

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Publications from USDA-ARS / UNL Faculty

U.S. Department of Agriculture: Agricultural
Research Service, Lincoln, Nebraska

7-6-2019

Structure and genetic diversity in wild and cultivated populations of Zapote mamey (*Pouteria sapota*, Sapotaceae) from southeastern Mexico: its putative domestication center

Jaime Martínez-Castillo

Centro de Investigación Científica de Yucatán (CICY), jmartinez@cicy.mx

Nassib H. Blancarte-Jasso

Centro de Investigación Científica de Yucatán (CICY)

Gabriel Chepe-Cruz

Centro de Investigación Científica de Yucatán (CICY)

Noemí G. Nah-Chan

Centro de Investigación Científica de Yucatán (CICY)

Matilde M. Ortiz-García

Centro de Investigación Científica de Yucatán (CICY)

See next page for additional authors

Follow this and additional works at: <https://digitalcommons.unl.edu/usdaarsfacpub>

Martínez-Castillo, Jaime; Blancarte-Jasso, Nassib H.; Chepe-Cruz, Gabriel; Nah-Chan, Noemí G.; Ortiz-García, Matilde M.; and Arias, Renee S., "Structure and genetic diversity in wild and cultivated populations of Zapote mamey (*Pouteria sapota*, Sapotaceae) from southeastern Mexico: its putative domestication center" (2019). *Publications from USDA-ARS / UNL Faculty*. 2200.
<https://digitalcommons.unl.edu/usdaarsfacpub/2200>

This Article is brought to you for free and open access by the U.S. Department of Agriculture: Agricultural Research Service, Lincoln, Nebraska at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Publications from USDA-ARS / UNL Faculty by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Jaime Martínez-Castillo, Nassib H. Blancarte-Jasso, Gabriel Chepe-Cruz, Noemí G. Nah-Chan, Matilde M. Ortiz-García, and Renee S. Arias



Structure and genetic diversity in wild and cultivated populations of Zapote mamey (*Pouteria sapota*, Sapotaceae) from southeastern Mexico: its putative domestication center

Jaime Martínez-Castillo¹ · Nassib H. Blancarte-Jasso¹ · Gabriel Chepe-Cruz¹ · Noemí G. Nah-Chan¹ · Matilde M. Ortiz-García¹ · Renee S. Arias²

Received: 13 August 2018 / Revised: 10 June 2019 / Accepted: 30 June 2019 / Published online: 6 July 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

This document is a U.S. government work and is not subject to copyright in the United States.

Abstract

Tropical fruit trees are an important component of the human diet; however, little is known about their genetic diversity levels. Zapote mamey (*Pouteria sapota*) is a tree native to southeastern Mexico and Central America, and Mexico is the leading producer in the world. Studies of the genetic diversity of Zapote mamey have been based on cultivated materials using morphological and biochemical characterization or dominant molecular markers. To gain a deeper understanding about the conservation status of Zapote mamey in its center of origin and domestication, we collected 188 individuals from eight wild and five cultivated populations in southeastern Mexico and characterized them using eight microsatellite loci. STRUCTURE, 3D-PCoA, and neighbor-joining analyses showed three groups in the wild gene pool and one group in the cultivated gene pool. F_{ST} values were significant between wild and cultivated gene pools, among the four groups observed and among the 13 populations collected (0.13, 0.25, and 0.36, respectively). Overall, we found low levels of genetic diversity ($A = 2.77$, $H_O = 0.29$, $H_E = 0.39$), permutation tests did not show significant differences between wild and cultivated gene pools. The Garza–Williamson index showed low values in both gene pools (wild = 0.16, cultivated = 0.11) and the Bottleneck program indicated a decrease in genetic diversity in both gene pools (wild, $P = 0.027$; cultivated, $P = 0.054$); both analyses suggest a potential genetic bottleneck within this species. This study can help to generate adequate sampling techniques and to develop effective management strategies for Zapote mamey of southeastern Mexico.

Keywords Bottleneck · Microsatellite markers · Sapotaceae · SSR · Yucatán Peninsula

Introduction

Perennial fruit species are a basic component of the human diet (Haq et al. 2008; Schreckenber et al. 2006). Tropical regions in particular harbor a great diversity of fruit trees that for centuries have provided food and medicine to the mankind; these trees include at least 1000 species reported in America, 1200 in Africa, and 500 in Asia (Awodoyin et al.

2015; Normah et al. 2013; Paull and Duarte 2012). Despite their great importance, little is known about the genetic diversity of many tropical fruit trees (Jamnadass et al. 2009). Such information is not only critical to understand the genetic vulnerability of these species, but also could be used to implement management and conservation strategies, particularly in the centers of origin of the tropical fruit crops where wild and cultivated populations coexist.

Zapote mamey (*Pouteria sapota* (Jacq.) H.E. Moore & Stearn) is a Neotropical species of the family Sapotaceae. The lowlands of southeastern Mexico and Central America have been considered as its plausible center of domestication (Azurdía 2006; Bruner and Morales-Payan 2011). The natural habitat of Zapote mamey is the medium or high evergreen forests at altitudes not exceeding 1300 m, located along the Gulf of Mexico in the states of Veracruz and Tabasco, Mexico, and along the Pacific, in the state of Jalisco, Mexico, extending through Panama (Azurdía 2006; León 1987). Zapote

Communicated by C. Dardick

✉ Jaime Martínez-Castillo
jmartinez@cicy.mx

¹ Centro de Investigación Científica de Yucatán (CICY), Calle 43 No. 130. Col. Chubumá de Hidalgo, CP 97200 Mérida, Yucatán, Mexico

² USDA-ARS-National Peanut Research Laboratory, 1011 Forrester Dr., SE, Dawson, GA 39842-0509, USA

mamey is cross-pollinated (Villarreal-Fuentes et al. 2015) and, although its specific pollinators are not known, it has been reported that flowers of the genus *Pouteria* are pollinated by bees and other insects (Pennington 1990; Ortiz and Cabello 1991; Knight et al. 1993). Among the dispersers of fruits and seeds of Zapote mamey are listed mammals such as *Agouti paca nelsoni* and *Potos flavus prehensilis* (Brewer and Rejmanek 1999; Martínez-Gallardo and Sánchez-Cordero 1997). Zapote mamey can be propagated sexually or asexually, though in natural conditions, reproduction by seed is rare due to animal foraging (Nava-Cruz and Ricker 2004). Under cultivation, reproduction occurs by seed, by cuttings, or by grafting, the latter being typically employed in small plantations (Bruner and Morales-Payan 2011).

The fruit of the Zapote mamey is highly nutritious and has distinct organoleptic characteristics, which increase its commercial value (Gordon et al. 2011; Gulyas-Fekete et al. 2013; Moo-Huchin et al. 2013; Murillo et al. 2011). Mexico is the leading producer of Zapote mamey in the world (SIAP 2015), and it is commonly grown in small plantations and orchards or harvested from wild trees when local people can distinguish between wild and cultivated individuals based on fruit quality (Bruner and Morales-Payan 2011). In 2014, 1651 ha of Zapote mamey were planted in 15 states in Mexico, yielding a total production of 17,586 t of fruit (SIAP 2015). More than 800 ha are planted under commercial production in the state of Yucatán, aspect that, in addition to the production in backyards in the Yucatán Peninsula (Yucatán, Campeche, and Quintana Roo states) destined for self-consumption or for local scale sale, converting this region in the main producer of Zapote mamey in Mexico (Villegas-Monter et al. 2016). Zapote mamey is currently cultivated in other tropical areas such as Florida (United States of America), Central America, Colombia, and the Caribbean islands. Additionally, there is growing interest in cultivating it in Australia, Israel, the Philippines, Vietnam, Spain, and Venezuela (Balerdi and Crane 2015).

