12-15-2021

# Genetic diversity of Huaya India (*Melicoccus oliviformis* Kunth), a neglected Neotropical fruit crop

Mónica I. Jiménez-Rojas
*Tecnológico Nacional de México*, monica.jimenez@itconkal.edu.mx

Rubén H. Andueza-Noh
*Tecnológico Nacional de México*, r_andueza81@hotmail.com

Obed I. Noh-Ake
*Tecnológico Nacional de México*, generral_keiki@hotmail.com

Daniel Potter
*University of California, Davis*, dpotter@ucdavis.edu

Matilde M. Ortiz-García
*Centro de Investigacion Cientifica de Yucatan*, arimat@cicy.mx

*See next page for additional authors*

Scientia Horticulturae 290 (2021) 110535

Contents lists available at ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

# Genetic diversity of Huaya India (*Melicoccus oliviformis* Kunth), a neglected Neotropical fruit crop

Mónica I. Jiménez-Rojas [a], Rubén H. Andueza-Noh [a], Obed I. Noh-Ake [a], Daniel Potter [c], Matilde M. Ortiz-García [b], Renee S. Arias [d], Jaime Martínez-Castillo [b,*]

[a] Tecnológico Nacional de México/I.T. Conkal, Av. Tecnológico S/N, 97345, Conkal, Yucatán, México
[b] Centro de Investigación Científica de Yucatán (CICY), Calle 43 No. 130. Col. Chuburná de Hidalgo, Mérida, Yucatán, México CP 97200
[c] Department of Plant Sciences, University of California, Davis, One Shields Avenue, Davis, California, 95616, USA
[d] USDA-ARS-National Peanut Research Laboratory, 1011 Forrester Dr., SE, Dawson, GA 39842-0509, USA

ARTICLE INFO

*Keywords:*
Domestication
Genetic structure
Maya Lowlands of Mexico
Microsatellite molecular markers
Perennial fruit species
Sapindaceae
SSR markers
Yucatan Peninsula

ABSTRACT

Currently, some species of Sapindaceae are important fruit crops worldwide. The Huaya India (*Meliccocus oliviformis*, Sapindaceae) is a neglected Neotropical fruit tree consumed locally in the Maya Lowlands of Mexico, where it exists in both wild and domesticated forms. Our objective was to evaluate the genetic diversity of the Huaya India in its possible domestication area and thus generate knowledge that serves as the basis for a commercial management. A total of 450 individuals collected from 15 natural vegetation sites and 15 Maya villages, were characterized using nine microsatellite loci and population genetics approaches were applied. STRUCTURE, Neighbor-Joining and PCoA analyses suggested the existence of three main groups: 1) one composed by 14 natural vegetation sites, 2) one integrated by 10 Maya villages plus one natural vegetation site, 3) one composed by five Maya villages. At the species level, genetic differentiation was high ($F_{ST} = 0.562$) and gene flow was low (Nm = 0.395); between genetic groups, differentiation was low and gene flow was high. Genetic diversity was low at the level species ($H_E = 0.19$) and higher in the group composed for only natural vegetation sites. When we considered only two groups (natural vegetation sites vs Maya villages) to explore a possible bottleneck as a consequence of human management, the natural vegetation sites showed higher, and significant, genetic diversity ($H_E = 0.231$) than the Maya villages ($H_E = 0.152$). This study can serve as a basis to develop management strategies for Huaya India in the Maya Lowlands of Mexico, but without compromising its conservation.

## 1. Introduction

Tropical regions are home to a large number of perennial fruit tree species; however, only a few, such as banana (*Musa paradisiaca* L.), mango (*Mangifera indica* L.), papaya (*Carica papaya* L.) and pineapple (*Ananas comosus* L.), are well-known crops worldwide (Paull and Duarte, 2012). This is largely due to the fact that more than 90% of the tropical fruits are consumed locally (FAO, 2010). In international markets, minor tropical fruits are still regarded as a novelty or niche product, though market opportunities have been developing rapidly in China on the back of income growth and urbanization, and demand is also on an upward trajectory in other key markets such as the United States and European Union, mainly in response to increasing health awareness and changing dietary preferences (Altendorf, 2018). Several tropical fruits considered as neglected until a few decades ago are now important in international markets. These include several Asian species of Sapindaceae family such as lychee (*Litchi chinensis* L.), rambutam (*Nephelium lappaceum* L.) and longan (*Dimocarpus longan* Lour.) (Altendorf, 2018). However, most tropical fruit species can be considered neglected, with little is known about their genetic diversity. Such information is essential to prevent inappropriate management and exploitation from compromising their conservation, particularly in their centers of origin and domestication (Martínez-Castillo et al., 2019b).

*Melicoccus* (Sapindaceae) is a small genus native to the Neotropics that includes 10 species of fruit trees, with a geographical distribution extending from the Yucatan peninsula in Mexico to South America, with

* Corresponding author.
*E-mail addresses:* monica.jimenez@itconkal.edu.mx (M.I. Jiménez-Rojas), r_andueza81@hotmail.com (R.H. Andueza-Noh), generral_keiki@hotmail.com (O.I. Noh-Ake), dpotter@ucdavis.edu (D. Potter), arimat@cicy.mx (M.M. Ortiz-García), renee.arias@usda.gov (R.S. Arias), jmartinez@cicy.mx (J. Martínez-Castillo).

https://doi.org/10.1016/j.scienta.2021.110535
Received 12 March 2021; Received in revised form 8 July 2021; Accepted 18 August 2021
Available online 27 August 2021

the latter region considered its center of diversity (Acevedo-Rodríguez, 2003). The two most consumed species in the genus are the Spanish Lime (*M. bigugatus* Jacq.) and the Huaya India (*M. oliviformis* Kunth) (Martínez-Castillo et al., 2019a). Huaya India presents a disjunct geographical distribution, with wild populations present in northern South America (Colombia, Venezuela, and Peru) as well as in the Yucatan peninsula, Mexico, but absent in Central America and the Caribbean (Acevedo-Rodríguez, 2003). In the Yucatan peninsula, which is part of the geographic-cultural area known as the Maya Lowlands, Huaya India trees are found in sub-evergreen forest and derived secondary vegetation (Carnevali et al., 2010). Furthermore, trees of this species are a common component of the backyards of the Maya villages, whose inhabitants have consumed their fruits for 2000 years (Colunga-GarcíaMarín and Zizumbo-Villarreal, 2004). Currently, the fruits are sold in local and regional markets, resulting in an economic income for Maya families, who reported that Huaya India trees do not require many inputs and agricultural care and are resistant to drought, making the species a good candidate for cultivation and integration into a commercial production system (Jiménez-Rojas et al., 2019).

The existence of wild populations of Huaya India only in the north of South America, as well as in the Yucatan peninsula, suggests that it is possibly native to these places. It is also possible however, that the Maya people who have occupied the Yucatan peninsula for many centuries may have introduced the species from northern South America where it is also native (Acevedo-Rodríguez, 2003), a hypothesis that is supported by the fact that the latter region is the center of diversity of the genus *Melicoccus*. To date, there are no published studies of the reproduction system of the Huaya India; however, *Melicoccus* species are propagated by seed and are considered dioecious, with individual trees bearing either staminate or pistillate flowers (Acevedo-Rodríguez, 2003). This last characteristic allows the Maya people to discriminate between male and female trees (Jiménez-Rojas et al., 2018).

The presence of Huaya India trees in both natural areas and within Maya villages and the long history of use of the species in the region suggest that the Maya Lowlands of Mexico could be an area of genetic diversity and possible center of domestication of this species. Previously, we reported finding significant morphological variation in the characteristics of Huaya India fruits collected from eight Maya villages from four geographical-cultural zones of the Yucatan Peninsula (Jiménez-Rojas et al., 2018); also, we reported the existence of limited yet significant differences between the fruit characteristics of Huaya India trees collected from areas of natural vegetation and those collected from Maya villages (Jiménez-Rojas et al., 2019). These observations suggest that this species is in an incipient stage of domestication (as defined by Clement 1989), resulting from thousands of years of conscious and/or unconscious management and selection.

Microsatellites (also known as SSR, Simple Sequence Repeats) are among the most commonly molecular markers used in genetic population studies due to their reproducibility, transferability between closely related species, and multi-allelic, co-dominant nature (Vieira et al., 2016). Currently, no specific microsatellite markers exist for Huaya India; however, SSRs for the closely related species Spanish Lime were recently developed by Martínez-Castillo et al. (2019a). These authors evaluated 31 polymorphic loci in 25 Spanish Lime trees collected from the state of Yucatan, Mexico, and found relatively low levels of genetic diversity ($H_E$ = 0.38). In the same work, though transferability and polymorphism were observed on SSR markers tested on DNA of five Huaya India trees, the sample size was too small to estimate the genetic diversity of this species.

