environmental conditions, making DNA-level genetic diversity particularly relevant. Since DNA sequences are not affected by environmental factors, molecular markers serve as a key tool for assessing this diversity. Molecular markers are specific DNA sequences with a known genomic location, indicating the presence or proximity of a gene or regulatory region involved in the expression of a trait. Single-nucleotide polymorphisms, or SNPs, are a type of molecular marker that enables fine-scale assessments of genetic variation. While most SNPs do not have direct functional effects, some contribute to genotype expression, whether located in coding regions or not (Ousmael *et al.,* 2023).

SNPs are highly abundant in the genome, have a low mutation rate, and can be identified using various high-throughput platforms (Teklemariam *et al.,* 2022). Genotyping-by-sequencing now enables the detection of genome-wide polymorphisms at the single-nucleotide level, allowing for comprehensive assessments of genetic diversity across thousands of loci simultaneously (Nadeem *et al.,* 2018). Beyond diversity analysis, SNPs enable genomic selection, the development of genome-wide association maps, phylogenetic studies, and precise insights into the evolutionary history of target populations (Yirgu *et al.,* 2023). Their application has allowed the identification of favorable variants, which can later be used to select plants with desirable agronomic traits, such as enhanced nutrient concentration, greater resistance to pathogens, or improved adaptation to restrictive environmental conditions such as drought, high temperatures, nutrient limitations, soil salinization, or even flooding (Huster *et al.,* 2021; Elbasyoni *et al.,* 2022). Given that understanding the genetic diversity of moringa populations in Mexico can serve as a foundation for the species' genetic improvement, this study aimed to evaluate the genetic diversity and structure of *M. oleifera* populations cultivated in Mexico using SNP markers.

Revista Bio ciencias
ISSN 2007-3380

## Material and Methods

### Biological material

Fourteen moringa populations were selected across 11 states in Mexico (Veracruz, Chiapas, Guerrero, Sinaloa, Oaxaca, Guanajuato, San Luis Potosí, Hidalgo, Yucatán, Michoacán, and Estado de México) (Table 1). The selected populations were intended for the commercial sale of moringa products and seeds and were geographically distant from one another. A total of 500 g of seeds were collected from each population. For germination, 30 seeds per population were selected, disinfected with 10% chlorine for 5 minutes, and sown in 22 cm × 22 cm nursery bags filled with a soil and compost substrate in a 4:1 ratio. Thirty days after germination, 30 mg of leaf and stem tissue were collected from 10 individuals per population. This resulted in a final composite sample of 300 mg per population, which was stored at -80°C until DNA extraction. A map was created using the coordinate reference system and the state layer from INEGI with the QGIS software.

**Table 1. Geographic location of the evaluated populations of *M. oleifera* in Mexico.**

| No. | Population | State | Municipality | Locality | Longitude | Latitude |
|---|---|---|---|---|---|---|
| **1** | *CHI* | *Chiapas* | *Tuxtla Gutiérrez* | *Colonia La Salle* | *-93.0868889* | *16.7429444* |
| **2** | *EMEX* | *Estado de México* | *Texcoco* | *San Pedro Santa Úrsula* | *-98.898385* | *19.497082* |
| **3** | *GTO* | *Guanajuato* | *Soledad de Gasca* | *Celaya* | *-100.8146904* | *20.502528* |
| **4** | *HGO* | *Hidalgo* | *San Felipe Orizatlán* | *Ahuatitla* | *-98.6660845* | *21.1630165* |
| **5** | *MICH1* | *Michoacán* | *Benito Juárez* | *El Rodeo* | *-100.4708226* | *19.3055772* |
| **6** | *MICHMP* | *Michoacán* | *Múgica* | *Múgica* | *-102.180997* | *18.928047* |
| **7** | *NL* | *Nuevo León* | *Escobedo* | *Francisco I. Madero* | *-100.2847444* | *25.78544732* |
| **8** | *OAX* | *Oaxaca* | *Santa Cruz Xoxocotlán* | *San Juan Bautista La Raya* | *-96.7280556* | *16.9791667* |
| **9** | *SLP* | *San Luis Potosí* | *Tanlajas* | *Guayajox* | *-98.73666667* | *21.71444444* |
| **10** | *SIN* | *Sinaloa* | *Culiacán* | *Valle* | *-107.37889* | *24.985* |
| *11* | *YUC1* | *Yucatán* | *Peto* | *Teshan* | *-88.62125* | *20.1486389* |
| **12** | *YUC2* | *Yucatán* | *Baca* | *Felipe Carrillo Puerto* | *-89.60700993* | *20.9954688* |
| **13** | *OAXH* | *Oaxaca* | *Santa María Huatulco* | *La Herradura* | *-96.3658333* | *15.7772222* |
| **14** | *GRO* | *Guerrero* | *Tecpan de Galeana* | *Mitla* | *-99.89343517* | *16.87894246* |

Source: Self made.