The study of the genetic diversity of Zapote mamey has been mainly based on the morphological and biochemical characterization of cultivated materials (Bañuelos-Jimenez and Ochoa 2006; Espinosa-Zaragoza et al. 2005; Gaona-García et al. 2008; Nascimento et al. 2008). Previous studies have been conducted on Zapote mamey using molecular markers to determine the genetic relationships and geographic origin of the cultivated plants applying dominant molecular markers such as AFLPs and RAPDs (Arellano-Durán 2012; Carrara et al. 2004; Ibarra-Estrada 2012; Rodríguez-Rojas et al. 2012). Crucially, these studies did not include wild individuals in their analysis, thus limiting its scope. Recently, the first microsatellite codominant markers (SSR—simple sequence repeats) specific to Zapote mamey were published by Arias et al. (2015). To determine the level of polymorphism and the potential utility of these molecular markers, these authors tested 205 SSR loci in 20 wild and nine cultivated individuals of Zapote mamey from

southeastern Mexico, reporting slightly higher levels of genetic diversity in the cultivated gene pool, but also finding some evidence of a bottleneck in this gene pool. Considering the importance of Zapote mamey for Mexico, our goal was to evaluate the structure and genetic diversity of Zapote mamey of southeastern Mexico, analyzing a large number of wild (115) and cultivated (73) individuals with eight SSR loci reported as polymorphic by Arias et al. (2015).

Material and methods

LocI information

Eight SSR loci reported as polymorphics to Zapote mamey by Arias et al. (2015) were used for the molecular characterization (Table 1). Linkage disequilibrium (LD) analyses among pairs of loci were done using a likelihood-ratio test, whose empirical distribution is obtained by a permutation procedure (Slatkin and Excoffier 1996). LD analyses were done with 1000 permutations using Arlequin 3.0 (Excoffier et al. 2005). Allele size range (in base pairs—bp), number of alleles per locus (*A*), and minor allele frequency (*MAF*) were calculated globally across the entire data set of samples analyzed in GeneMapper 4.0 (Applied Biosystems, Foster City, CA) (Table 1).

Plant material

Eight wild and five cultivated populations of Zapote mamey were collected in southeastern Mexico (Fig. 1). For the purpose of this study, we defined a wild population as a group of individuals growing in a localized site in their natural habitat without human care, and a cultivated population as a group of individuals under any level of human care within the same town. In the case of wild populations, individuals collected were more than 20 m apart and throughout the population distribution area. For the cultivated populations, one individual was collected per orchard, trying to cover the entire town. In total, 188 individuals were collected: 115 wild and 73 cultivated individuals, with an average of 15 individuals from each population.

DNA extraction and microsatellite technique

DNA was extracted from young leaves using of ChargeSwitch gDNA Plant Kit, following manufacturer instructions (Invitrogen). The quality of extracted DNA was verified by electrophoresis in gels of 1% agarose, stained with ethidium bromide. DNA quantification was performed in a NanoDrop 1000 spectrophotometer (Thermo Scientific). DNA amplification was done in 20- μ L-volume polymerase chain reaction (PCR) consisting of 10 \times PCR buffer, 50 mM of MgCl₂, 10 mM of dNTP mix, 10 mM of primers, 5 u/ μ L of Taq

Table 1 Characteristics of the eight microsatellite loci used in the analysis of structure and genetic diversity of 13 populations of Zapote mamey (*Pouteria sapota*) from southeastern Mexico

Locus code	5' to 3'	Primer sequence	TM	Allele size range (bp)	A	MAF
280_a	Forward	TTCGAGTTTCTCACTTTGGTTTCC	59°	140–158	4	0.313
	Reverse	AGTCGACTCTGTCTGAATTGATGC				
588_a	Forward	CTCGGTAAAACCTCGGCTCAAATAC	59°	138–146	2	0.113
	Reverse	CTGACGAGTTCCTCCCTCCTC				
756_c	Forward	ACCAAGTCCTTCTCCTCTCCAAAC	59°	153–177	4	0.185
	Reverse	ACAGGAGAAGAAAGGGAAAACCAG				
1078_a	Forward	GAACCACTTACCAAGCTCCAACCT	58°	89–95	2	0.092
	Reverse	GGGTATTGGAAGTAAAGAAACCAGG				
2496_a	Forward	TGCAAACTCTGTGCTAGTTGATTTTC	59°	110–125	3	0.211
	Reverse	ATATTCAAAGACTTTTGGGGAGGG				
2670_a	Forward	AGGGTTCCTTATGACCTTTGCATTC	58°	126–136	3	0.332
	Reverse	GGACAGATGGATGTTTCTATGTATG				
3313_a	Forward	TTCACTCTTTCCGGAATCAAAG	58°	160–170	2	0.085
	Reverse	GTTTTCAACCTCCAATGTCAATCC				
3976_a	Forward	TTCGTAATGATTTGATAGTAGTTGTTGG	58°	125–133	4	0.209
	Reverse	AAATTTCCCTATTGCAAGGAGATG				

TM, annealing temperature in °C; bp, base pairs; A, number of alleles per locus; MAF, minor allele frequency