In order to contribute to the knowledge of the genetic diversity and domestication of Huaya India in the Maya Lowlands of Mexico, we evaluated the genetic diversity and structure of 450 trees of this species collected from natural areas and from Maya villages of the Yucatan peninsula, Mexico, using nine microsatellite loci previously reported as polymorphic for Huaya India (Martínez-Castillo et al., 2019a). The knowledge generated in this study, together with the ethnobotanical and morphological information previously reported, will help lay the foundations for future commercial management of the Huaya India, taking advantage of the market niches already opened by Asian species of Sapindaceae, without compromising the conservation of the species.

## 2. Materials and methods

### 2.1. Study area

This study was carried out in the Mexican part of the Yucatan peninsula, which includes the States of Campeche, Yucatan and Quintana Roo, a region that is part of the Maya Lowlands of Mexico. Plant material was collected from 450 trees of Huaya India in 30 sites: 225 trees in 15 Maya villages and 225 trees in 15 natural vegetation sites, with an average of 15 trees per site (Table 1). The selected trees in natural areas were distributed throughout each collection site and the trees from Maya villages were collected in backyards belonging to different families, with the goal of obtaining the greatest possible genetic diversity. We did not classify trees as cultivated/domesticated or wild a priori; we only divided them according to the collection site (natural vegetation sites vs Maya villages). We adopted this approach in order to avoid misclassification of the material, since the species is a long-lived tree and do not know the recent history of the trees in many of the collection sites, including the Maya villages and backyards where half of the total trees were collected. Fresh leaves were collected from each tree, which were stored in paper bags and transferred to the Molecular Markers laboratory of the Centro de Investigación Científica de Yucatán, located in Mérida, Yucatan, Mexico.

### 2.2. DNA extraction and microsatellite technique

DNA extraction was performed using the QIAGEN DNeasy Plant Mini Kit, following the manufacturer's instructions. The quality of the DNA was verified on 1% agarose gels stained with 10 mg/ml ethidium bromide (EtBr), in a 0.5 X Tris-Borate-EDTA (TBE) buffer solution. Electrophoresis was carried out in a horizontal Life Technologies – Horizon 11-14 equipment with an EC-105 power source, at a load of 100 volts for 30 minutes. Agarose gels were observed on a UV light transilluminator (Transilluminator UV, Dinco & Rhenium Industries).

Initially, all 450 individuals were genotyped using 12 nuclear SSR loci developed for *M. bijugatus* (Martínez-Castillo et al., 2019b). PCR amplifications were performed using the polymerase chain reaction technique, in a GeneAmp PCR System 9700 thermal cycler (Applied Biosystems, Foster City, USA). The PCR program consisted of 35 cycles, each including a denaturation step of 2 minutes at 94°C, an annealing step of 1 minute at temperature depending on the primers used, and an extension step of 5 minutes at 72°C, with an additional final extension of 5 min at 72°C. A volume of 5 μL of formamide containing 0.45% bromophenol blue and 0.25% xylene cyanol was added to the PCR product and was denatured for a period of 5 min at 94°C, then 5 μL of this reaction product were loaded on 5% polyacrylamide gels (19:1 acrylamide:bisacrylamide) containing 5 mol/L urea and 0.5 × TBE buffer. Electrophoresis was performed at 60 W constant power for 45 min for 1 h using an SQ3 sequencer (Hoeffer Scientific Instruments, San Francisco, California, USA). The amplification products were visualized with the silver staining technique (Bassam et al., 1991), the sizes of the amplified fragments were determined visually in base pairs (bp) using a 10 bp molecular marker as a reference (Invitrogen, Life Technologies, Brazil). Finally, we compared the sizes of the alleles found in the individuals of each of the 30 collected sites to ensure a correct reading of the data. After this last step, we decided to use for the final analysis only the data obtained from the nine loci that generated the best reads among the fragments found in each individual (Table 2).

**Table 1**
Data of the 30 populations of Huaya India (*Melicocus oliviformis* Kunth) collected in the Yucatan Peninsula, Mexico, for in this study.

| Population | Category | State | Latitude | Longitude |
|---|---|---|---|---|
| Laguna Azul | Natural vegetation | Quintana Roo | 19° 50.294 | -88° 04.669 |
| Laguna Ocom | Natural vegetation | Quintana Roo | 19° 28.630 | -88° 03.303 |
| Mujer Escondida | Natural vegetation | Campeche | 19° 59.476 | -89° 45.864 |
| Miguel Colorado | Natural vegetation | Campeche | 18° 48.727 | -90° 44.853 |
| Lol-Tún | Natural vegetation | Yucatán | 20° 14.336 | -89° 27.663 |
| Noh-Bec | Natural vegetation | Yucatán | 20° 03.763 | -89° 06.601 |
| Chichen Itzá | Natural vegetation | Yucatán | 20° 41.093 | -88° 48.945 |
| Tixcacal Cupul | Natural vegetation | Yucatán | 20° 31.900 | -88° 16.738 |
| Cobá | Natural vegetation | Quintana Roo | 20° 28.514 | -87° 44.238 |
| Bacalar | Natural vegetation | Quintana Roo | 18° 53.403 | -88° 14.217 |
| Champotón | Natural vegetation | Campeche | 19° 15.067 | -90° 40.671 |
| Calakmul | Natural vegetation | Campeche | 18° 30.703 | -89° 54.034 |
| Tulum | Natural vegetation | Quintana Roo | 19° 02.784 | -88° 12.643 |
| Temozón Norte | Natural vegetation | Yucatán | 21° 08.817 | -88° 12.921 |
| Edzná | Natural vegetation | Campeche | 19° 48.144 | -89° 49.419 |
| X-Hazil | Maya village | Quintana Roo | 19° 28.630 | -88° 03.303 |
| Señor | Maya village | Quintana Roo | 19° 50.294 | -88° 04.669 |
| Calderitas | Maya village | Quintana Roo | 19° 02.784 | -88° 12.643 |
| Chumpón | Maya village | Quintana Roo | 20° 42.093 | -88° 48.645 |
| Cobá Pueblo | Maya village | Quintana Roo | 19° 46.986 | -87° 53.468 |
| Yotholin | Maya village | Yucatán | 20° 14.336 | -89° 27.663 |
| Becanchén | Maya village | Yucatán | 20° 03.763 | -89° 06.601 |
| Libre Unión | Maya village | Yucatán | 20° 42.093 | -88° 48.645 |
| Calotmul | Maya village | Yucatán | 21° 00.817 | -88° 10.749 |
| Xocén | Maya village | Yucatán | 20° 31.900 | -88° 16.738 |
| X-Pujil | Maya village | Campeche | 18° 30.020 | -89° 23.933 |
| Pueblo Nuevo | Maya village | Campeche | 19° 48.144 | -89° 49.419 |
| Vicente Guerrero | Maya village | Campeche | 19° 15.067 | -90° 40.671 |
| Bolonchén | Maya village | Campeche | 19° 59.476 | -89° 45.864 |
| Escárcega | Maya village | Campeche | 18° 48.727 | -90° 44.853 |

### 2.3. Data analysis

#### 2.3.1. Confidence of data

To check confidence in the accuracy of genotyping, SSR data were analyzed with the MICRO-CHECKER program v.2.2.3 (Van Oosterhout et al., 2004) to determine the existence of null alleles (alleles that fail to amplify during PCR), stuttering (slight changes that occur in the allele size due to errors during PCR), and dropout alleles (large alleles do not amplify as efficiently as small alleles).

#### 2.3.2. Clustering pattern and genetic structure

To determine how the genetic diversity of Huaya India is distributed in the Yucatan peninsula, data were analyzed using three clustering methods. 1) An individual assignment test was implemented using the STRUCTURE program (Pritchard et al., 2000); we used the admixture model with correlated allele frequencies, 200,000 as a period of burn-in and 400,000 iterations after burn-in to allow the Markov chain Monte Carlo to reach seasonality. Ten independent simulations were run for each value of K, evaluating from K = 1 to K = 32. The results generated were used to obtain the optimal K, following Evanno's method (Evanno et al., 2005), implemented in the STRUCTURE HARVESTER program (Earl and vonHoldt, 2012). The STRUCTURE program was run again considering the results obtained by Evanno's method and the ancestry coefficients obtained by collection site were shown as circular graphs on a map according to the geographic coordinates of each collection site. 2) Neighbor-Joining analysis (Saitou and Nei, 1987), using Nei's genetic distances (Nei, 1972) and 1000 boostraps with NTSYSpc v.2.2 (Rohlf, 2008). 3) Principal Coordinate Analysis (PCoA) was performed with the GenAlex program ver. 6.502 (Peakall and Smouse, 2017).