Revista Bio ciencias
ISSN 2007-3380

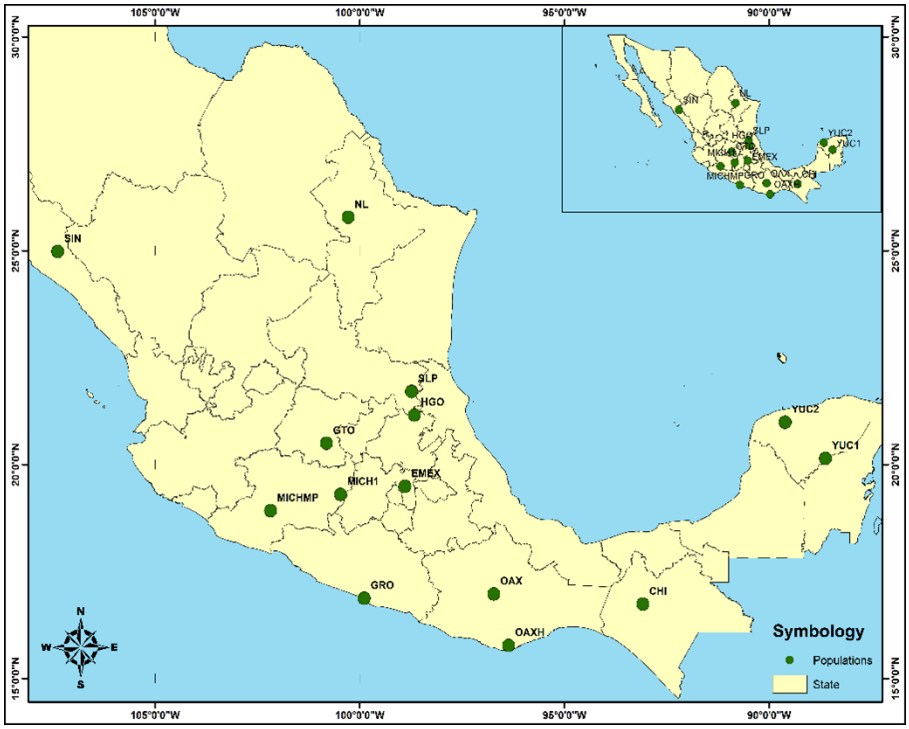


**Figure 1. Geographic location of *M. oleifera* populations cultivated in Mexico scanned in this study.**

Source: Self-made.

## DNA extraction

Composite samples were ground using a polypropylene pestle and liquid nitrogen in 2 mL polypropylene tubes. The resulting material was used for DNA extraction following a small-scale CTAB-based protocol (CIMMYT, 2006). Briefly, 1 mL of lysis solution (100 mM Tris-HCl, 700 mM NaCl, 50 mM EDTA, 2 % CTAB, and 1 % β-mercaptoethanol) was added and mixed until complete tissue resuspension, straightaway, the mixture was incubated at 65 °C in a water bath for 60 minutes, with inversion mixing every 10 minutes. After incubation, samples were centrifuged at 11,500 rpm for 10 minutes at room temperature (RT) to separate the three phases. Then, 800 µL of the upper phase was transferred to a clean, sterile 2.0 mL tube. DNA purification was performed by adding 500 µL of chloroform:isoamyl alcohol (24:1), followed by constant inversion mixing for 10 minutes at RT. The mixture was centrifuged at 11,500 rpm for 10 minutes at RT, and 600 µL of the supernatant was transferred to a sterile 1.5 mL tube. DNA was precipitated by adding 300 µL of isopropanol and incubating for 1 hour at -20 °C. The precipitated DNA was centrifuged at 11,500 rpm for 10 minutes at RT and washed with 1 mL of cold 70 % ethanol. The DNA pellet was recovered by centrifugation at 8,500 rpm for 5 minutes at RT. Finally, the supernatant was removed,

Revista Bio ciencias
ISSN 2007-3380

and the pellet was air-dried for 2 hours at RT before being dissolved in 70 µL of HPLC-grade water. DNA concentration was measured using a microvolume spectrophotometer (NanoDrop™ 2000, Thermo Scientific, Waltham, MA, USA). DNA integrity was assessed through electrophoresis on a 1 % agarose gel (Sigma, MO, USA), using 2 µL of DNA, migrated at 80 volts for 40 minutes. The gel was visualized under UV light using a MiniBisPro 16 mm® transilluminator (DNR Bio-Imaging Systems®, Israel).

### Genotyping

Genotyping was conducted by the Genotyping Service for Agriculture at CIMMYT using the DArTseq protocol. For this process, 60 µL of DNA at a concentration of 50 ng $\mu L^{-1}$ was sent for analysis. The proprietary DArT genotyping method (https://www.diversityarrays.com/services/dartseq/) was applied, utilizing the restriction enzymes *Pst*I (CTGCA/G) and *Ase*I (AT/TAAT) for DNA fragmentation, followed by next-generation sequencing on a NovaSeq™ 6000 platform (Illumina). SNP calling was performed using DArTsoft (http://www.diversityarrays.com/software.html#dartsoft). The SNP sequences were received in a double-row data matrix with scores of "-", "0", and "1". This matrix was recoded, assigning "0" to the dominant homozygous genotype, "1" to the recessive homozygous genotype, and "2" to the heterozygous genotype.

### Data analysis

The SNP matrix was cleaned to retain high-quality sequences by removing those with 25 % missing data and a minor allele frequency below 5 % (MAF < 0.05). From this matrix dataset, expected heterozygosity (He), observed heterozygosity (Ho), effective allele number (Ae), rare allele number (Ar), Shannon diversity index, and the percentage of polymorphic loci were calculated using BIO-R (Biodiversity Analysis with R for Windows) version 3.2 (Pacheco-Gil *et al.,* 2016). Genetic relationships among populations were visualized using principal coordinate analysis (PCoA) in PAST version 3.0. A cluster analysis was performed based on Euclidean distance matrices using JMP® version 17.2.0. Genetic structure analysis was conducted using STRUCTURE version 2.3.4 (Pritchard *et al.,* 2000) under the following parameters: burn-in period of 10,000, Markov Chain Monte Carlo (MCMC) replications after burn-in of 250,000, Admixture model, K values ranging from 1 to 10, and 100,000 iterations. Delta K values were visualized using Structure Harvester version 0.6.94 (Earl & vonHoldt, 2012).

## Results and Discussion

### Genetic diversity

The genotyping of 14 Mexican moringa populations identified 9,862 SNPs. Based on these SNPs, genetic diversity statistics were calculated. He ranged from 0.18 (OAXH) to 0.43 (EMEX), with an average of 0.33, while Ho varied from 0.21 (OAXH) to 0.62 (EMEX), with an average of 0.44. The Ae number ranged from 1.23 to 1.74, averaging 1.51. The Shannon diversity index fluctuated between 0.21 and 0.62, averaging 0.43. Additionally, the Ar ranged from 67.41 to 167.79, with a

Revista Bio ciencias
ISSN 2007-3380

mean of 119.90, indicating a substantial presence of rare alleles (Table 2). Among the analyzed populations, OAXH exhibited the lowest diversity values (He = 0.18, Ho = 0.21), whereas EMEX had the highest diversity values (He = 0.43, Ho = 0.62). Overall, heterozygosity levels were found to be moderate to high, with OAXH representing the most homogeneous population and EMEX being the most genetically diverse.