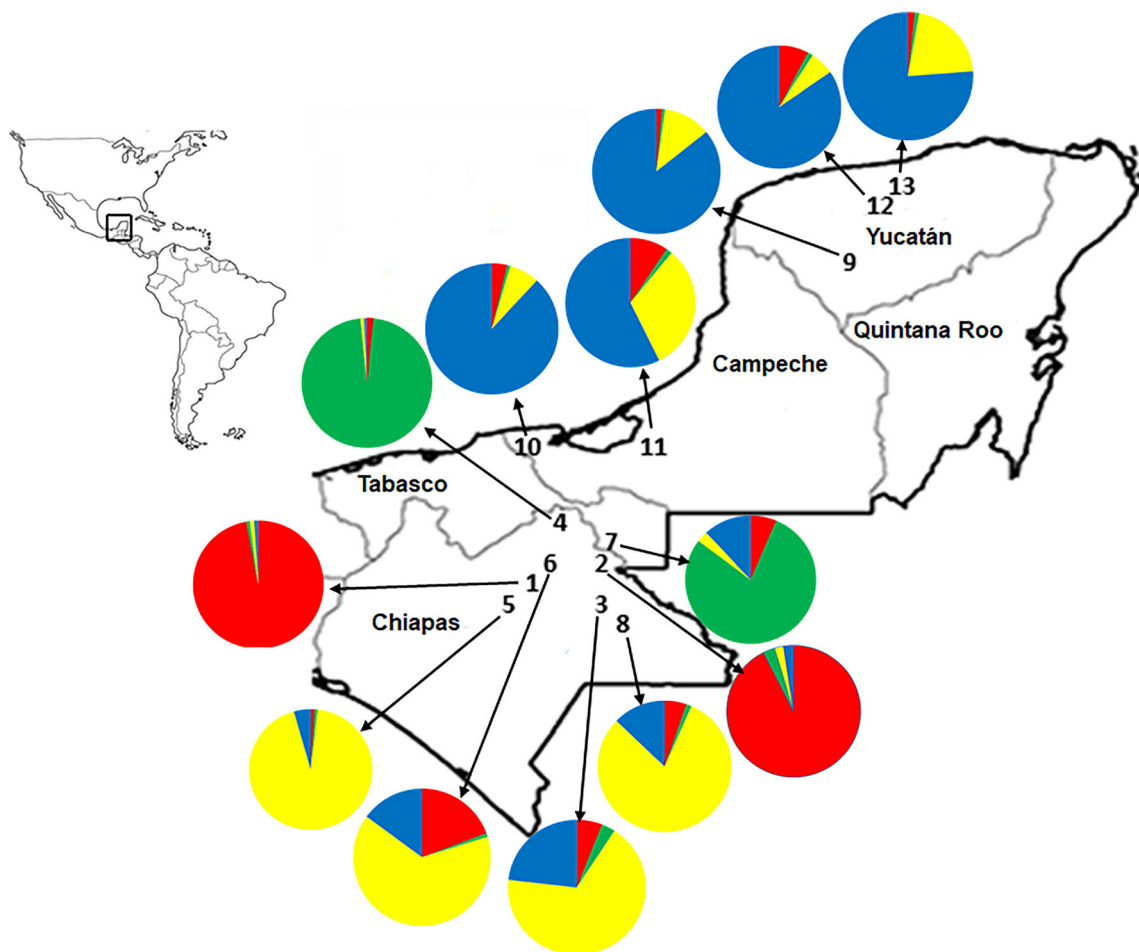


Fig. 1 Collection area of the 13 populations of Zapote mamey (*Pouteria sapota*) analyzed in the present study. Wild populations: Yajalon (1), Ursula Galvan (2), Welib-Ha (3), Palenque (4), Jose Castillo (5), Agua Azul (6), Penjamo (7), Nueva Palestina (8). Cultivated populations:

Oxcutzcab (9), Atasta (10), Chekubul (11), Tibolon (12), Dzidzantun (13). Pie charts for each population show the results of the test of assignment of individuals, based on K optimal = 4, using the STRUCTURE program

polymerase, and 10 ng of DNA. The PCR program consisted of 35 cycles, each included a denaturing step of 2 min at 94 °C, an alignment step of 1 min (at a temperature of alignment depending on the primers employed), an extension step of 5 min at 72 °C, and a final extension for 5 min at 72 °C. The PCR amplification was performed in a thermocycler GeneAmp PCR System 9700 (Applied Biosystems, Foster City, USA). 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 was loaded on 5% polyacrylamide gels (19:1 acrylamide-bisacrylamide) containing 5 M urea and 0.5 × Tris-borate EDTA (TBE) buffer. Electrophoresis was performed at 60-W constant power for 45 min to 1 h (SQ3 sequencer, Hoeffer Scientific Instruments, San Francisco, CA). The amplified products were visualized with the silver staining technique Bassam et al. 1991). The size of the fragments were determined visually in base pairs (bp), using as reference a 10-bp molecular marker (Invitrogen, Life Technologies, Brazil).

Data analysis

Confidence of data

To verify the accuracy of the genotyping executed, SSR data was analyzed using the MICRO-CHECKER program v.2.2.3 (Van Oosterhout et al. 2004). This program detects the presence of null alleles, alleles that failed to amplify during PCR; stuttering, slight changes that have occurred in the allele size during PCR; and dropout alleles, large alleles that do not amplify as efficiently as smaller alleles. Null alleles, stuttering, and dropout alleles can affect the estimation of the diversity and genetic structure of populations when utilizing the microsatellite technique. Also, considering the limitations of the silver staining method used to determine the size of the bands (alleles) observed in the polyacrylamide gels, we checked the specific values (bp) of the alleles found in each of the 13 populations studied.

Genetic structure

To determine how the genetic diversity was organized in the wild and cultivated populations of Zapote mamey studied, data were analyzed using three grouping methods. The first was a Bayesian approach implemented with the STRUCTURE program (Pritchard et al. 2000), using the admixture model with correlated allele frequencies, with 100,000 as a period of burn-in and 200,000 iterations after burn-in to allow the Markov chain to reach stationarity. Ten independent simulations were run for each value of K , ranging from $K = 1$ to $K = 14$, checking for consistency across outputs.

The outfiles of the STRUCTURE were used in the STRUCTURE HARVESTER program (Earl and vonHoldt 2012) to obtain the optimal K value according to Evanno et al. (2005). A summary of the ancestry coefficients generated by the STRUCTURE were displayed by making pie charts for each of the sample sites, using the K optimal obtained (Fig. 1). The second approach utilized a three-dimensional principal coordinate analysis (3-D PCoA); it was performed with the computer package NTSYS-pc version 2.1 (Rohlf 2001). The third approach used a neighbor-joining (N-J) tree of all individuals as implemented with the Populations 1. 2.30 software (Langella 2002). The N-J was constructed using Nei's standard genetic distance (Nei 1987). The tree topology was displayed and edited with MEGA 6 software (Tamura et al. 2013). Then, genetic differentiation was assessed calculating F_{ST} values between wild and cultivated gene pools, among gene groups defined on the basis of the grouping methods before mentioned, and among populations. For this, we applied 1000 permutations using Arlequin 3.5 (Excoffier et al. 2005). Finally, recent gene flow was estimated among gene groups using BayesAss program (Wilson and Rannala 2003). While the STRUCTURE uses a Bayesian probabilistic model to assign individuals to clusters, the BayesAss estimates the posterior probability of an individual's migratory history and thus allows estimating the rate and direction of recent dispersal. Unlike estimators of long-term gene flow, the BayesAss makes relatively few assumptions about demography and can be applied to populations that are not in the Hardy-Weinberg equilibrium. The MCMC method was run for 20,000,000 iterations with a burn-in period of 1,000,000 and a sampling frequency of 2000 iterations.

Genetic diversity

Genetic diversity was estimated at four levels: (a) at species, (b) within each of the wild and cultivated gene pools, (c) within the different groups defined by the three grouping methods, and (d) within each individual population. To allow comparisons between groups and populations with different sample size, the average number of alleles per locus (A) was calculated using the rarefaction method implemented in HP-Rare version 1 (Kalinowski 2005). Observed heterozygosity (H_O) and expected heterozygosity (H_E) were calculated with a level of polymorphism of 95% using Arlequin 3.0 (Excoffier et al. 2005). In order to compare H_O and H_E between wild and cultivated gene pools, the FSTAT program (Goudet 2002) was used to compute a one-sided group comparison test with 1000 permutations. Finally, two methods were used to explore the potential existence of a bottleneck in Zapote mamey: (1) the modified Garza–Williamson index (M) (Garza and Williamson 2001; Excoffier et al. 2005) computes the mean ratio of the number of alleles at a given locus (k) with respect to the range of allele size (r). Studies of several natural

populations have shown that M is greater than 0.82 for populations that have not suffered a known reduction in population size and less than 0.70 for those that have undergone a bottleneck (Garza and Williamson 2001). This index was estimated with the Arlequin 3.0 (Excoffier et al. 2005); (2) using the Bottleneck program v1.2.02 (Luikart and Cornuet 1997), we performed a Wilcoxon sign test ($\alpha = 0.05$) to determine if a significant number of loci featured an excess of heterozygosity, which is indicative of a bottleneck, assuming a two-phase mutation (TPM, model of microsatellite mutation) model and using 1000 permutations.