Genetic structure was determined evaluating $F_{ST}$, hierarchical Analysis of Molecular Variance (AMOVA), and historical gene flow (Nm), between the genetic groups found and between the natural vegetation sites vs Maya villages. Also, a Mantel test was applied to test the hypothesis of isolation by distance among collection sites. All of these analyses were performed with Arlequin ver. 3.5 program (Excoffier and Lischer, 2010).

#### 2.3.3. Genetic diversity

First, genetic diversity was estimated at three levels: a) species, b) genetic groups found, c) collection sites. The calculated estimators were: allelic richness (*A*), observed heterozygosity ($H_O$), expected heterozygosity ($H_E$) and Fixation index (*F*), with a polymorphism level of 95% using GenAlex ver. 6.502 (Peakall and Smouse, 2017). In order to explore a possible decrease in genetic diversity as a result of the management of the trees collected in the Maya villages, we estimated *A*, $H_O$ and $H_E$ in the 15 natural vegetation sites and the 15 Maya villages and compared them with a one-sided group comparison test and 1000 permutations, using FSTAT version 1.2 (Goudet, 2002). Also, the percentage of reduction in genetic diversity between natural vegetation sites and Maya villages was calculated following the formula $\%r = (H_{G1}\text{-}H_{G2})/H_{G2}$, (where: $H_{G1}$ = genetic diversity of group1 and $H_{G2}$ = genetic diversity of group 2).

## 3. Results

### 3.1. Confidence of data

Of all individuals in the data set, none showed amplification problems and the percentage of missing data was 0%. Only two accessions showed evidence of null alleles in two of its loci, and these were excluded in the final analysis. Accessions showed no evidence of stuttering or dropout alleles.

### 3.2. Organization of genetic diversity and genetic structure

Evanno's method indicated an optimal value of K = 4 (Delta K value = 11.340802) for the 30 sites analyzed; however, this method also showed two other relatively high and very close Delta K values for K = 2 and K = 19 (Delta K values = 7.216799 and 6.484439, respectively) (Fig. 1-A). It has been pointed out that evaluating different K values can facilitate detection of different genetic and demographic processes, thus ensuring a better biological interpretation of the data (Meirmans, 2015). When STRUCTURE was run with K = 4 (Fig. 1-B), 215 individuals collected from natural vegetation sites were clustered into two genetic groups: the Green group, containing 120 individuals collected from eight sites; and the Yellow group, formed by 90 individuals collected

**Table 2**
Characteristics of the nine microsatellite loci used in the analysis of genetic diversity of Huaya India (*Melicoccus oliviformis* Kunth) on the Yucatan peninsula, Mexico.

| Code | 5' to 3' | Primer sequence | TM | Fragmentsize range (bp) | *A* |
|---|---|---|---|---|---|
| 06579_a | Forward | ACAAAACAGAGCTGACTCCAAACC TTGGTGTTTCTGGTCATGAAAATG | 55 | 147-162 | 5 |
| | Reverse | | | | |
| 02129_a | Forward | ACGATGTTTTTGCTGTTGACTTTG TTTCATAAATGTTACGCATGTCACG | 54.5 | 123-142 | 7 |
| | Reverse | | | | |
| 00301_a | Forward | TCAACCATCATTCACCTCAGTGTC | 55 | 167-190 | 2 |
| | Reverse | GCGAAATTGAATCCAGAAGAAGAG | | | |
| 04505_a | Forward | ATCGGATCTCTTGGATCTGTTTTG | 54 | 99-106 | 3 |
| | Reverse | CTCATTCTCATTTCTCCAATTCCC | | | |
| 00695_a | Forward | TTAACTCAACTTTCCGACAGCAGC TCTTGGTAGAGAAGTGAAGCCAGC | 55 | 127-146 | 5 |
| | Reverse | | | | |
| 06603_a | Forward | GCCATTTTCCGTTAAGGAGAGTTC TCTCTATTTAGAACCCCCACCACC | 58 | 163-164 | 5 |
| | Reverse | | | | |
| 05271_a | Forward | TTGTTTGGATGGATGTAATGTAAGG CAGCAGCTAGATAAGCAAAGTTCAG | 53 | 105-113 | 3 |
| | Reverse | | | | |
| 03511_a | Forward Reverse | CAGAAAAAGCGAAATTGAAACCTG GTACTCTCACCGTCACCGAAGG | 57 | 118-138 | 5 |
| 01689_a | Forward | TCCAGTTCCTATTCAGCTTGAACC | 55 | 119-151 | 9 |
| | Reverse | TTCTGGTTCTCACTCATCAAGACG | | | |

**TM**, Temperature Melting; **bp**, base pairs; **A**, number of alleles

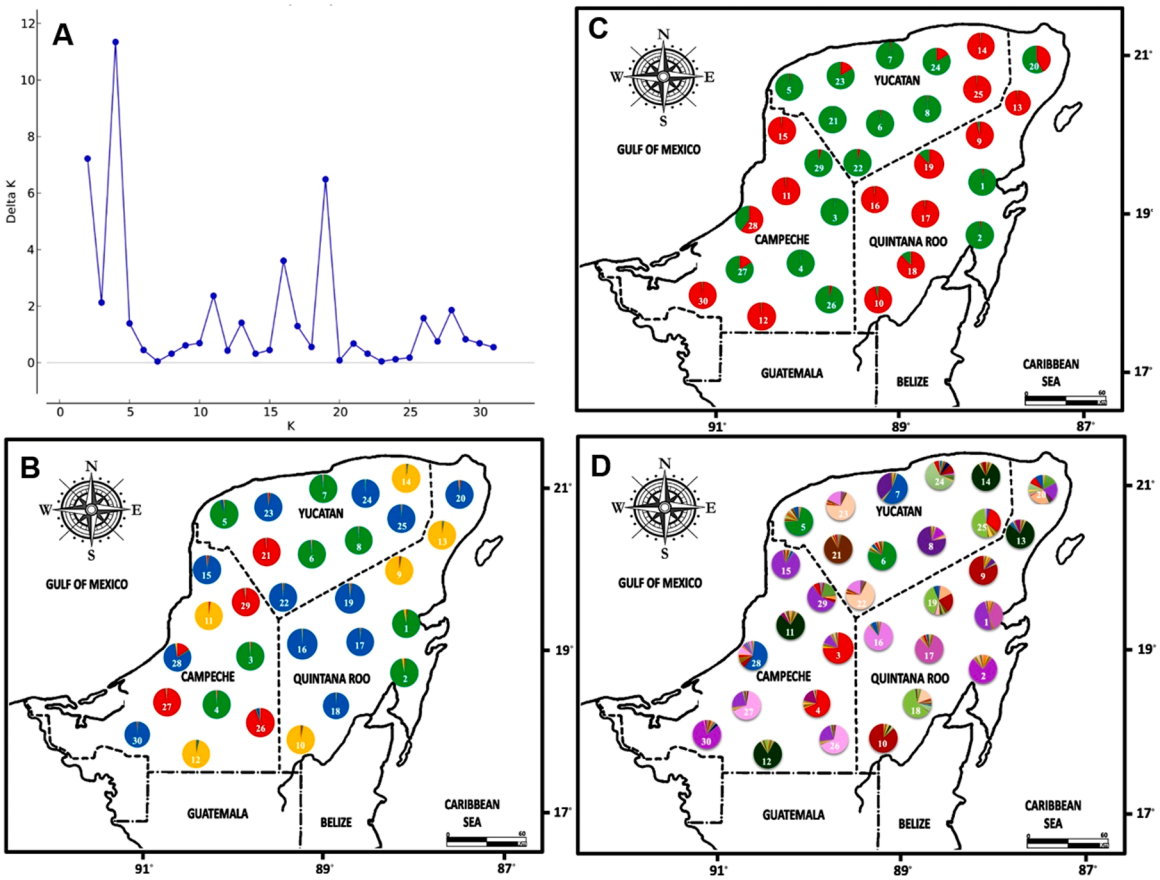


**Fig. 1.** STRUCTURE analysis of 450 trees of Huaya India (*Melicoccus oliviformis*) from the Maya Lowlands of Mexico. A, Delta K plot; B, geographical distribution of the percentages of ancestry estimated with K = 4; C, geographical distribution of the percentages of ancestry estimated with K = 2; D, geographical distribution of the percentages of ancestry estimated with K = 19.

from six sites. Separately, the 225 individuals collected from Maya villages were clustered into two genetic groups: the Blue group, that included individuals collected from 10 Maya villages plus Edzná, a site considered natural vegetation but located in an archaeological zone; and the Red group, formed by individuals collected from five Maya villages. With the exception of some individuals collected from the Maya villages of Vicente Guerrero and Calotmul, individuals collected from each site exhibited high genetic uniformity representative of the genetic group of which they were part. The grouping pattern generated with STRUCTURE and K = 2 (Fig. 1-C) no longer showed a clear difference between natural vegetation sites versus Maya villages. The Green group was integrated by most of the individuals collected in eight natural vegetation sites (Green group, according to K = 4), but also for individuals collected in seven Maya villages. Whereas the Red group was integrated by individuals collected from six natural vegetation sites, but also for individuals collected from six Maya villages. In both genetic groups, individuals with shared ancestry (admixed individuals) were found; particularly, the Maya villages of Cobá and Vicente Guerrero

showed many admixed individuals, and for this reason they were not included in either of the two groups. The grouping pattern generated with STRUCTURE and K = 19 (Fig. 1-D) increased the complexity of the observed groups, as a result of the shared ancestry of many individuals with various collection sites. However, some of the groups observed did show relatively high genetic uniformity. Examples of this were the genetic group integrated by the natural vegetation sites of Cobá and Bacalar; the genetic group formed by the natural vegetation sites of Champotón, Calakmul, Tulum and Temozón Norte; and the genetic group that included the villages of X-Hazil and Señor.