**Table 2. Diversity parameters in *M. oleifera* populations cultivated in Mexico using SNP.**

| No. | Population | He | Ho | Ae | Shannon index | Ar | % NA |
|---|---|---|---|---|---|---|---|
| **1** | *CHI* | *0.42* | *0.60* | *1.72* | *0.60* | *163.33* | *0.00* |
| **2** | *EMEX* | *0.43* | *0.62* | *1.74* | *0.62* | *167.79* | *0.00* |
| **3** | *GTO* | *0.38* | *0.50* | *1.60* | *0.50* | *121.51* | *0.00* |
| **4** | *HGO* | *0.41* | *0.57* | *1.69* | *0.57* | *157.83* | *0.00* |
| **5** | *MICH1* | *0.37* | *0.49* | *1.58* | *0.49* | *119.01* | *0.00* |
| **6** | *MICHMP* | *0.38* | *0.51* | *1.61* | *0.51* | *120.89* | *0.00* |
| **7** | *NL* | *0.35* | *0.46* | *1.54* | *0.46* | *124.69* | *0.00* |
| **8** | *OAX* | *0.33* | *0.42* | *1.50* | *0.42* | *99.84* | *0.00* |
| **9** | *SLP* | *0.23* | *0.26* | *1.30* | *0.26* | *75.45* | *0.00* |
| **10** | *SIN* | *0.40* | *0.54* | *1.65* | *0.54* | *153.84* | *0.00* |
| **11** | *YUC1* | *0.26* | *0.31* | *1.36* | *0.31* | *88.39* | *0.00* |
| **12** | *YUC2* | *0.29* | *0.36* | *1.41* | *0.36* | *95.74* | *0.00* |
| **13** | *OAXH* | *0.18* | *0.21* | *1.23* | *0.21* | *67.41* | *0.00* |
| **14** | *GRO* | *0.24* | *0.28* | *1.32* | *0.28* | *80.92* | *0.00* |
| **Average** | | *0.33* | *0.44* | *1.52* | *0.44* | *116.90* | *0.00* |

He: Expected heterozygosity; Ho: Observed heterozygosity; Ae: Number of effective alleles; Shannon index: Shannon diversity index; Ar: Number of rare alleles; % NA: Proportion of missing data.

Source: Prepared by the authors based on the output of the BIO-R software version 3.2.

This study aimed to determine the extent of genetic diversity and relationships among 14 cultivated moringa populations in Mexico using SNP markers, thus, it represents the first report utilizing this type of polymorphic marker for this species at this country. A total of 15,191 SNPs were identified, of which 9,862 SNPs were used to assess genetic diversity and relationships among the moringa cultivated populations. The broad genomic distribution of these polymorphisms is attributed to the advantages of using the DArTseq platform (Gawroński *et al.,* 2016), which enables the evaluation of unexplored germplasm collections or agro-food-relevant cultivars.

At the population level, the highest observed He was 0.42, indicating moderate diversity, while the lowest He was 0.18, representing low diversity according to the scale proposed by

Revista Bio ciencias
ISSN 2007-3380

Botstein *et al.* (1980). One of the factors influencing heterozygosity within moringa populations is cross-pollination. Wu *et al.* (2018) reported that the self-fertilization rate in this species is 27%, and pollen dispersal can reach up to 24.7 meters, these characteristics ensure gene flow among individuals within a population. The heterozygosity values obtained in this study indicate that the 14 populations exhibit genetic diversity, which could be utilized for several purposes.

Regarding SNP locus diversity statistics, after data filtering, 65 % of the loci were found to be polymorphic, resulting in a mean He of 0.313, Ho of 0.438, Ae of 1.507, and a Shannon diversity index of 0.695 (Table 3).

**Table 3. Genetic diversity statistics of *M. oleifera* from 9,862 SNP-GBS markers detected in 14 cultivated populations of *M. oleifera*.**

| Parameter | Value | Standard deviation |
|---|---|---|
| *% of polymorphic loci* | *0.645* | |
| *Expected heterozygosity (He)* | *0.313* | *0.002* |
| *Observed heterozygosity (Ho)* | *0.438* | *0.003* |
| *Number of effective alleles (Ae)* | *1.507* | *0.004* |
| *Shannon diversity index* | *0.695* | *0.003* |

Source: Prepared by the authors based on the output of the BIO-R software version 3.2.

The polymorphic information content (PIC) of the SNPs was found to be 0.64 after data filtering. Gouda *et al.* (2021) reported a PIC of 0.23 when evaluating *Oryza* species using 158 SNPs, while Tomar *et al.* (2021) obtained an average PIC of 0.201 across 14,563 SNPs in *Triticum aestivum*. According to Dube *et al.* (2023), a high PIC value allows for an efficient assessment of genetic diversity in plant materials. Additionally, Muli *et al.* (2022) emphasized the importance of marker filtering to obtain a more objective determination of genetic diversity in plant populations. Teklemariam *et al.* (2022) reported that the total number of generated SNPs provides insights into population relationships, and data filtering is crucial for reducing dataset volume while retaining the most informative SNPs. Furthermore, the Shannon diversity index for SNPs in this study was 0.44, a value comparable to the 0.47 described for *Triticum aestivum* (Tomar *et al.,* 2021).