Results

Loci information and confidence in the SSR data

The analysis of linkage disequilibrium between loci pairs showed evidence of linkage only between loci 00756_c and 01078_a ($P = 0.165$) and loci 00756_c-02496_a ($P = 0.300$). The number of alleles per locus (A) varied from two at 588_a, 3313_a, and 1078_a loci, to four at 756_c and 280_a loci; minor allele frequency (MAF) varied from 0.085 to 0.332 (Table 1).

The percentage of missing data was of 0.7%. Three populations showed evidence of null alleles in one of their loci (1078_a locus in Atasta and Tibolon populations, and 2670_a locus in the Tibolon population); therefore, these loci were excluded from the final analysis. None of the analyzed populations showed evidence of stuttering or dropout alleles. The vast majority of alleles found had a difference in size of 5 or more base pairs from each other, a result that gives reliability to reading of the bands obtained by the silver staining method used in this work.

Genetic structure and grouping pattern

The value of delta K was four, suggesting the existence of only four genetic groups among the 13 populations sampled and identified by their source location. The analysis with the STRUCTURE showed that there were three groups within the wild gene pool: red, yellow, and green groups; all cultivated populations were grouped into the blue group (Fig. 1). 3D-PCoA supported the existence of the groups observed utilizing the STRUCTURE (Fig. 2). The first dimension explained 34.04% of the total variation, the second 24.21%, and the third 22.62%. Together, these three dimensions explained 80% of the total genetic variation. The N-J analysis also showed the presence of four groups (Fig. 3): nine individuals from the blue group were included in the yellow group, while 16 individuals from the yellow group were included in the blue group. Four individuals from the red group and only one from

the green group were included in the blue group. No individuals from the blue group were included in the red and green groups. Only one individual from the yellow group was included in the green group. These three analyses supported the closest genetic relationship between the blue group (cultivated) and the yellow group (wild). F_{ST} value between wild and cultivated gene pools was of 0.13. Among gene groups, F_{ST} was of 0.25: minor F_{ST} was between the yellow and blue groups (0.15), major F_{ST} was between the red and blue groups (0.36). Among populations, F_{ST} was of 0.36: minor F_{ST} was between Oxcutzcab- and Atasta-cultivated populations (0.07), major F_{ST} was between Palenque wild population and Tibolon-cultivated population (0.62). All P values obtained for the F_{ST} values were less than 0.05. Although very low, the BayesAss indicated a recent gene flow, mainly from the yellow to the blue group ($m = 0.03$).

Genetic diversity

The eight SSR loci used generated a total of 23 alleles. All the loci were polymorphic in the wild gene pool, two loci (588_a and 3313_a) were monomorphic in the cultivated gene pool. The estimators of genetic diversity are shown in Table 2. At the level of wild and cultivated gene pools, the allelic richness (A) and the expected heterozygosity (H_E) were higher in the wild gene pool ($A = 3.1$, $H_E = 0.41$) than in the cultivated gene pool ($A = 2.0$, $H_E = 0.38$). The observed heterozygosity (H_O) was very similar in both gene pools (wild, $H_O = 0.29$; cultivated $H_O = 0.30$). When all these diversity estimators were compared between wild and cultivated gene pools using permutation tests, the P values did not indicate significant differences between both gene pools in any of the three evaluated estimators (P , $A = 0.205$; $H_O = 0.103$; $H_E = 0.293$). At the group level, the highest A was observed in the green group ($A = 3.89$). The green group also showed the highest H_O (0.45) and H_E (0.47). At the population level, the greater H_O occurred in Palenque, Penjamo, and Dzidzantun ($H_O = 0.59$, 0.48, and 0.47, respectively), where Palenque and Penjamo were wild populations and Dzidzantun was cultivated. The highest H_E was found in Dzidzantun (cultivated) and Welib-Ha (wild) ($H_E = 0.49$ and 0.48, respectively). In relation to the possible existence of a bottleneck in Zapote mamey, the modified Garza–Williamson index (M) was very low but similar for the wild and cultivated gene pools, although the wild gene pool showed a slightly higher value ($M = 0.16 \pm 0.09$) than the cultivated gene pool ($M = 0.11 \pm 0.09$). The Bottleneck program indicated an excess of heterozygosity in both gene pools (wild, $P = 0.02$; cultivated, $P = 0.05$), but when this program was run at the group level, only the blue group (cultivated) showed an excess of heterozygosity.

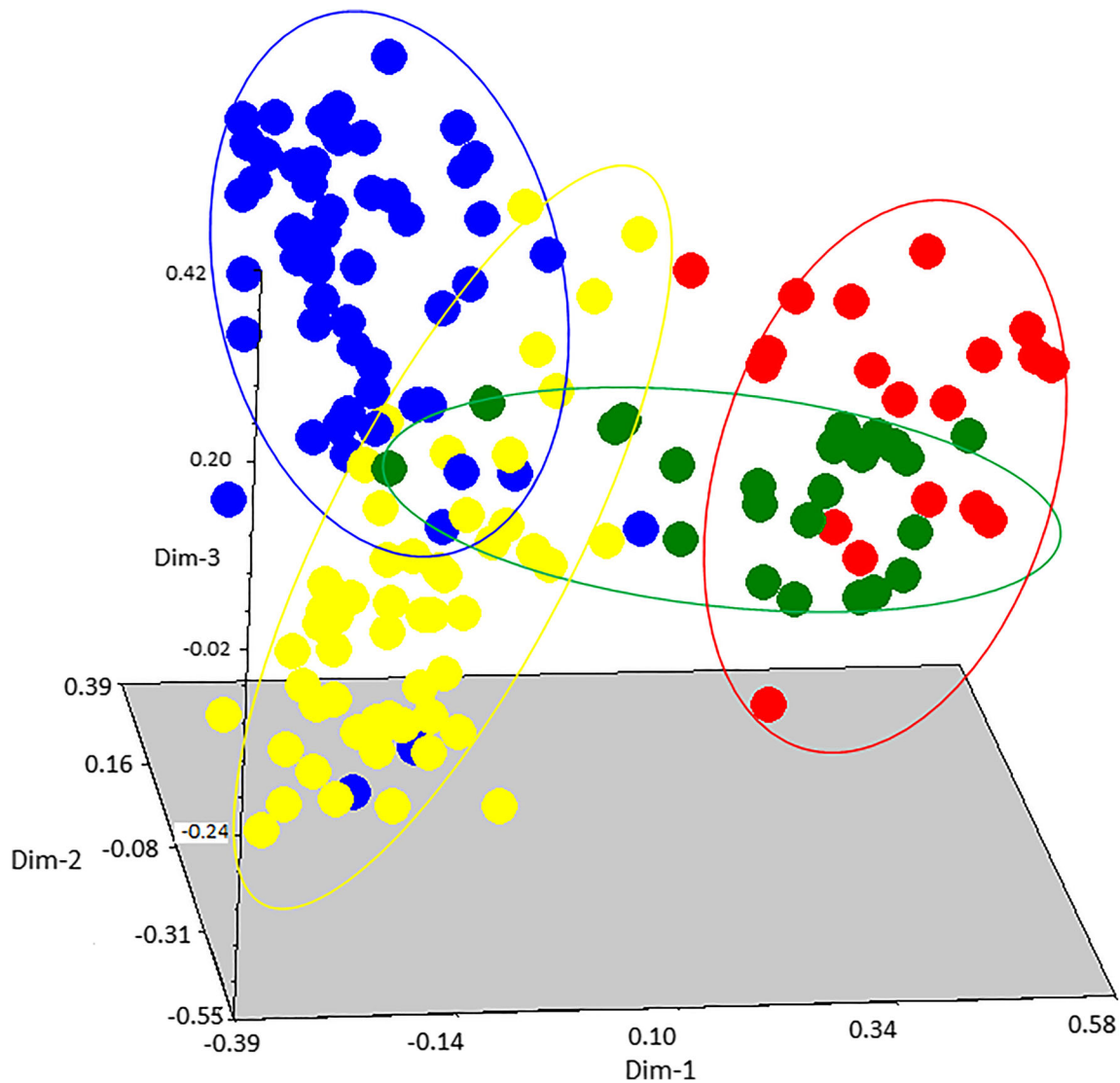


Fig. 2 Three-dimensional principal coordinates analysis (3-D PCoA) of 13 populations of *Pouteria sapota* from southeastern Mexico, using eight microsatellite loci. Wild populations: Yajalon (1), Ursula Galvan (2), Welib-Ha (3), Palenque (4), Jose Castillo (5), Agua Azul (6), Penjamo