The Neighbor-Joining (N-J) analysis showed two main groups (Fig. 2). Group 1 included 14 natural vegetation sites; even though within this group two subgroups were observed, these did not correspond to the Green and Yellow groups observed with STRUCTURE (K = 4). Group 2 included 15 Maya villages and Edzná, a natural vegetation site; within this group, two subgroups were observed: subgroup 2-a included 10 Maya villages and Edzná, this group is identical to the Blue group of STRUCTURE (K = 4); subgroup 2-b was formed by five Maya villages; this group is identical to the Red group of STRUCTURE (K = 4). The bootstrap values were less than 50% for these groups, so they were not indicated in the dendrogram.

The Principal Coordinate Analysis (PCoA) showed the presence of four major groups and one minor one (Fig. 3). The first and second principal coordinates explained 21.78% and 16.65% of the total variation, respectively. Group 1 integrated the majority of individuals collected from nine natural vegetation sites plus one individual from Edzná. Group 2 integrated all individuals collected from six natural vegetation sites. Group 3 is a minor group formed from only five individuals collected from three natural vegetation sites. Compared with the STRUCTURE results for K = 4, these three groups comprise only individuals from the Yellow and Green groups; compared with N-J analysis, these three groups comprise only individuals of Group 1. Fig. 2 shows that all these individuals can be considered as part of a single large group (large purple circle). On the other hand, Group 3 contained all individuals collected from five Maya villages; compared with the results generated by STRUCTURE (K = 4), this group comprises only individuals from the Red group; compared with N-J analysis, this group is identical to subgroup 2-a. Group 4 included all individuals collected from 10 Maya villages, as well as 14 individuals of Edzná; compared with the results generated by STRUCTURE (K = 4), this group comprises only individuals from the Blue group; compared with N-J analysis, this group is identical to subgroup 2-b.

The clustering patterns generated by Structure, N-J and PCoA analyses together suggest the existence of three main genetic groups (henceforth named according to the colors used in the PCoA): 1) Purple group, made up of individuals from 14 of the natural vegetation sites (Laguna Azul, Laguna Ocom, Mujer Escondida, Miguel Colorado, Lol-Tún, Noh-Bec, Chichen Itzá, Tixcacal Cupul, Cobá, Bacalar, Champotón, Calakmul, Temozón Norte and Tulum), plus one individual of Edzná; 2) Blue group, made up of individuals from 10 Maya villages (X-Hazil, Señor, Calderitas, Chumpón, Cobá Pueblo, Becanchén, Libre Unión, Xocén, Vicente Guerrero and Pueblo Nuevo) and 14 individuals from Edzná; and 3) Red group, made up of individuals from five Maya villages (Yotholín, Calotmul, X-Pujil, Escárcega and Bolonchén).

At the species level, genetic differentiation was high ($F_{ST}$ = 0.562) and historical gene flow was low (Nm= 0.395). Considering the existence of the three main groups mentioned above, the genetic differentiation between the Purple and Blue groups ($F_{ST}$ = 0.136) was less than that observed between the Purple and Red and the Blue and Red groups ($F_{ST}$ = 0.181 in both cases). When the $F_{ST}$ values were compared between groups using permutation tests, the P values did not indicate significant differences between any of the paired comparisons (Purple/Blue, P = 0.48; Purple/Red, P = 0.44; Blue/Red, P = 0.48). The historical gene flow was high, and higher between the Purple and Blue groups (Nm = 17.28), than between the Purple and Red and Blue and Red Groups (Nm = 4.67 and 13.13, respectively). AMOVA showed that most of the genetic variation is found among collected sites within genetic groups (30% of variation, $F_{SC}$ = 0.42, P = 0.000) and among genetic groups (27.3% of variation, $F_{CT}$ = 0.27, P = 0.000), which together accounts for 57.3 % of total variation. The Mantel test did not indicate a process of isolation by distance among populations (r = -0.014, P = 0.585).

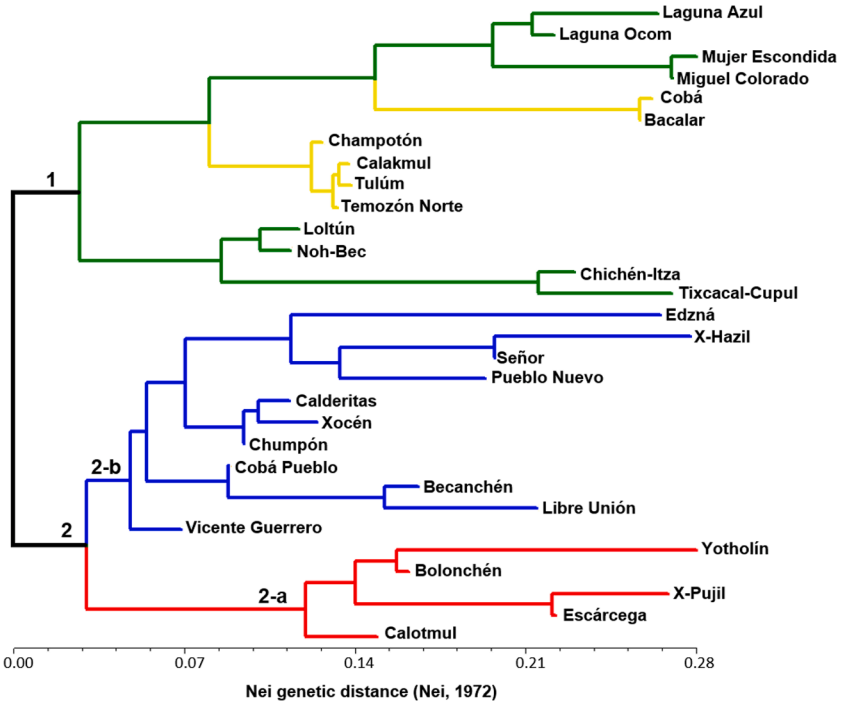


**Fig. 2.** Neighbor-Joining Analysis of 450 trees of Huaya India (*Melicoccus oliviformis*) from the Maya Lowlands of Mexico. The yellow, green, red and blue colors correspond to those used in STRUCTURE and a K = 4 (Fig. 1-B).

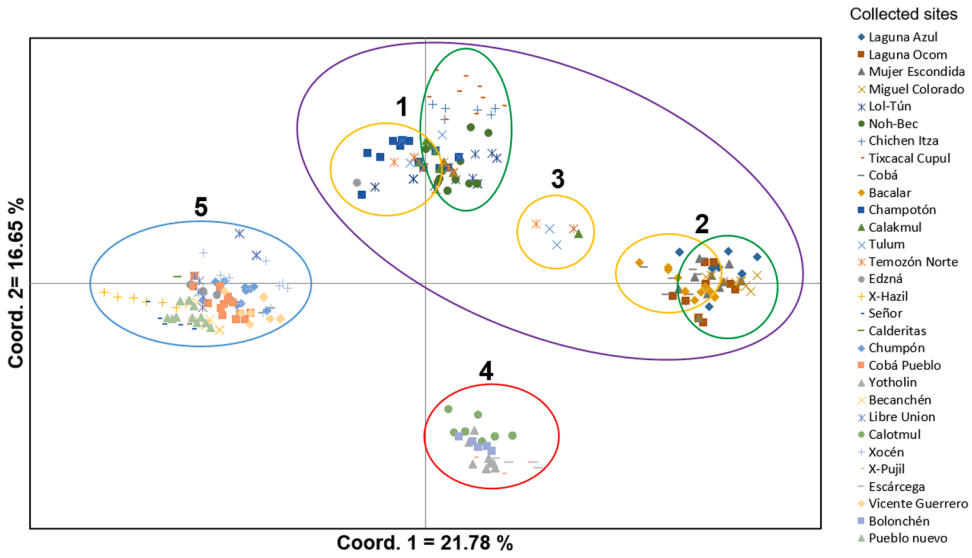


**Fig. 3.** Principal Coordenate Analysis of 450 trees of Huaya India (*Melicoccus oliviformis*) from the Maya Lowlands of Mexico. The yellow, green, red and blue circles correspond to the grouping pattern found with STRUCTURE and a K = 4 (Fig. 1-B). The purple circle integrates all the natural vegetation collection sites, with the exception of Edzná.