Genetic diversity has been reported as a key factor in enabling natural populations to better withstand the adverse effects of pests, diseases, and climate change (Zhang *et al.,* 2017). At the research level, knowledge of intra- and interpopulation genetic variation allows for efficient management of plant genetic resources that benefit farmers (Nadeem *et al.,* 2018), as well as improved planning for conservation and utilization of highly polymorphic ecotypes or those containing rare alleles at low frequencies. In this regard, Das & Mishra (2021) suggested that

Revista Bio ciencias
ISSN 2007-3380

genetic diversity could be applied in targeted crosses, hybridization experiments, heterotic group assignments, and germplasm conservation. Genetic diversity serves as a reservoir of alleles that influence plant responses to environmental stressors, resistance to biotic and abiotic stresses, and food production capacity (Begna, 2022). By evaluating and identifying the best allelic variants in moringa, agro-food systems with a higher degree of sustainability can be developed (Peter *et al.,* 2023). Moreover, utilizing genetic diversity in crop improvement programs could enhance the nutritional quality of food, ultimately improving human well-being.

Genetic diversity in moringa may explain variations in the nutritional content of the studied Mexican populations. Previous research has documented differences in mineral composition (Ndhlala & Tshabalala, 2023) and protein content ranging from 13.67 % to 23.50 % in moringa populations cultivated under identical edaphoclimatic and agronomic conditions (Ruiz-Hernández *et al.,* 2022b). Pasha *et al.* (2020) identified 36 genes involved in the secondary metabolites production. It would be of interest to examine potential associations between the identified SNPs and genes responsible for metabolite biosynthesis to develop cultivars with favorable alleles for increased metabolite concentrations. The number of SNPs identified in this study provides a foundation for further exploration of nutritionally significant traits, as moringa has 18,541 reported gene-coding sequences (Chang *et al.,* 2018). Populations exhibiting higher genetic diversity could serve as the basis for selecting individuals with superior traits, facilitating variety development while preserving the genetic reservoir in a germplasm bank.

The genetic diversity identified in these populations through SNP analysis may also support the evaluation of cultivars under temperate and cold climates to select materials with enhanced acclimatization capacity. Pasha *et al.* (2020) identified 2,326 transcription factors, of which 32 % were associated with responses to biotic and abiotic stress. Similarly, Zhang *et al.* (2019) reported the presence of WRKY proteins in moringa, which play roles in growth and development, mediate responses to various stress factors (biotic and abiotic), and regulate secondary metabolite production.

## Genetic relationships

The genetic relationships among the 14 populations characterized by SNP markers were visualized through various analytical approaches. The first method employed was hierarchical clustering based on a distance matrix using Ward's method. Molecular profiles revealed that the most genetically distant population was GRO, while the closest populations were GTO and MICHMP (Figure 2). This clustering analysis identified two major groups: Group I included the closest populations, with GTO and MICHMP exhibiting the shortest genetic distance, followed closely by MICH1; the OAX population also clustered within this group but was positioned at a greater distance, indicating that part of its genetic profile differs from the others. Group I similarly grouped EMEX and CHI as closely related, followed by SIN, NL, and HGO, the latter being the most genetically distant within this group and in the overall clustering analysis. Within Group II, YUC1 and YUC2 were identified as highly similar, followed by SLP, OAXH, and GRO, which were integrated into the group in that order based on increasing genetic distance.

Revista Bio ciencias
ISSN 2007-3380

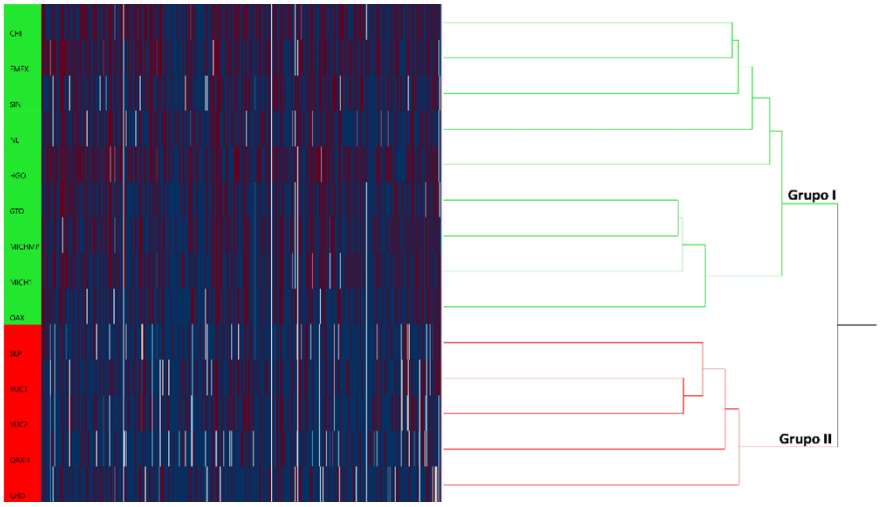


**Figure 2. Dendrogram of 14 *M. oleifera* populations cultivated in Mexico, molecularly characterized with 9,862 SNPs.**

The dendrogram shows a heat map composed of 9,862 SNPs where the color similarity corresponds to the same allele type. Blue color: 0 = AA; white color: 1 = aa; and red color: 2 = Aa.

Source: Own elaboration based on the output of the JMP software.

The visualization of relationships through principal coordinate analysis (PCoA) also detected the formation of two separate groups in the plane of the two main axes. In this case, axis 1 explained 18 % of the observed genetic variation, while axis 2 accounted for 11.33 %. The groups identified corresponded to those found in the hierarchical clustering analysis. The group structure was as follows: Group I included the populations MICH1, MICHMP, and GTO, which clustered closely together, while population OAX was integrated separately from the first three. The populations EMEX, HGO, NL, CHI, and SIN were also part of Group I but were more dispersed. These populations were distributed separately within the two-axis plane. Group II also included the populations OAXH, SLP, YUC1, YUC2, and GRO, which clustered closely together (Figure 3).