(7), Nueva Palestina (8). Cultivated populations: Oxcutzcab (9), Atasta (10), Chekubul (11), Tibolon (12), Dzidzantun (13). Colors of the circles correspond to those showed in Figs. 1 and 3

Discussion

Genetic structure

Various methods (STRUCTURE, 3D-PCoA, N-J, F_{ST}) were used to analyze how the genetic diversity of Zapote mamey is distributed in southeastern Mexico. Four groups were identified according to their genetic distances, three in the wild gene pool, and one in the cultivated gene pool. Yajalon and Ursula Galvan (red group), Palenque (green group), and Jose Castillo (yellow group) wild populations were the ones with the lowest degree of admixture of the 13 populations studied, perhaps because these four populations were collected in highly conserved vegetation areas. In the wild gene pool, eight populations were collected from a relatively small geographic area

(in the states of Chiapas and Tabasco, Mexico) with heterogeneous environmental conditions, which could explain the resultant genetic variation in the three different genetic groups. Understanding the distribution of genetic diversity in the wild relatives of domesticated species is important to devise adequate sampling techniques and to develop effective management strategies (Hamrick 1983; Rao and Hodgkin 2002). All of the cultivated populations (blue group) were collected from the Yucatán Peninsula (in the states of Campeche and Yucatán), the primary region of Zapote mamey production in Mexico, where there are no wild populations reported of this species. This pattern of clustering of the cultivated populations could be explained by a common origin of individuals and their subsequent clonal propagation. Unfortunately, it was not possible to verify the specific origin of each individual

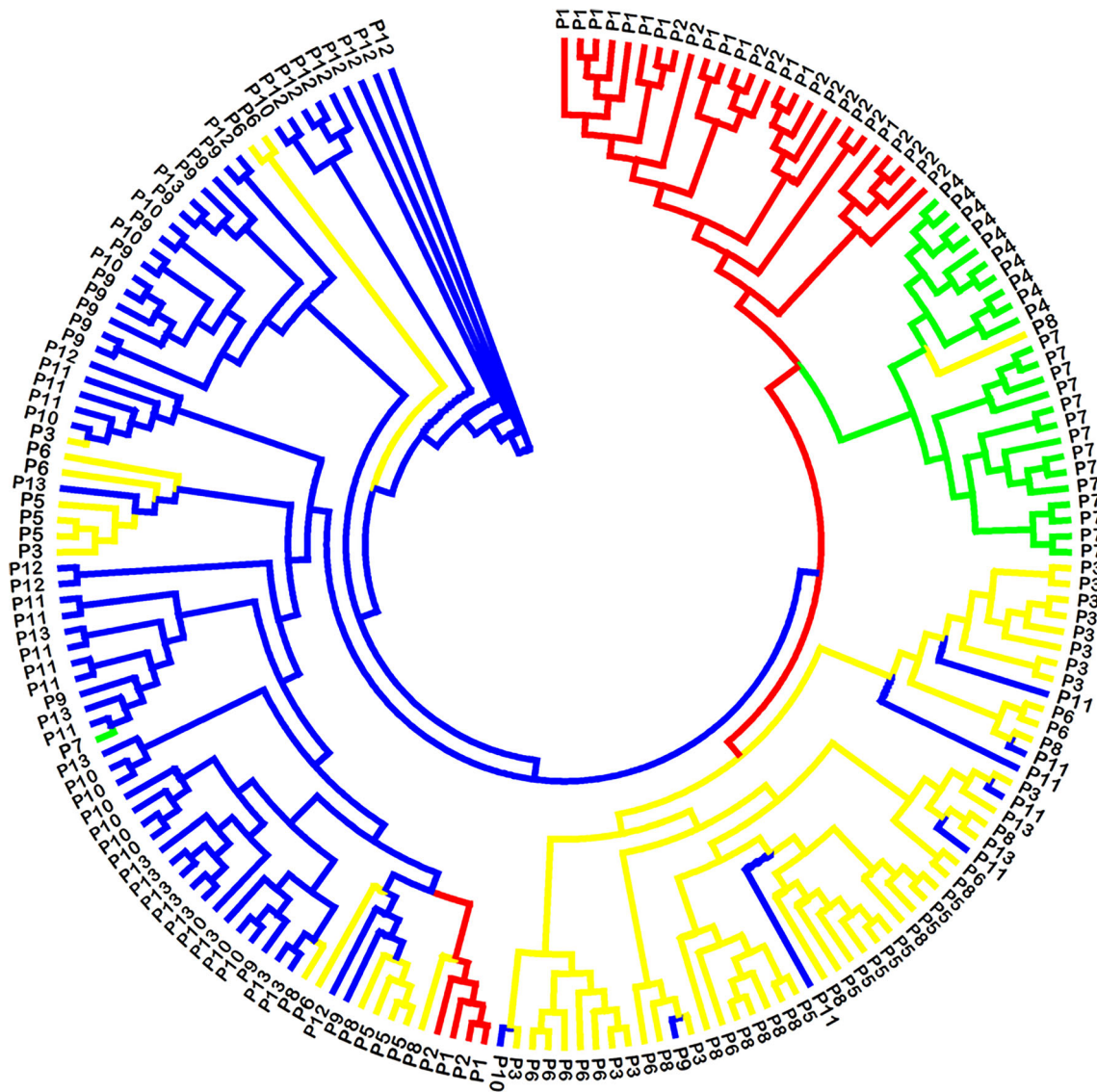


Fig. 3 Neighbor-joining (N-J) analysis of 13 populations of *Pouteria sapota* from southeastern Mexico, using eight microsatellite loci. Wild populations: Yajalon (1), Ursula Galvan (2), Welib-Ha (3), Palenque (4), Jose Castillo (5), Agua Azul (6), Penjamo (7), Nueva Palestina (8).

Cultivated populations: Oxcutzcab (9), Atasta (10), Chekubul (11), Tibolon (12), Dzidzantun (13). Colors of the lines correspond to those showed in Figs. 1 and 2

collected in the Yucatán Peninsula since the owners of the orchards from where the collections were made no longer remembered that origin, or the orchards had been planted by previous generations. Though, as indicated by Bruner and Morales-Payan (2011), reproduction of Zapote mamey by grafting is the typical method employed in small plantations.