### *3.3. Genetic diversity*

In total, 44 alleles were found with the nine loci used. Locus 9 presented the highest number of alleles (9) and locus 3 presented the lowest number of alleles (2). At the species level, the allelic richness (*A*) was 1.67, the observed heterozygosity ($H_O$) was 0.22, the expected heterozygosity ($H_E$) was 0.19 and the inbreeding coefficient was negative (F = -0.15) (Table 3). The Purple group showed the highest levels of genetic diversity for *A*, $H_O$ y $H_E$ (1.74, 0.27 and 0.22, respectively), while the Red group showed the lowest values for these estimators (*A*= 1.56, $H_O$= 0.17, $H_E$= 0.14, respectively), the three groups showed negative values of F and these were close to zero (Table 3). When *A*, $H_O$ and $H_E$ were compared between genetic groups using permutation tests, the P values showed significant differences between Purple and Blue groups (P: *A*= 0.004, $H_O$= 0.005, $H_E$= 0.01) and between Purple and Red groups (P: *A*= 0.021, $H_O$= 0.0013, $H_E$= 0.003); but not between Blue and Red groups (P: *A*= 0.634, $H_O$ 0.402, $H_E$= 0.215). The collection sites that presented higher levels of genetic diversity were from the Purple group, an example of this are Miguel Colorado (*A* = 1.78, $H_O$ = 0.34, $H_E$ = 0.26), Lol-Tún (*A* = 1.78, $H_O$ = 0.30, $H_E$ = 0.26) and Noh-Bec (*A* = 2.00, $H_O$ = 0.32, $H_E$ = 0.26). Edzná, belonging to the Blue group, also presented high genetic diversity (*A* = 1.78, $H_O$ = 0.36, $H_E$= 0.26). The collection sites that showed lower levels of genetic diversity were the Maya villages of Escárcega (*A* = 1.33, $H_O$ = 0.07, $H_E$ = 0.06) and X-Pujil (*A* = 1.44, $H_O$ = 0.12, $H_E$ = 0.09), both from the Red group; X-Hazil (*A* = 1.22, $H_O$ = 0.10, $H_E$ = 0.09) and Becanchén (*A* = 1.33, $H_O$ = 0.12, $H_E$ = 0.11), both from the Blue group (Table 3). The inbreeding value (F) presented a range between -0.53 (Bolonchén) to 0.11 (Chichen Itzá).

When the analyses were conducted grouping natural vegetation sites versus Maya villages, genetic differentiation was similar within groups ($F_{ST}$ = 0.445 and 0.527, respectively) and permutation tests did not show significant differences between the two groups (P = 0.972). AMOVA showed that most of the genetic variation is found among collected sites within groups (38.85% of variation, $F_{SC}$ = 0.47, P = 0.000) and among groups (18.74% of variation, $F_{CT}$ = 0.18, P = 0.000), which together accounts for 57.59% of total variation. The natural vegetation sites showed higher values of genetic diversity (*A* = 1.74, $H_O$ = 0.27, $H_E$ = 0.23) than the Maya villages (*A* = 1.51, $H_O$ = 0.165, $H_E$ = 0.15) (Table 3). When all these diversity estimators were compared between both groups using permutation tests, the P values showed significant differences for the three estimators (*A* = 0.002, $H_O$ = 0.001, $H_E$ = 0.001). The percentage of reduction in genetic diversity in the Maya villages was high, especially for expected and observed heterozygosity (*A* = 15.2%, $H_O$ = 65.4 %, $H_E$ = 51.9%).

## 4. Discussion

### *4. .1 Organization of genetic diversity and genetic structure*

Various methods (STRUCTURE, PCoA, N-J, AMOVA, $F_{ST}$, Mantel test) were used to analyze how the genetic diversity of Huaya India is distributed in the Maya Lowlands of Mexico, which is a possible center of domestication for this species. Three main groups were identified according to their genetic distances: one including only individuals from natural vegetation sites and the other two made up mainly of individuals from Maya villages. In the analysis, most individuals collected from natural vegetation in Edzná clustered together with those of the Maya villages (Blue group), indicating that these individuals are probably descendants of trees previously subjected to human selection by the Maya.

When the grouping pattern observed in this study was compared with the morphological data of the fruits of Huaya India previously reported for the same collection sites by our working group (Jiménez-Rojas et al., 2019), we detected that the Purple group contains trees with small fruits (18.22-23.89 mm), thick epicarp (0.59-0.80 mm), pH of 3.29- 4.97, and a content of total soluble solids of 7.59-15.16 ($^O$Brix); these features seem to be associated with characteristics observed in the wild. Also, we observed that the Blue group presents medium size fruits (24.90-25.07 mm), thin epicarp (0.45-0.40 mm), pH of 5.00-5.10 and a content of total soluble solids of 14.07 to 16.45 ($^O$Brix), whereas the Red group is made up of trees harboring large fruits (25.89-30.34 mm), thin epicarp (0.38-0.19 mm), pH of 5.15-6.59 and a content of total soluble solids of 16.53-18.95 ($^O$Brix); the characteristics observed in both groups, but in particular in the Red group, seem to be associated with the domestication syndrome.

Combining morphological and molecular data of useful species in wild populations and agroecosystems that coexist in the same geographic region is key to understanding the domestication process (Chen et al., 2017; Mastretta-Yanes et al., 2018). In the present study, the three main genetic groups of Huaya India found seems to be a genetic continuum related to the management and selection to which this species has been subjected by the native communities of the Maya

**Table 3**
Genetic diversity estimators of the Huaya India (*Meliccocus oliviformis* Kunth) from the Yucatan Peninsula, Mexico, using nine microsatellite loci.