Revista Bio ciencias
ISSN 2007-3380

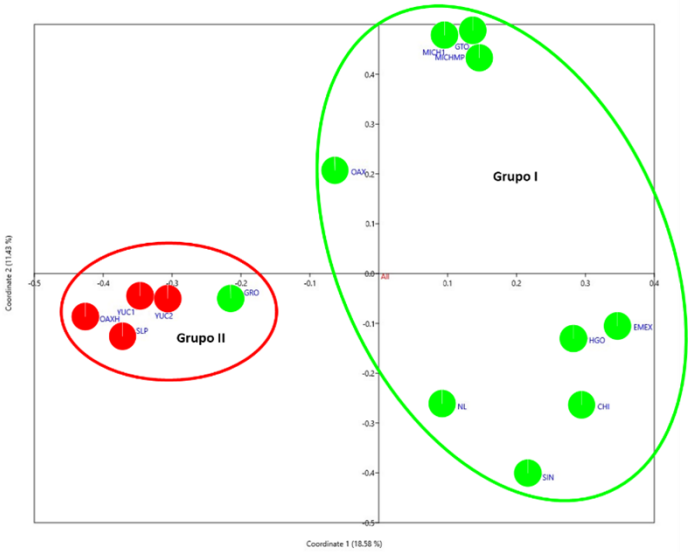


**Figure 3. Scatter plot of the first two PCoA coordinates of 14 *M. oleifera* populations cultivated in Mexico characterized with 9862 SNPs.**

The colors indicate the proportion shared by the populations of the original genetic groups (I: green and II: red) obtained with the STRUCTURE program.

Source: Own elaboration based on output of the PAST software.

The classification of the 14 moringa populations via cluster analysis and PCoA consistently identified two groups: MICH1, MICHMP, GTO, OAX, NL, SIN, CHI, HGO, and EMEX formed Group I, while OAXH, SLP, YUC1, YUC2, and GRO clustered into Group II. Both PCoA and clustering analysis demonstrated that the geographic distance between the evaluated populations does not indicate their genetic similarity levels. Moringa has experienced a commercial resurgence in recent years, leading to anthropogenic distribution and the reintroduction of new materials in Mexico. This study observed that the recent reintroduction of materials into Mexico has increased the genetic pool of moringa. Investigating genetic diversity in plant populations allows for the identification of genetic relationships, germplasm mixing, and the selection of progenitors based on genetic distance (Delfini *et al.,* 2021).

## Genetic structure

The Bayesian analysis identified K = 2 as the most probable number of original genetic groups, as the highest rate of change in the likelihood function concerning K (delta K) was obtained. Consequently, each population was assigned to one of the two genetic groups. Figure 4 presents the calculated Delta K values, computed using Structure Harvester (Evanno *et al.,* 2005). Since hierarchical clustering and PCoA grouped the populations into three clusters, Bayesian analysis was extended, confirming the presence of only two original genetic populations (ΔK = 2) among

Revista Bio ciencias
ISSN 2007-3380

the 14 evaluated moringa populations. Figure 5 illustrates the percentage of genetic contribution shared among the 14 moringa populations based on K = 2.

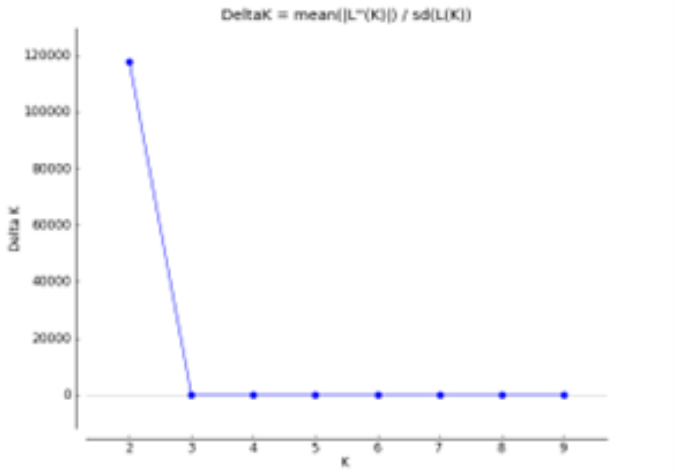


**Figure 4. Delta K values for the STRUCTURE analysis of *M. oleifera* populations cultivated in Mexico.**

Source: Prepared by the authors based on the results of Structure Harvester version 0.6.94.

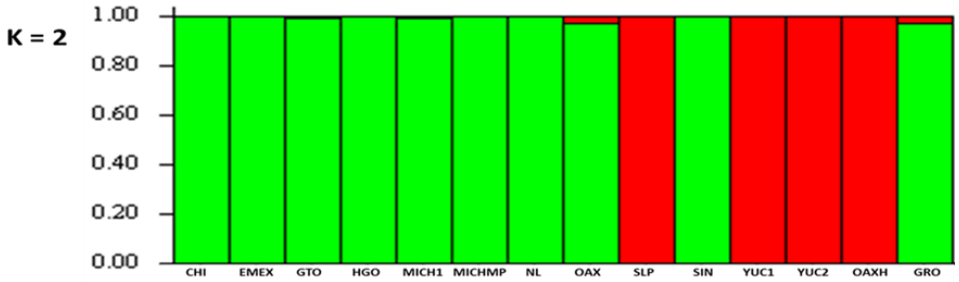


**Figure 5. Barplot clustering of 14 *M. oleifera* populations cultivated in Mexico characterized with 9,862 SNPs according to STRUCTURE.**

The distribution of two original populations (K = 2) is observed. The numbers on the axis indicate the ancestry coefficient.

Source: Own elaboration based on results obtained from STRUCTURE version 2.3.4.

The analysis of genetic ancestry distribution using STRUCTURE allowed for identifying the shared proportion among moringa populations. Populations with scores equal to or greater than 0.800 are considered to have a high degree of purity, while those with values below 0.800 are

Revista Bio ciencias
ISSN 2007-3380

classified as heterozygous. Table 4 presents the proportion of the original genetic group shared by the Moringa populations collected from cultivated fields in Mexico.