We found low, but significant, genetic differences between wild and cultivated gene pools of Zapote mamey in southeastern Mexico. Arias et al. (2015), using the neighbor-joining and 3D-PCoA methods, reported no clear separation between wild and cultivated individuals of Zapote mamey. An important characteristic that can affect the genetic variation between cultivated species and their wild relatives is the mode of reproduction (Abbo et al. 2014). Zapote mamey can be propagated sexually or

asexually (Nava-Cruz and Ricker 2004). In domesticated species that can be clonally propagated, individuals with interesting traits can result in rapid rates of change because these ones can be selected and reproduced identically and in large numbers (McKey et al. 2010), favoring the genetic differentiation between gene pools. However, this reproductive characteristic can also limit the genetic differences between wild and cultivated plants if human groups continue to favor the gene flow through the movement of cuttings and the agricultural practice of grafting. Test of assignment of individuals (STRUCTURE) applied in the present study suggested the existence of a recent gene flow, primarily between the yellow wild group and the cultivated gene pool, finding was also supported for the PCoA and N-J analyses. But the BayesAss program showed that this gene flow is very

Table 2 Genetic diversity estimators of Zapote mamey (*Pouteria sapota*) from southeastern Mexico, using eight microsatellite loci

	<i>N</i>	<i>A</i>	$H_O \pm SD$	$H_E \pm SD$
Species	178	2.77	0.29 ± 0.06	0.39 ± 0.14
Gene pools				
Wild	115	3.10	0.29 ± 0.12	0.41 ± 0.19
Cultivated	73	2.00	0.30 ± 0.09	0.38 ± 0.17
Groups				
Red	30	1.98	0.30 ± 0.31	0.24 ± 0.17
Yellow	60	2.60	0.27 ± 0.09	0.37 ± 0.18
Green	25	3.89	0.45 ± 0.28	0.47 ± 0.20
Blue	73	2.00	0.30 ± 0.09	0.38 ± 0.17
Populations				
Yajalon	15	1.96	0.44 ± 0.34	0.32 ± 0.17
Ursula Galvan	15	2.00	0.33 ± 0.31	0.27 ± 0.17
Welib-Ha	15	1.95	0.23 ± 0.21	0.48 ± 0.17
Palenque	10	2.00	0.59 ± 0.35	0.39 ± 0.17
José Castillo	15	2.53	0.37 ± 0.21	0.35 ± 0.18
Agua Azul	15	1.99	0.33 ± 0.22	0.31 ± 0.17
Penjamo	15	1.99	0.48 ± 0.29	0.43 ± 0.15
Nueva Palestina	15	3.00	0.28 ± 0.17	0.28 ± 0.17
Oxcutzcab	15	2.00	0.34 ± 0.19	0.33 ± 0.15
Atasta	15	2.00	0.41 ± 0.14	0.42 ± 0.09
Chekubul	15	1.91	0.24 ± 0.23	0.30 ± 0.25
Tibolon	13	1.86	0.23 ± 0.17	0.25 ± 0.21
Dzidzantun	15	2.60	0.47 ± 0.11	0.49 ± 0.11

N, sample size; *A*, average number of alleles per locus calculated using rarefaction methods; H_O , heterozygosity observed; H_E heterozygosity expected; *SD*, standard deviation

low (less than 0.03) among all groups, although higher between yellow and blue groups. Although not conclusive, our results could be indicating that the ancestors of the Zapote mamey currently cultivated in the Yucatán Peninsula and sampled in this study originated from materials of the wild populations of the yellow group. Considering that there are no wild populations of Zapote mamey in the Yucatán Peninsula, the gene flow we observed was likely generated by an influx of seeds or cuttings, presumably taken from wild populations originating from the states of Chiapas, and transferred to the Yucatán Peninsula and maintained or multiplied through clonal propagation. Many perennial species are highly heterozygous (Petit and Hampe 2006); thus, while clonal propagation maintains the heterozygosity at the individual level, it promotes genetic homogeneity at the population level (Zohary and Spiegel-Roy 1975). This phenomenon was observed, in part, in our study, where some cultivated populations had high levels of heterozygosity (Dzidzantun, $H_E = 0.49$; $H_O = 0.47$) compared with many of the wild populations; however, all cultivated populations were part of the same genetic group (blue). Future studies emphasizing sampling at the population level, of both cultivated species and their wild relatives, as

considered in the present study, will shed light on the extent to which genes are transferred between wild and cultivated perennial crops clonally propagated.

Genetic diversity

In recent years, several studies within the Sapotaceae family have reported high levels of genetic diversity in wild species such as: *Argania spinosa* L. Skeels ($H_E = 0.67$, Yatrib et al. 2017); *Manilkara maxima* ($H_E = 0.67$, Ganzhorn et al. 2015); *Pouteria reticulata* (Engl.) Eyma ($H_E = 0.71$, Schroeder et al. 2014); *Manilkara zapota* ($H_E = 0.95$, González–Hernández et al. 2012); and *Manilkara huberi* (Ducke) Standl ($H_E = 0.86$, Azevedo et al. 2007; $H_E = 0.78$, Rennó 2007). These studies suggest that the existence of high levels of genetic diversity is a common pattern within the Sapotaceae family. However, our study showed low genetic diversity for wild *P. sapota* ($H_E = 0.41$), similar to that reported by Toledo (2009) for *Sideroxylon portoricense* subs. *minutiflorum* (Pittier) ($H_E = 0.35$), other species of Sapotaceae of Mexico. Arias et al. (2015) also found low levels of genetic diversity ($H_E = 0.35$) in a sample of 20 wild individuals of seven wild populations of *P. sapota* from southeastern Mexico. Although these authors analyzed a low number of individuals, they indicated that this low diversity is correctly estimated given the large number of loci (205) analyzed and the accuracy of allele calling (1 bp differences detected) when using capillary electrophoresis. It should be noted that the individuals used by these authors were also included in our work. A factor that can explain the low levels of genetic diversity found in the wild populations of Zapote mamey from southeastern Mexico is the existence of a bottleneck due to genetic drift caused by fragmentation of habitat due to natural and human causes. We observed a low Garza–Williamson index value for the wild gene pool of Zapote mamey ($M = 0.16$). As previously indicated, studies of several natural populations have shown that *M* is less than 0.70 for populations that have experienced a bottleneck event (Garza and Williamson 2001). Also, the Bottleneck program used in our study indicated an excess of heterozygosity in this gene pool, plausibly indicating a bottleneck (Luikart and Cornuet 1997). Finally, we found evidence of linkage disequilibrium between two pairs of loci. Although linkage disequilibrium can be affected by both biological and historical factors, one of these factors is the existence of a bottleneck (Gaut and Long 2003). So, all this evidence suggest the existence of a bottleneck in the wild gene pool of Zapote mamey of southeastern Mexico. This result is crucial when considering the development of conservation strategies for this species when the region where the wild populations of Zapote mamey were collected is a hot spot for potential species extinction as a result of urbanization, agriculture, and other land uses (De Baan et al. 2013).