| | *N* | *A* ± *SD* | $H_O$ ± *SD* | $H_E$ ± *SD* | *F* ± *SD* |
|---|---|---|---|---|---|
| **Species** | 450 | 1.63 ± 0.05 | 0.22 ± 0.02 | 0.17 ± 0.01 | -0.15 ± 0.04 |
| **Genetic groups** | | | | | |
| Purple | 210 | 1.74 ± 0.08 | 0.27 ± 0.03 | 0.22 ± 0.02 | -0.17 ± 0.05 |
| Blue | 165 | 1.52 ± 0.08 | 0.18 ± 0.03 | 0.16 ± 0.02 | -0.09 ± 0.05 |
| Red | 75 | 1.56 ± 0.14 | 0.17 ± 0.05 | 0.14 ± 0.03 | -0.19 ± 0.05 |
| **Collection sites** | | | | | |
| Natural vegetation | 225 | 1.74 ± 0.03 | 0.27 ± 0.06 | 0.23 ± 0.04 | -0.17 ± 0.07 |
| Maya villages | 225 | 1.51 ± 0.06 | 0.17 ± 0.05 | 0.15 ± 0.04 | -0.10 ± 0.07 |
| Laguna Azul | 15 | 2.00 ± 0.37 | 0.22 ± 0.10 | 0.24 ± 0.06 | 0.06 ± 0.19 |
| Laguna Ocom | 15 | 1.56 ± 0.24 | 0.18 ± 0.09 | 0.14 ± 0.07 | -024 ± 0.05 |
| Mujer Escondida | 15 | 1.67 ± 0.24 | 0.32 ± 0.12 | 0.24 ± 0.08 | -0.31 ± 0.17 |
| Miguel Colorado | 15 | 1.78 ± 0.32 | 0.34 ± 0.14 | 0.26 ± 0.09 | -0.37 ± 0.21 |
| Lol-Tún | 15 | 1.78 ± 0.28 | 0.30 ± 0.12 | 0.26 ± 0.09 | -0.21 ± 0.21 |
| Noh-Bec | 15 | 2.00 ± 0.37 | 0.32 ± 0.13 | 0.26 ± 0.09 | -0.14 ± 0.19 |
| Chichen Itzá | 15 | 1.56 ± 0.29 | 0.16 ± 0.11 | 0.19 ± 0.09 | 0.11 ± 0.32 |
| Tixcacal Cupul | 15 | 1.56 ± 0.29 | 0.13 ± 0.09 | 0.15 ± 0.08 | 0.09± 0.23 |
| Cobá | 15 | 1.78 ± 0.32 | 0.33 ± 0.13 | 0.23 ± 0.08 | -0.38 ± 0.15 |
| Bacalar | 15 | 1.67 ± 0.24 | 0.33 ± 0.13 | 0.24 ± 0.08 | -0.36 ± 0.19 |
| Champotón | 15 | 1.67 ± 0.33 | 0.28 ± 0.12 | 0.24 ± 0.09 | -0.23 ± 0.19 |
| Calakmul | 15 | 1.67 ± 0.24 | 0.23 ± 0.13 | 0.18 ± 0.08 | -0.03 ± 0.27 |
| Tulum | 15 | 1.89 ± 0.31 | 0.29 ± 0.12 | 0.25 ± 0.07 | -0.11 ± 0.23 |
| Temozón Norte | 15 | 1.78 ± 0.22 | 0.30 ± 0.14 | 0.24 ± 0.07 | -0.09 ± 0.27 |
| Edzná | 15 | 1.78 ± 0.22 | 0.36 ± 0.15 | 0.26 ± 0.08 | -0.19 ± 0.29 |
| X-Hazil | 15 | 1.22 ± 0.15 | 0.10 ± 0.07 | 0.09 ± 0.06 | -0.13 ± 0.08 |
| Señor | 15 | 1.44 ± 0.24 | 0.20 ± 0.11 | 0.14 ± 0.07 | -0.38 ± 0.14 |
| Calderitas | 15 | 1.44 ± 0.24 | 0.13 ± 0.07 | 0.14 ± 0.08 | 0.09 ± 0.13 |
| Chumpón | 15 | 1.67 ± 0.44 | 0.14 ± 0.07 | 0.16 ± 0.09 | 0.02 ± 0.09 |
| Cobá Pueblo | 15 | 1.56 ± 0.29 | 0.21 ± 0.11 | 0.19 ± 0.10 | -0.11 ± 0.15 |
| Yotholín | 15 | 1.67 ± 0.29 | 0.25 ± 0.13 | 0.22 ± 0.10 | -0.09 ± 0.29 |
| Becanchén | 15 | 1.33 ± 0.24 | 0.12 ± 0.08 | 0.11 ± 0.07 | -0.12 ± 0.18 |
| Libre Unión | 15 | 1.67 ± 0.24 | 0.22 ± 0.09 | 0.23 ± 0.07 | 0.13 ± 0.20 |
| Calotmul | 15 | 1.67 ± 0.37 | 0.16 ± 0.10 | 0.16 ± 0.08 | 0.01 ± 0.27 |
| Xocén | 15 | 1.56 ± 0.24 | 0.22 ± 0.09 | 0.19± 0.08 | -0.08 ± 0.19 |
| X-Pujil | 15 | 1.44 ± 0.29 | 0.12 ± 0.09 | 0.09 ± 0.06 | -0.17 ± 0.23 |
| Pueblo Nuevo | 15 | 1.56 ± 0.29 | 0.19 ± 0.12 | 0.15 ± 0.08 | -0.29 ± 0.06 |
| Vicente Guerrero | 15 | 1.44 ± 0.24 | 0.11 ± 0.08 | 0.12 ± 0.06 | 0.10 ± 0.27 |
| Bolonchén | 15 | 1.67 ± 0.37 | 0.24 ± 0.14 | 0.15 ± 0.09 | -0.53 ± 0.15 |
| Escárcega | 15 | 1.33 ± 0.26 | 0.07 ± 0.05 | 0.06 ± 0.04 | -0.19 ± 0.05 |

***N,*** Number of individuals; ***A,*** Average number of alleles observed per locus; $H_O$, Observed heterozygosity; $H_E$, Expected heterozygosity; F, Fixation Index; SD, Standard Deviation.

Lowlands of Mexico for several thousand years (Colunga-GarcíaMarín and Zizumbo-Villarreal, 2004; Jiménez-Rojas et al., 2018), which has brought this species to a stage of incipient domestication, according to Clement's (1989) classification. Because traditional agricultural habitats can include domesticated plants as well as propagules derived directly from wild individuals, it has been suggested that these populations may represent one portion of a continuum of genetic differentiation ranging from wild to domesticated variants (Harris, 1989); this could be the case of the Huaya India trees collected in the Maya villages that were part of the blue genetic group. This situation has been reported for several species in Mesoamerica, such as tempesquistle (*Sideroxylon palmeri* (Rose) T. D. Penn; González-Soberanis and Casas, 2004), jocote (*Spondias purpurea* L.; Miller and Schaal, 2006), cacao (*Theobroma cacao* L.; Chumacero de Schawe et al., 2013) and ramon (*Brosimum alicastrum* Sw.; Ferrer et al., 2020).

In the Huaya India from the Maya Lowlands of Mexico, the genetic continuum wild-domesticated can be a consequence of high levels of gene flow. A high historical gene flow was observed at the species level, being significantly higher among the Purple/Blue groups than among the Purple/Red groups. These high levels of gene flow are consistent with the general observation that tropical trees exhibit high levels of interbreeding and gene flow (Ward et al., 2005; Petersen et al., 2014). Huaya India is a dioecious species whose trees usually have either staminate or pistillate flowers (Acevedo-Rodríguez, 2003), an aspect that conditions their reproduction to the presence of gene flow. Although there are no studies on the pollination of the Huaya India, different species of flying insects are probably involved, including bees. This dioecious nature of the Huaya India allows people to differentiate between male and female trees, and though the main exploited part of this species are its fruits, Maya people tend to leave some male trees standing on their plots, noting that these provide more shade than female trees (Jiménez-Rojas et al., 2019). This cultural practice, added to the reproduction mode of the species and the existence of Huaya India trees in natural vegetation sites surrounding many of the Maya villages, increase the existence of wild-domesticated gene flow.

Another possible explanation of the genetic continuum wild-domesticated observed in Huaya India is the origin of the villages in the Maya Lowlands of Mexico. When a village is founded by the Maya people, they leave on their backyards the trees that are useful for their fruits, wood, forage, shade, etc. This also happens when villages expand, as people often build their new homes in the surrounding natural vegetation areas, often tropical forest, leaving useful trees standing (Barrera, 1981; Casas et al., 2007). This management of natural vegetation allows the selected trees to stop competing for light, water and nutrients (Parker et al., 2010). In the particular case of Huaya India, this has allowed a better growth of the trees in the backyards and with it the development of different characteristics in architecture (e. g. greater tree girth and lower height) and physiology (e. g. greater production of flowers and fruits), compared to trees found in areas of natural vegetation (Jiménez-Rojas et al., 2018; Jiménez-Rojas et al., 2019). Considering this and the natural distribution of Huaya India in the Yucatán Peninsula (Acevedo-Rodríguez, 2003); probably, the Blue group is represented, in part, by wild individuals that were not eliminated at the time of the establishment and/or expansion of the Maya villages. This hypothesis is supported by the fact that, during the collection of plant material, many owners of the plots indicated that the Huaya India trees were present before the construction of their houses (Jiménez-Rojas et al., 2018). In addition, many Maya villages in the Yucatan peninsula were recently founded, in particular the villages of central Quintana Roo and some regions of Campeche, which occurred after the Caste War toward the end of the 19th century and beginning of the XXth century (Reed, 1971; González-Navarro, 1979).

### 4.2. Genetic diversity

The low genetic diversity found in the Huaya India of the Yucatan peninsula could be evidence of its introduction from northern South America by the Maya people many centuries ago; however, the lack of studies on the genetic diversity of South American populations did not allow us to test this hypothesis. The genetic diversity levels of Huaya India reported here were relatively low when compared to those reported for Spanish Lime collected in the Yucatan Peninsula ($A = 2.61$, $H_O = 0.39$, $H_E = 0.38$, Martínez-Castillo et al., 2019b). These differences in genetic diversity between the Spanish Lime (an introduced species to Mexico with only cultivated individuals) and the Huaya India (a species with both wild and cultivated individuals) raise concerns about the conservation of genetic resources of the Maya Lowlands. The greater genetic diversity found in the Spanish Lime may be due to the greater number of loci evaluated in that species (31 versus 9 loci used for the Huaya India) and the use of automatic sequencers that can detect single-nucleotide changes during the collection of molecular data versus manual sequencers used in the present work. However, it is also important to note that for the Spanish Lime only 25 cultivated trees were analyzed, while in this work 450 trees of Huaya India were evaluated, 225 of which were collected from natural vegetation sites. A possible explanation for the low genetic diversity found in trees collected from natural vegetation sites may lie in the history of vegetation management in the Yucatan Peninsula; it has been pointed out that the jungles of this part of Mexico are, for the most part, secondary vegetation resulting from their management for thousands of years by the Maya people (Barrera et al., 1977).

In addition to the published work on Spanish Lime (Martínez-Castillo et al., 2019a), there is currently no more information on the genetic diversity of species within the genus *Melicoccus* with which to compare our results; in fact, there are few studies on the genetic diversity of the Sapindaceae that have used co-dominant markers such as microsatellites, and these studies have been carried out mainly in species of commercial importance. Compared with Huaya India, high genetic diversity levels have been reported in lychee ($H_E = 0.53$, Madhou et al., 2013; $H_E = 0.454$-$0.782$, Tran et al., 2019), rambutan ($H_E = 0.631$, Ab Razak et al., 2020), and longan ($H_E = 0.46$, Yen et al., 2020). Only in Pulasan (*Nephelium ramboutan-ake* L.), a native Javanese species very similar to rambutam but not of commercial importance, the genetic diversity values reported ($H_E = 0.161$, Puhili et al., 2016) were similar to those found in Huaya India; however, the low genetic diversity observed in Pulasan was based on a much smaller sample size (N = 62) compared to that used here for Huaya India (N = 450).