**Table 4. Estimated proportion of ancestry of the 14 populations of *M. oleifera* Lam.**

| No. | Population | Genetic group | |
|---|---|---|---|
| | | I | II |
| *1* | *CHI* | *1.000* | *0.000* |
| *2* | *EMEX* | *1.000* | *0.000* |
| *3* | *GTO* | *1.000* | *0.000* |
| *4* | *HGO* | *1.000* | *0.000* |
| *5* | *MICH1* | *1.000* | *0.000* |
| *6* | *MICHMP* | *1.000* | *0.000* |
| *7* | *NL* | *1.000* | *0.000* |
| *8* | *OAX* | *0.994* | *0.006* |
| *9* | *SLP* | *0.000* | *1.000* |
| *10* | *SIN* | *1.000* | *0.000* |
| *11* | *YUC1* | *0.000* | *1.000* |
| *12* | *YUC2* | *0.003* | *0.997* |
| *13* | *OAXH* | *0.000* | *1.000* |
| *14* | *GRO* | *0.994* | *0.006* |

Source: Self-made.

The genetic structure analysis revealed two original genetic populations (K = 2) and showed that the populations in Group I from the PCoA and clustering analysis share the same original genetic ancestry. It is known that moringa materials arrived in Mexico via the voyages of the Nao de China in the 1500s (Olson & Fahey, 2011), and at least among the 14 populations examined, their origin appears to trace back to the same ecotypes or their propagation. Given that moringa populations have been distributed through human activity, a low level of genetic diversity was expected. Liu *et al*. (2023) mention that geographic distance between populations increases genetic differentiation among collected samples. However, this was not the case for the examined moringa populations.

Nonetheless, while the populations in Group I share the same original genetic contribution, it is noteworthy that population OAX exhibits a minimal contribution from original population II, adding genetic variation. In Group II, a genetic contribution from original population I was observed in population GRO. The genetic diversity of original populations I and II should serve as the foundation for conservation and targeted management plans. Zhang *et al.* (2017) state

Revista Bio ciencias
ISSN 2007-3380

that the genetic structure of populations is influenced by domestication and selection processes. Another factor affecting gene flow is migration (Muli *et al.,* 2022). However, this does not appear to be the case for Mexican moringa populations. These populations have yet to undergo selection and breeding processes, and understanding their genetic diversity and structure provides insights into which populations could be candidates for improvement. The exchange of germplasm or the creation of populations with two contrasting genetic materials increases genetic diversity and the number of polymorphic alleles. However, it limits the formation of well-defined groups in the genetic structure (Raatz *et al.,* 2019). Genetic diversity in populations used for food production is influenced by human management, and the selection of individuals leads to gene loss. Understanding genetic structure helps identify distribution patterns in plant populations (Cheng *et al.,* 2020).

## Conclusions

The 14 moringa populations cultivated in various states of Mexico exhibit genetic diversity. The SNP markers used for this analysis revealed genetic diversity ranging from low to moderate, with the latter being predominant. It was also identified that this diversity is structured into different genotypes, as the Ho was higher than the He. These populations formed two genetically related groups, within which some populations were more genetically distant. These groupings aligned with the number of original ancestral populations. The evaluated moringa populations hold the potential for developing new genetic combinations that could enhance traits of productive and nutritional interest, as well as improve tolerance to environmental and biological factors.

## Author contributions

Work conceptualization: Ruiz-Hernández, R. and Hernández-Rodríguez, M.; methodology development: Ruiz-Hernández, R., Hernández-Rodríguez, M., and Pérez-Vázquez, A.; software handling: Ruiz-Hernández, R. and Hernández-Rodríguez, M.; experimental validation: Ruiz-Hernández, R. and Hernández-Rodríguez, M.; data analysis: Ruiz-Hernández, R. and Hernández-Rodríguez, M.; data management: Ruiz-Hernández, R. and Hernández-Rodríguez, M.; manuscript writing and preparation: Ruiz Hernández, R. and Hernández-Rodríguez, M.; drafting, reviewing, and editing: Ruiz-Hernández, R., Hernández-Rodríguez, M., and Pérez-Vázquez, A.; project administration: Rayas-Amor, A.A.; funding acquisition: Rayas-Amor, A.A.

All authors of this manuscript have read and approved the published version.

## Funding

This research was self-funded.

Revista Bio ciencias
ISSN 2007-3380

## Acknowledgments

The authors thank the Colegio de Postgraduados, Campus Veracruz, and Campus Montecillo, for their support in conducting this study.

## Conflict of interest

The authors declare no conflict of interest.

## References

Begna, T. (2022). Role of plant genetic diversity conservation and utilization. *International Journal* of *Novel Research in Life Sciences*, 9(1), 1-11. https://www.noveltyjournals.com/upload/paper/ROLE%20OF%20PLANT%20GENETIC-01012022-7.pdf

Botstein, D., White., R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *The American Journal of Human Genetics*, 32(3), 314-331.

Calicioglu, O., Flammini, A., Bracco, S., Bellù, L., & Sims, R. (2019). The future challenges of food and agriculture: an integrated analysis of trends and solutions. *Sustainability*, 11(1), 222. https://doi.org/10.3390/su11010222

Chang, Y., Liu, H., Liu, M., Liao, X., Sahu, S. K., Fu, Y., Song, B., Cheng, S., Kariba, R., Muthemba, S., Hendre, P. S., Mayes, S., Ho, W. K., Yssel, A. E. J., Kendabie, P., Wang, S., Li, L., Muchugi, A., Jamnadass, R., Lu, H., Peng, S., Van Deynze, A., Simons, A., Yana-Shapiro, H., Van de Peer, Y., Xu, X., Yang, H., Wang, J., & Liu, X. (2018). The draft genomes of five agriculturally important African orphan crops. *GigaScience*, 8(3), 1-16. https://doi.org/10.1093/gigascience/giy152

Cheng, J., Kao, H., & Dong, S. (2020). Population genetic structure and gene flow of rare and endangered *Tetraena mongolica* Maxim. revealed by reduced representation sequencing. *BMC Plant Biology*, 20(1), Article 391. https://doi.org/10.1186/s12870-020-02594-y

Centro Internacional de Mejoramiento de Maíz y Trigo [CIMMYT]. (2006). Protocolos de laboratorio: Laboratorio de Genética Molecular Aplicada del CIMMYT. (3a ed.). México, D.F. https://repository.cimmyt.org/bitstream/handle/10883/593/91224.PDF?sequence=4&isAllowed=y