We found higher levels of genetic diversity in the wild gene pool of Zapote mamey, compared with the cultivated one. However, permutation tests did not indicate significant differences between both gene pools. Arias et al. (2015) found that the cultivated gene pool of Zapote mamey had slightly higher genetic diversity (9 individuals, $H_E = 0.41$) than its wild counterpart (20 individuals, $H_E = 0.35$). Miller and Gross (2011) emphasized that higher levels of genetic diversity in cultivated populations, relative to their wild counterparts, could be due to (a) insufficient sampling of wild populations; (b) cultivated populations that represent the descendants of controlled crosses between geographically and genetically distinct individuals, yielding new variants carrying novel combinations of alleles, not found in the wild; (c) a loss, or extinction, of wild plants, following the establishment of cultivated populations due to habitat destruction or other reasons; and (d) somatic mutations in clonally propagated cultivars which contribute to elevated levels of genetic variation. Another factor to explain our results is that Zapote mamey is in a process of incipient domestication, which would help explain, in part, not having found significant differences in genetic diversity between the wild and the cultivated gene pools. Although in this work we use only 8 loci, these loci were expected to be polymorphic and detect variations. The Garza–Williamson index and Bottleneck program indicated a bottleneck in the cultivated gene pool of Zapote mamey from southeastern Mexico. This finding is supported by the report of Arias et al. (2015) that (a) using the Bottleneck software, they found a reduction in genetic diversity in the cultivated gene pool; (b) at 62 SSR loci (of 205 analyzed), one or more alleles were observed in the wild individuals that were not detected in cultivated individuals. Considering the strength of the population sampling of our study and the robustness of the genomic sampling of Arias et al. (2015), all available evidence supports the existence of a weak, but significant, bottleneck in the cultivated Zapote mamey of southeastern Mexico.

Conclusions

Our study provides the first analysis at a population level about the structure and genetic diversity of Zapote mamey incorporating a broad sampling of wild and cultivated individuals collected from southeastern Mexico, its putative domestication area, and characterized using specific codominant SSR markers. We found genetic differentiation between both gene pools and evidence of a low recent gene flow between cultivated populations from Yucatán Peninsula and wild populations from Chiapas State. Our results indicated low levels of genetic diversity in wild and cultivated populations of Zapote mamey, but did not demonstrate significant differences between both gene pools. The information provided in this study can be used to determine adequate sampling techniques and to develop effective management strategies for Zapote mamey of southeastern Mexico.

Acknowledgments This research was done with the assistance of the Molecular Markers Laboratory, in the Natural Resources Department, of CICY. The authors thank Paulino Sima for his dedicated fieldwork assistance, Dr. Ruben Andueza for doing the N-J analysis, and Laura Mills for providing the English revision.

Data archiving statement The above research work does not involve any novel molecular markers and hence no data or sequences have been submitted to public databases.

Funding information The second author wishes to thank CONACYT-Mexico for their generous Master's scholarship and the academic advice of Dr. Javier Mijangos, Dr. Eduardo Morales, and Dr. Rafael Durán.

References

- Abbo S, Phinasi van-Oss R, Gopher A, Saranga Y, Ofner I, Peleg Z (2014) Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends Plant Sci* 19: 351–360
- Arellano-Durán N (2012) Caracterización morfológica, estudio molecular y determinación de azúcares en zapote mamey (*Pouteria sapota* (Jacq.) H.E. Moore & Stearn). Dissertation, Colegio de Postgraduados
- Arias RS, Martínez-Castillo J, Sobolev VS, Blancarte-Jasso NH, Simpson SA, Ballard LL, Duke MV, Liu XF, Irish BM, Scheffler BE (2015) Development of a large set of microsatellite markers in Zapote mamey (*Pouteria sapota* (Jacq.) H.E. Moore & Stearn) and their potential use in the study of the species. *Molecules* 20:11400–11417
- Awodoyin R, Olubode O, Ogbu J, Balogun R, Nwawuisi J, Orji K (2015) Indigenous fruit trees of tropical Africa: status, opportunity for development and biodiversity management. *Agric Sci* 6:31–41
- Azevedo VCR, Kanashiro M, Ciampi AY, Grattapaglia D (2007) Genetic structure and mating system of *Manilkara huberi* (Ducke) A. Chev., a heavily logged amazonian timber species. *J Hered* 98:646–654
- Azurdía C (2006) Tres especies de Zapote en América Tropical (*Pouteria campechiana*, *P. sapota*, *P. viridis*). Universidad de Southampton, Southampton
- Balerdi CF, Crane JH (2015) El mamey sapote en Florida. Miami: the Institute of Food and Agricultural Sciences-Extension. University of Florida, USA
- Bañuelos-Jimenez JS, Ochoa I (2006) Caracterización morfológica de zapote mamey (*Pouteria sapota* Jacq. H.E. Moore & Stearn) del centro occidente de Michoacán, México. *Rev Fitotec Mex* 29:9–17
- Bassam BJ, Anollés GC, Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem* 196:80–83
- Brewer SW, Rejmanek M (1999) Small rodents as significant dispersers of tree seeds in a Neotropica forest. *J Veg Sci* 10(2):165–174
- Bruner BR, Morales-Payan JP (2011) Soils, plant growth and crop production - Sapote, Sapodilla and Star apple. In: *Encyclopedia of life support systems*. Paris, France: UNESCO. <http://www.colss.net/sample-chapters/c10/e1-05a-53-00.pdf>. (Accessed 13 January 2018)
- Carrara S, Campbell R, Schnell R (2004) Genetic variation among cultivated selections of mamey sapote (*Pouteria* spp. [Sapotaceae]). *P Fl St Hortic Soc* 117:195–200
- De Baan L, Mutel CL, Curran M, Hellweg S, Koellner T (2013) Land use in life cycle assessment: global characterization factors based on regional and global potential species extinction. *Environ Sci Technol* 47:9281–9290
- Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4(2):359–361