Although Evanno's method showed a high Delta K value for K = 2, the analysis performed with STRUCTURE (K = 2) did not show a grouping pattern based on the existence of natural vegetation sites vs Maya villages groups, since individuals collected in both collection sites were intermixed. Permutation tests comparing $F_{ST}$ values also did not support the existence of these two groups, since no significant genetic differentiation was found between them. However, in order to provide more information on the genetic diversity of the Huaya India in the Maya Lowlands, we analyzed its diversity by grouping the 450 individuals collected according their collection site.

The significant differences in genetic diversity found in this work between the natural vegetation sites versus Maya villages suggest the existence of a bottleneck in the trees of Huaya India collected from the Maya villages. This reduction in genetic diversity could be result of a founder effect due to domestication; however, this should be taken with caution due to the relatively small number of loci used (9). Studies based in a larger number of SSRs loci or based in Single Nucleotide Polymorphism (SNPs) markers, which provide broader sampling of the genome, are necessary to address this question thoroughly. We found a high reduction in genetic diversity in the trees collected in Maya villages, higher than that commonly reported in perennial fruit species. Reviews on this subject indicate that, in general, domesticated individuals of perennial fruit species retain an average of 91.4 to 99.9% of the genetic variability contained in their wild progenitors (Miller and Gross, 2011). This pattern has been reported mainly in temperate climate species, though it is consistent with that reported for perennial species in the early stages of domestication (Pickersgill, 2007; Hollingsworth et al., 2005), as seems to be the case of the Huaya India (Jiménez-Rojas et al., 2019). Few studies carried out in native tree fruit species of Mesoamerica have addressed this issue. In avocado (*Persea americana* L.), it was reported that domesticated individuals retained 80% of the diversity present in wild progenitor populations (Chen et al., 2009); and in Jocote (*Spondias purpurea* L.), it was found that the cultivated populations retained about 90% of the diversity contained in their wild relatives (Miller and Schaal, 2006). Although is possible that the reduction in genetic diversity in domesticated species may be the consequence of a founder effect due to domestication, other factors can generate bottlenecks, such as the effective size of the population, demographic or environmental events, and human management (Bouzat, 2010).

### 4.3. Implications to commercial management

The information about the Huaya India from Maya Lowlands of Mexico reported here and in our previous publications allows us to visualize the commercial potential of this species. The current production of Huaya India fruits is limited to harvest from trees present in the backyards of Maya villages. Incorporation of this species into large-scale production systems would allow exploration of its potential for commercialization in international markets such as the United States of America. Etnhobotanical information indicate that the Huaya India trees are highly productive (with two fruiting seasons per year) and that they are resistant to drought and do not require a large investment in agricultural inputs (Jiménez-Rojas et al., 2019), aspects that could be beneficial for sustainable commercial management. Also, although currently the Huaya India fruits have a low commercial importance at the regional level, in some parts of the Yucatan peninsula they can reach high prices, in particular fruits with larger sizes and sweeter flavors. Since there are no commercial Huaya India cultivars, the results reported here and those of our previous studies (Jiménez-Rojas et al. 2018, 2019) are of great importance for identifying individuals that show promise for commercial plantations and breeding programs, such as those that were part of the Red group in this study. On the other hand, if future plantations and breeding programs of Huaya India based on elite individuals (for example, those individuals that produce fruits with larger sizes, sweeter flavors, and longer shelf lives) are to be considered, it will also be important to take into account the dynamics of wild-cultivated gene flow, for which the results of this work should be very useful.

## 5. Conclusion

Our study provides a population-level analysis of the Huaya India incorporating a broad sampling of trees collected in natural vegetation sites and Maya villages from the Maya Lowlands of Mexico, using microsatellite markers. We found three main groups: one formed by trees collected only from areas of natural vegetation, the other two groups composed, mainly, of trees collected from Maya villages. These groups appear to behave as a continuum from wild to domesticated individuals that could be the consequence of a high genetic flow and/or the permanence and maintenance of wild trees at the time of the foundation and expansion of the Maya villages. Huaya India presented low levels of genetic diversity: 1) for trees collected from areas of natural vegetation, this could be due to bottlenecks caused by vegetation management for 2000 years, 2) for trees collected from some Maya villages, this could be due to a founder effect due to domestication. Whereas it would be necessary to increase the extent of genomic sampling to test these hypotheses thoroughly, the information generated in this study, together

with the other data already generated by our group, can be useful for future commercial management programs, taking advantage of the market niches opened by other Asian species of Sapindaceae, without compromising the conservation of the species.

## Declaration of Competing Interest

The authors declare that there are no conflicts of interest.

*Acknowledgements*

The first author thanks CONACYT-Mexico for the scholarship for her postgraduate studies. The authors thank Gabriel R. Dzib for the support obtained in the field.

*Funding*

This research was funding by UC MEXUS-CONACYT (Mexico). DP's contributions were partly supported by funding from the USDA National Institute of Food and Agriculture, Hatch project number CAD-PLS-6273-H. This work was supported in part by USDA-ARS project 6044-21000-005-00D."

*Author contribution*

Conceptualization: J.M-C., D.P. and R.H.A-N; Laboratory work: M.I. J-R, O.I.N-A, M.M.O-G and R.S.A; Microsatellite primers design: R.S.A.; Field work: M.I.J-R and O.I.N-A; Formal Analysis: M.I.J-R, J.M-C., D.P. and R.H.A-N; Writing – original draft preparation: M.I.J-R and J.M-C.; Writing – review & editing: All the authors; Supervision and project administration: J.M.C, D.P and R.H.A-N; Funding Acquisition: J.M-C. and D.P. All authors have read and agreed to the published version of the manuscript.

*Submission declaration*

This work has not been published previously

## References

Ab Razak, S., Mad Radzuan, S., Mohamed, N., Nor Azman, N.H.E., Abd Majid, A.M., Ismail, S.N., Nasir, K.H., 2020. Development of novel microsatellite markers using RAD sequencing technology for diversity assessment of rambutan (*Nephelium lappace*um L.) germplasm. J. Heliyon. 6 (9), e05077. https://doi.org/10.1016/j.heliyon.2020.e05077.

Acevedo-Rodríguez, P., 2003. *Mellicocceae* (*Sapindaceae*): *Melicoccus* and *Talisia*. Flora neotropical monograph 87. Botanical. Garden Bronx, New York 27–59 págs.

Altendorf, S., 2018. Minor tropical fruits: Mainstreaming a niche market. Food Outlook. http://www.fao.org/3/a-I8080e.pdf (Accessed 2 February 2021).

Barrera, A., 1981. Sobre la unidad de habitación tradicional campesina y el manejo de recursos bióticos en el área maya yucatanense. Árboles y arbustos de los huertos familiares. Biotica 5, 115–128.

Barrera, A., Gómez-Pompa, A., Vázquez-Yanes, C., 1977. El manejo de las selvas por los Mayas: Sus implicaciones silvícolas y agrícolas. Biótica 2 (2), 47–61.

Bassam, B.J., Anollés, G.C., Gresshoff, P.M., 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal. Biochem. 196, 80–83. https://doi.org/10.1016/0003-2697(91)90120-i.

Bouzat, J., 2010. Conservation genetics of population bottlenecks: The role of chance, selection, and history. Conserv. Genet. 11, 463–478. https://doi.org/10.1007/s10592-010-0049-0.

Carnevali, F.C.G., Tapia-Muñoz, J.L., Duno de Stefano, R., Ramírez-Morillo, I., 2010. Flora Ilustrada de la Península de Yucatán: Listado Florístico. Centro de Investigación Científica de Yucatán, Yucatán, México, 9686077823070. A.C. ISBN.

Casas, A., Otero-Arnaiz, A., Perez-Negron, E., Valiente-Banuet, A., 2007. In situ management and domestication of plants in mesoamerica. Ann. Bot. 100 (5), 1101–1115. https://doi.org/10.1093/aob/mcm126 https://.

Chen, H., Morrell, P.L., Ashworth, V.E.T.M., De la Cruz, M., Clegg, M., 2009. Tracing the Geographic Origins of Major Avocado Cultivars. J. Heredity. 100 (1), 56–65. https://doi.org/10.1093/jhered/esn06.

Chen, Y.H., Shapiro, L.R., Benrey, B., Cibrián-Jaramillo, A., 2017. Back to the origin: in situ studies are needed to understand selection during crop diversification. Front Ecol Evol 5, 125–134. https://doi.org/10.3389/fevo.2017.00125.