Das, D., & Mishra, R. (2021). SNP markers and its impact on crop improvement. *Asian Journal of Biological and Life Sciences*, 10(3), 539-548. https://doi.org/10.5530/ajbls.2021.10.72

Delfini, J., Moda-Cirino, V., dos Santos Neto, J., Ruas, P.M., Sant'Ana, G.C., Gepts, P., & Gonçalves, L. S. A. (2021). Population structure, genetic diversity and genomic selection signatures among a Brazilian common bean germplasm. *Scientific Reports*, 11(1), 2964. https://doi.org/10.1038/s41598-021-82437-4

Dube, S. P., Sibiya, J., & Kutu, F. (2023). Genetic diversity and population structure of maize inbred lines using phenotypic traits and single nucleotide polymorphism (SNP) markers. *Scientific Reports*, 13(1), 17851. https://doi.org/10.1038/s41598-023-44961-3

Earl, D. A., & vonHoldt, B. M. (2012). Structure harvester: a website and program for visualizing

STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Elbasyoni, I. S., Eltaher, S., Morsy, S., Mashaheet, A. M., Abdallah, A. M., Ali, H. G., Mariey, S. A., Baenziger, P. S., & Frels, K. (2022). Novel single-nucleotide variants for morpho-physiological traits involved in enhancing drought stress tolerance in barley. *Plants*, 11(22), Article 3072. https://doi.org/10.3390/plants11223072

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, 14(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Gawroński, P., Pawełkowicz, M., Tofil, K., Uszyński, G., Sharifova, S., Ahluwalia, S., Tyrka, M., Wędzony, M., Kilian, A., & Bolibok-Brągoszewska, H. (2016). DArT markers effectively target gene space in the rye genome. *Frontiers in Plant Science*, 7(6), 1600. https://doi.org/10.3389/fpls.2016.01600

Gouda, A. C., Warburton, M. L., Djedatin, G. L., Kpeki, S. B., Wambugu, P. W., Gnikoua, K., & Ndjiondjop, M. N. (2021). Development and validation of diagnostic SNP markers for quality control genotyping in a collection of four rice (*Oryza*) species. *Scientific Reports*, 11(1), 18617. https://doi.org/10.1038/s41598-021-97689-3

Hendre, P. S. & Van Deynze, A. (2015). How a consortium is changing Africa's food systems? http://www.africanorphancrops.org/category/resources/blog/page/3/

Hendre, P. S., Muthemba, S., Kariba, R., Muchugi, A., Fu, Y., Chang, Y., Song, B., Liu, H., Liu, M., Liao, X., Sahu, S. K., Wang, S., Li, L., Lu, H., Peng, S., Cheng, S., Xu, X., Yang, H., Wang, J., Liu, X., Simons, A., Shapiro, H. Y., Mumm, R. H., Van Deynze, A., & Jamnadass, R. (2019). African Orphan Crops Consortium (AOCC): status of developing genomic resources for African orphan crops. *Planta*, 250(3), 989-1003. https://doi.org/10.1007/s00425-019-03156-9

Huster, A. R., Wallace, L. T., & Myers, J. R. (2021). Associated SNPs, Heritabilities, trait correlations, and genomic breeding values for resistance in snap beans (*Phaseolus vulgaris* L.) to root rot caused by *Fusarium solani* (Mart.) f. sp. *phaseoli* (Burkholder*). Frontiers in Plant Science*, 12(2021), 697615. https://doi.org/10.3389/fpls.2021.697615

Liu, H., Wang, Z., Zhang, Y., Li, M., Wang, T., & Su, Y. (2023). Geographic isolation and environmental heterogeneity contribute to genetic differentiation in *Cephalotaxus oliveri*. *Ecology and Evolution*, 13(3), e9869. https://doi.org/10.1002/ece3.9869

Muhl, Q. E., Toit, E. S. D., & Robbertse, P. J. (2011). *Moringa oleifera* (Horseradish Tree) leaf adaptation to temperature regimes. *International Journal of Agriculture and Biology*, 13(6), 10-21. https://www.up.ac.za/media/shared/Legacy/sitefiles/file/48/2056/31fijab1112510211024.pdf

Muli, J. K., Neondo, J. O., Kamau, P. K., Michuki, G. N., Odari, E., & Budambula, N. L. M. (2022). Genetic diversity and population structure of wild and cultivated *Crotalaria* species based on genotyping-by-sequencing. *PLoS ONE*, 17(9), e0272955. https://doi.org/10.1371/journal.pone.0272955

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ahmad, F., Alsaleh, A., Labhane, N., Özkan, H., Chung, G., & Baloch, F. S. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment*, 32(2), 261–285. https://doi.org/10.1080/13102818.2017.1400401

Ndhlala, A. R., & Tshabalala, T. (2023). Diversity in the nutritional values of some *Moringa oleifera* Lam. cultivars. *Diversity*, 15, 834. https://doi.org/10.3390/d15070834

Olson, M. E., & Fahey, J. W. (2011). *Moringa oleifera*: un árbol multiusos para las zonas tropicales

seca. *Revista Mexicana de Biodiversidad*, 82(4), 1071-1082. https://doi.org/10.22201/ib.20078706e.2011.4.678

Ousmael, K., Whetten, R. W., Xu, Nielsen, U. B., Lamour, K., & Hansen O. K. (2023). Identification and high-throughput genotyping of single nucleotide polymorphism markers in a non-model conifer (*Abies nordmanniana* (Steven) Spach). *Scientific Reports,* 13(2023), 22488. https://doi.org/10.1038/s41598-023-49462-x

Pacheco-Gil, R. Á., Alvarado; G. Rodríguez; F., & Burgueño, J. (2016). BIO-R (Biodiversity analysis with R for Windows) Version 3.3. CIMMYT Research Data & Software Repository Network, 14. https://hdl.handle.net/11529/10820