- Espinosa-Zaragoza S, Saucedo VC, Villegas MA, Ibarra EME (2005) Caracterización de frutos de zapote mamey (*Pouteria sapota* (Jacq.) H. E. Moore & Stearn) en Guerrero, México. *Proc Interamerican Soc Trop Hortic* 48:135–138
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinforma* 1:47–50
- Ganzhorn SM, Thomas WW, Gaiotto FA, Lewis JD (2015) Spatial genetic structure of *Manilkara máxima* (Sapotaceae), a tree species from the Brazilian Atlantic forest. *J Trop Ecol* 31:437–447
- Gaona-García A, Alia-Tejagal I, López-Martínez V, Andrade-Rodríguez M, Colinas-León MT, Villegas-Torres O (2008) Caracterización de frutos de zapote mamey (*Pouteria sapota*) en el suroeste del estado de Morelos. *Revi Chapingo Ser Hortic* 14:41–47
- Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Mol Ecol* 10:305–318
- Gaut BS, Long AD (2003) The lowdown on linkage disequilibrium. *Plant Cell* 15:1502–1506
- González-Hernández D, García-Pérez E, Guntin-Marey P (2012) Genetic characterization of *Manilkara zapota* from Veracruz, México, with SSR markers. *Agrociencia* 46:663–675
- Gordon A, Jungfer E, Da Silva BA, Maia JGS, Marx F (2011) Phenolic constituents and antioxidant capacity of four underutilized fruits from the Amazon region. *J Agric Food Chem* 59:7688–7699
- Goudet J (2002) Fstat version 1.2: a program to estimate and test gene diversities and fixation indices. *J Hered* 86: 485–486.
- Gulyas-Fekete G, Murillo E, Kurtan T, Papp T, Illyes T-Z, Drahos L, Visy J, Agocs A, Turcsi E, Deli J (2013) Cryptocapsinopoxide-type carotenoids from red mamey, *Pouteria sapota*. *J Nat Prod* 76:607–614
- Hamrick JL (1983) The distribution of genetic variation within and among natural plant populations. In: SchonewaldCox CM, Chambers SM, MacBryde B, Thomas L (eds) *Genetics and conservation*. Benjamin-Cummings, London, pp 335–348
- Haq N, Bowe C, Dunsiger ZE (2008) Challenges to stimulating the adoption and impact of indigenous fruit trees in tropical agriculture. In: Akinnifesi FK, RRB L, Ajayi OC, Sileshi G, Tchoundjeu Z, Matakala P, Kwesiga FR (eds) *Indigenous fruit trees in the tropics: domestication, utilization and commercialization*. CAB International, Wallingford, pp 50–69
- Ibarra-Estrada ME (2012) Morfología de frutos, perfil de ácidos grasos en semilla y diversidad genética molecular en selecciones de zapote mamey (*Pouteria sapota* (Jacq.) H. E. Moore & Stearn). Dissertation, Colegio de Postgraduados, México
- Jamnadas R, Lowe A, Dawson IK (2009) Molecular markers and the management of tropical trees: the case of indigenous fruits. *Trop Plant Biol* 2(1):1–12
- Kalinowski ST (2005) HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Mol Ecol Notes* 5: 187–189
- Knight RJ, Campbell CW, Campbell RJ (1993) Pollination requirements for successful fruiting of tropical fruit species. *Proc Interam Soc Trop Hortic* 37:167–170
- Langella O (2002) Populations 1.2.28: Population genetic software: individuals or populations distances based on allelic frequencies, phylogenetic trees, file conversions. Available at: http://bioinformatics.org/project/?group_id=84. Accessed 15 Mar 2019
- León J (1987) Sapotáceas. Botánica de los cultivos tropicales. Instituto Interamericano de Cooperación para la Agricultura. Second edn. San José de Costa Rica, pp 208–216
- Luikart G, Cornuet JM (1997) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conserv Biol* 12:228–237
- McKey D, Elias M, Pujol B, Duputié A (2010) The evolutionary ecology of clonally propagated domesticated plants. *New Phytol* 186:318–332
- Martínez-Gallardo R, Sánchez-Cordero V (1997) Historia natural de algunas especies de mamíferos terrestres. In: González SE, Dirzo R, Vogt R (eds) *Historia natural de los Tuxtlas*. Universidad Nacional Autónoma de México, México, pp 25–31
- Miller AJ, Gross BL (2011) From forest to field: perennial fruit crop domestication. *Am J Bot* 98:1389–1414
- Moo-Huchin V, Estrada-Mota I, Estrada-León R, Cuevas-Glory LF, Sauri-Duch E (2013) Chemical composition of crude oil from the seeds of pumpkin (*Cucurbita* spp.) and mamey sapota (*Pouteria sapota* Jacq.) grown in Yucatan, Mexico. *CyTA - J Food* 11:324–327
- Murillo E, MClean R, Britton G, Agocs A, Nagy V, Deli J (2011) Sapotexanthin, an A-provitamin carotenoid from red mamey (*Pouteria sapota*). *J Nat Prod* 74:283–285
- Nascimento VE, Martins ABG, Hojo RH (2008) Caracterização física e química de fruto de mamey. *Rev Bras Frutic* 30:953–957
- Nava-Cruz Y, Ricker M (2004) El Zapote mamey (*Pouteria sapota* (Jacq.) H. Moore y Stearn), un fruto de la selva Mexicana con alto valor comercial. In: Alexiades MN, Shanley P (eds) *Productos Forestales, Medios de Subsistencia y Conservación: Estudios de Caso Sobre Sistemas de Manejo de Productos Forestales no Maderables*, vol 3. Center for International Forestry Research (CIFOR), Bogor, pp 43–62
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York
- Normah MN, Malik SK, Chaudhury R, Salma I, Makeen MA (2013) Conservation of tropical fruit genetic resources. In: Normah M, Chin H, Reed B (eds) *Conservation of tropical plant species*. Springer, New York, pp 137–170
- Ortiz SFJ, Cabello GT (1991) Use of insects in the pollination of subtropical crops. *Agrícola Vergel* 10(119):692–695
- Paul RE, Duarte O (2012) *Tropical fruits*. CAB International, London
- Pennington TD (1990) *Flora neotropical monograph 52: Sapotaceae*. The New York Botanical Garden, Bronx
- Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annu Rev Ecol Evol Syst* 37:187–214
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Rao RV, Hodgkin T (2002) Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell Tiss Org* 68:1–19
- Rennó AVC (2007) Desenvolvimento e aplicações de microsatélites, análise cpDNA e modelagem computacional para estudos da estrutura e dinâmica genética de macaranduba –*Manilkara huberi* (Ducke) Standl. Sapotaceae. Dissertation. Universidade de Brasília, Brasil
- Rodríguez-Rojas T, Andrade-Rodríguez M, Alia-Tejagal I, López-Martínez V, Espinosa-Zaragoza S, Esquinca-Avilés H (2012) Caracterización molecular de Zapote mamey (*Pouteria sapota* (Jacq.) Moore & Stearn). *Rev Fac Agron Luz* 29:339–354
- Rohlf FJ (2001) NTSYS-pc: numerical taxonomy and multivariate analysis system, version 2.1. Exeter Publishing, Setauket
- SIAP (2015) Anuario estadístico de la producción de los Estados Unidos Mexicanos. Secretaría de Agricultura, Ganadería y Desarrollo Rural, México. www.siap.gob.mx. Accessed 15 December 2017
- Schreckenber K, Awono A, Degradé D, Bosso CM, Ndoye O, Tchoundjeu Z (2006) Domesticating indigenous fruit trees as a contribution to poverty reduction. *For Tree Livelihoods* 16:35–51
- Schroeder JW, Tran HT, Dick CW (2014) Fine scale spatial genetic structure in *Pouteria reticulata* (Engl.) Eyma (Sapotaceae), a dioecious, vertebrate dispersed tropical rain forest tree species. *Global Ecol Conserv* 1:43–49
- Slatkin M, Excoffier L (1996) Testing for linkage disequilibrium in genotypic data using the EM algorithm. *Heredity* 76:377–383

- Tamura K, Peterson D, Filipski A, Kumar S (2013) MEGA 6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729
- Toledo CL (2009) Efecto de la fragmentación en la estructura genética y en características de historias de vida de *Sideroxylon portoricense* (Sapotaceae) en la selva de los Tuxtlas Veracruz, México. Dissertation, UNAM, México
- Van Oosterhout TC, Hutchinson WF, Wills DP, Shipniew P (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4:535–538
- Villarreal-Fuentes JM, Alia-Tejagal I, Hernández E, Pelayo-Zaldivar C, Franco-Mora O (2015) Caracterización poscosecha de selecciones de zapote mamey (*Pouteria sapota* (Jacq.) H. E. Moore & Stearn) procedentes del Soconusco, Chiapas. *Ecosist Rec Agropec* 2(5): 217–224
- Villegas-Monter A, Escobar-Sandoval CM, Arrieta-Ramos G, Berdeja-Arbeu R (2016) Zapote mamey [*Pouteria sapota* (Jacq.) Moore & Stearn], diversity and uses. *AgroProductividad* 9(4):47–54
- Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163:1177–1191
- Yatrib C, Belkadi B, Medraoui L, Pakhrou O, Alami M, El Mousadik A, Ferradous A, Msanda F, El Modafar C, Souda-Kouraichi SI, Filali-Maltouf A (2017) Genetic diversity and population structure of the endangered argan tree (*Argania spinosa* L. Skeels) in Morocco as revealed by SSR markers: implication for conservation. *Aust J Crop Sci* 11(10):1304–1314
- Zohary D, Spiegel-Roy P (1975) Beginnings of fruit growing in the old world. *Science* 187:319–327

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.