12 Chumacero de Schawe, C., Durka, W., Tscharntke, T., Hensen, I., Kessler, M., 2013. Gene flow and genetic diversity in cultivated and wild cacao (*Theobroma cacao*) in Bolivia. Am. J. Bot. 100, 2271–2279. https://doi.org/10.3732/ajb.1300025.

Clement, C.R., 1989. A Center of Crop Genetic Diversity in Western Amazonia. Bioscience 39, 624–631.

Colunga-GarcíaMarín, P., Zizumbo-Villarreal, D., 2004. Domestication of plants in Maya Lowlands. Econ. Bot. 58, S101–S110. https://doi.org/10.1663/0013-0001(2004)58[S101:DOPIML]2.0.CO;2.

Earl, D.A., vonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4 (2), 359–361. https://doi.org/10.1007/s12686-011-9548-7.

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

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite Ver. 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x.

Ferrer, M.M., Montañez-Escalante, P.I., Ruenes-Morales, M.R., Estrada-Medina, H., Jiménez-Osornio, J., 2020. Growing out of the tropical forests: domestication syndrome of native Mesoamerican trees in Mayan homegardens. Genetic. Resour. Crop. Evol. 67, 587–604. https://doi.org/10.1007/s10722-019-00833-2.

González-Navarro, M., 1979. Raza y tierra. La guerra de castas y el henequén. Centro de Estudios Históricos, Nueva serie 10. El Colegio de México, México, D.F., p. 156 págs.

González-Soberanis, C., Casas, A., 2004. Traditional management and domestication of tempesquistle, *Sideroxylon palmeri* (Sapotaceae) in the Tehuacán-Cuicatlán Valley, Central Mexico. J. Arid Environ. 59, 245–258. https://doi.org/10.1016/j.jaridenv.2004.01.018.

Goudet, J., 2002. Fstat version 1.2: a program to estimate and test gene diversities and fixation indices. J. Hered. 86, 485–486.

Harris, D.R., 1989. An evolutionary continuum of plant–people interaction. In: Harris, DR, Hillman, GC (Eds.), Foraging and Farming: the Evolution of Plant Exploitation. Unwin-Hyman, Boston, pp. 11–26.

Hollingsworth, P.M., Dawson, I.K., Goodall-Copestake, W.P., Richardson, J.E., Weber, J. C., Sotelo Montes, C., Pennington, R.T., 2005. Do farmers reduce genetic diversity when they domesticate tropical trees? A case study from Amazonia. Mol. Ecol. 14, 497–501. https://doi.org/10.1111/j.1365-294X.2005.02431.x.

Jiménez-Rojas, M.I., Andueza-Noh, R.H., Martínez-Castillo, J., Potter, D., 2019. Management and cultivation of the huaya india (*Melicoccus oliviformis* Kunth) on the Yucatan Peninsula. Econ. Bot. 73, 429–442. https://doi.org/10.1007/s12231-019-09470-3.

Jiménez-Rojas, M.I., Martínez-Castillo, J., Potter, D., Dzib, G.R., Ballina-Goméz, H., Latournerie-Moreno, L., Andueza-Noh, R.H., 2018. Morphological diversity of Huaya India fruits (*Melicoccus oliviformis* Kunth) in the Maya Lowlands. Genet. Resour. Crop. Evol. 66, 513–522. https://doi.org/10.1007/s10722-018-00731-z.

Madhou, M., Normand, F., Bahorun, T., Hormaza, J.I., 2013. Fingerprinting and analysis of genetic diversity of litchi (*Litchi chinensis* Sonn.) accessions from different germplasm collections using microsatellite markers. Tree Genet. Genomes. 9, 387–396. https://doi.org/10.1007/s11295-012-0560-1.

Martínez-Castillo, J., Arias, R.S., Andueza-Noh, R.H., Ortiz-García, M.M., Irish, B.M., Scheffler, B.E., 2019a. Microsatellite markers in Spanish lime (*Melicoccus bijugatus*, Sapindaceae), a neglected Neotropical fruit crop. Genet. Resour. Crop Evol. 66, 1371–1377. https://doi.org/10.1007/s10722-019-00815-4.

Martínez-Castillo, J., Blancarte-Jasso, N.H., Chepe-Cruz, G., Nah-Chan, NG., Ortiz-García, M.M., Arias, R.S., 2019b. Structure and genetic diversity in wild and cultivated populations of Zapote mamey (*Pouteria sapota*, Sapotaceae) of southeastern Mexico: its supposed domestication center. Tree Genet. Genomes. 15, 1–11. https://doi.org/10.1007/s11295-019-1368-z.

Mastretta-Yanes, A., Acevedo Gasman, F., Burgeff, C., Cano Ramírez, M., Piñero, D., Sarukhán, J., 2018. An initiative for the study and uses of genetic diversity of domesticated plants and their wild relatives. Front. Plant Sci. 9, 209–216. https://doi.org/10.3389/fpls.2018.00209.

Meirmans, P.G., 2015. Seven common mistakes in population genetics and how to avoid them. Mol. Ecol. 24, 3223–3231. https://doi.org/10.1111/mec.13243.

Miller, A.J., Gross, B.L., 2011. From forest to field: perennial fruit crop domestication. Am. J. Bot. 98, 1389–1414. https://doi.org/10.3732/ajb.1000522.

Miller, A.J., Schaal, B.A., 2006. Domestication and the distribution of genetic variation in wild and cultivated populations of the Mesoamerican fruit tree *Spondias purpurea* L. (Anacardiaceae). Mol. Ecol. 15, 1467–1480. https://doi.org/10.1111/j.1365-294X.2006.02834.x https://doi.org/.

Nei, M., 1972. Genetic distance between populations. Am. Nat. 106, 283–292. https://doi.org/10.1086/282771.

Parker, I.M., López, I., Petersen, J., Anaya, N., Cubilla-Rios, L., Potter, D., 2010. Domestication syndrome in caimito (*Chrysophyllum cainito* L.): Fruit and seed characteristics. Econ. Bot. 64 (2), 161–175. https://doi.org/10.1007/S12231-010-9121-4 https://.

Paull, R.E., Duarte, O., 2012. Tropical fruits. CAB International, London, pp. 34–57 págs.

Peakall, R., Smouse, P., 2017. GenAlex 6.5: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6, 288–295. https://doi.org/10.1093/bioinformatics/bts460 https://.

Petersen, J.J., Parker, I.M., Potter, D., 2014. Domestication of the neotropical tree *Chrysophyllum cainito* from a geographically limited yet genetically diverse gene pool in Panama. Ecol. Evol. 4, 539–553. https://doi.org/10.1002/ece3.948.

Pickersgill, B., 2007. Domestication of Plants in the Americas: Insights from Mendelian and Molecular Genetics. Ann. Bot. 100 (5), 925–940. https://doi.org/10.1093/aob/mcm193.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Puhili, A.L., Chikmawatim, T., Djuita, N.R., 2016. Evaluation of Pulasan (*Nephelium ramboutan-ake*) genetic diversity in Bogor, West Java, using microsatellite markers. J. Trop. Life Sci. 6 (3), 184–189. https://doi.org/10.11594/jtls.06.03.09.

Reed, N., 1971. La guerra de castas de Yucatán. Ediciones Era, México, D.F., pp. 297–302 págs.

Rohlf, F.J., 2008. NTSYS-pc. Numerical taxonomy system version 2.2. Exeter Publishing, Ltd., Setauket.

Tran, H., Kanzaki, S., Triest, L., Hormaza, I., Kuk, N.J., Ming, R., Bousquet, J., Khasa, D., Van Damme, P., 2019. Analysis of genetic diversity of lychee (*Litchi chinensis* Sonn.) and wild forest relatives in the Sapindaceae from Vietnam using microsatellites. Genet. Resour. Crop Evol. 66, 1653–1669. https://doi.org/10.1007/s10722-019-00837-y.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P., 2004. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4, 535–538.

Vieira, M.L., Santini, L., Diniz, A.L., Munhoz, C., 2016. Microsatellite markers: what they mean and why they are so useful. Genet. Mol. Biol. 39, 312–328. https://doi.org/10.1590/1678-4685-GMB-2016-0027.

Ward, M., Dick, C., Gribel, R., Lowe, A,J., 2005. To self or not to self? A review of outcrossing and pollen mediated gene flow in neotropical trees. Heredity 95, 246–254. https://doi.org/10.1038/sj.hdy.6800712.

Yen, T.T.O., Thao, D.T.T., Nhu, T.T.T., Truong, N.N., Bay, D.T.B., Hoa, N.V., 2020. Genetic diversity of longan cultivars in Vietnam. Acta Hortic 1293, 105–112. https://doi.org/10.17660/ActaHortic.2020.1293.15.