Pasha, S. N., Shafi, K. M., Joshi, A. G., Meenakshi, I., Harini, K., Mahita, J., Sajeevan, R. S., Karpe, S. D., Ghosh, P., Nitish, S., Gandhimathi, A., Mathew, O. K., Prasanna, S. H., Malini, M., Mutt, E., Naika, M., Ravooru, N., Rao, R. M., Shingate, P. N., Sukhwal, A., Sunitha, M. S., Upadhyay, A. K., Vinekar, R. S., & Sowdhamini, R. (2020). The transcriptome enables the identification of candidate genes behind medicinal value of drumstick tree (*Moringa oleifera*). *Genomics*, 112(1), 621–628. https://doi.org/10.1016/j.ygeno.2019.04.014

Peter A. F., Wagiran, A., Rahmat Z., Yusop M. R., & Ridzuan, R. (2023). Exploring genetic variation and therapeutic properties of *Moringa oleifera*: progress and future potential for crop improvements. *Pharmacognosy Reviews*, 17(34), 426-438. https://doi.org/10.5530/phrev.2023.17.18

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

Raatz, B., Mukankusi, C., Lobaton, J. D., Male, A., Chisale, V., Amsalu, B., Fourie, D., Mukamuhirwa, F., Muimui, K., Mutari, B., Nchimbi-Msolla, S., Nkalubo, S., Tumsa, K., Chirwa, R., Maredia, M. K., & He, C. (2019). Analyses of African common bean (*Phaseolus vulgaris* L.) germplasm using a SNP fingerprinting platform: diversity, quality control and molecular breeding. *Genetic Resources and Crop Evolution*, 66(3), 707–722. https://doi.org/10.1007/s10722-019-00746-0

Ruiz-Hernández, R., Hernandez-Rodriguez, M., Cruz-Monterrosa, R. G., Diaz-Ramirez, M., Martinez-García, C. G., Garcia-Martinez, A., & Amor, A. A. R. (2022a). *Moringa oleifera* Lam.: una revisión de los factores ambientales y de manejo que influyen en el contenido nutricional de las hojas. *Tropical and Subtropical Agroecosystems,* 25(1), 1-15. https://doi.org/10.56369/tsaes.4053

Ruiz-Hernández, R., Hernández-Rodríguez, M., Pérez-Vázquez, A., Cruz-Monterrosa, R. G., & Rayas-Amor, A. A., (2022b). Chemical composition of twelve accessions of *Moringa oleifera* Lam. grown in Mexico. *Agro Productividad*. 15(11), 11-20. https://doi.org/10.32854/agrop.v15i10.2405

Ruiz-Hernández, R., Pérez-Vázquez, A., García-Pérez, E., Morales-Trejo, F., & Soto-Hernández, R. M. (2021). Caracterización morfológica de accesiones de *Moringa oleifera* provenientes del sur-sureste de México. *Revista Mexicana De Ciencias Agrícolas*, 12(7), 1210–2022. https://doi.org/10.29312/remexca.v12i7.2632

Ruiz-Hernández, R., Pérez-Vázquez, A., Hernández-Rodríguez, M., García-Pérez, E., Morales-Trejo, F., & Soto-Hernández, R. M. (2024). Genetic diversity of *Moringa oleifera* Lam. in southeastern Mexico. *Agrociencia*, 58(6), 1-17. https://doi.org/10.47163/agrociencia.v58i6.3101

Shyamli, P. S., Pradhan, S., Panda, M., & Parida, A. (2021). *De novo* whole-genome assembly of *Moringa oleifera* helps identify genes regulating drought stress tolerance. *Frontiers in Plant Science*, 12(1), 766999. https://doi.org/10.3389/fpls.2021.766999

Teklemariam, S. S., Bayissa, K. N., Matros, A., Pillen, K., Ordon, F., & Wehner, G. (2022). The genetic diversity of Ethiopian barley genotypes in relation to their geographical origin. *PLoS ONE*, 17(5),

e0260422. https://doi.org/10.1371/journal.pone.0260422

Tian, Y., Zeng, Y., Zhang, J., Yang, C., Yan, L., Wang, X., Shi, C., Xie, J., Dai, T., Peng, L., Zeng Huan, Y., Xu, A., Huang, Y., Zhang, J., Ma, X., Dong, Y., Hao, S., & Sheng, J. (2015). High quality reference genome of drumstick tree (*Moringa oleifera* Lam.), a potential perennial crop. *Science China Life Sciences*, 58(7), 627–638. https://doi.org/10.1007/s11427-015-4872-x

Tomar, V., Dhillon, G. S., Singh, D., Singh, R. P., Poland, J., Joshi, A. K., Tiwari B. S., & Kumar, U. (2021). Elucidating SNP-based genetic diversity and population structure of advanced breeding lines of bread wheat (*Triticum aestivum* L.). *PeerJ*, 9, e11593. http://doi.org/10.7717/peerj.11593

Wu, J. C., Zhang, Y. P., Zheng, Y. X., & Peng, X. M. (2018). Pollen mediated gene flow in a small experimental population of *Moringa oleifera* Lam. (Moringaceae). *Industrial Crops and Products*, 117(1), 28-33. https://doi.org/10.1016/j.indcrop.2018.02.077

Yirgu, M., Kebede, M., Feyissa, T., Lakew, B., Woldeyohannes, A. B. & Fikere, M. (2023). Single nucleotide polymorphism (SNP) markers for genetic diversity and population structure study in Ethiopian barley (*Hordeum vulgare* L.) germplasm. *BMC Genomic Data*, 24(1), 7. https://doi.org/10.1186/s12863-023-01109-6

Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., & Song, B. H. (2017). Back into the wild-apply untapped genetic diversity of wild relatives for crop improvement. *Evolutionary Applications*, 10(1), 5-24. https://doi.org/10.1111/eva.12434

Zhang, J., Yang, E., He, Q., Lin, M., Zhou, W., Pian, R., & Chen, X. (2019). Genome-wide analysis of the WRKY gene family in drumstick (*Moringa oleifera* Lam.). *PeerJ*, 7(1), e7063. https://doi.org/10.7717/peerj.7063