

RESEARCH ARTICLE

Analysis of the biodiversity of hawthorn (*Crataegus* spp.) from the morphological, molecular, and ethnobotanical approaches, and implications for genetic resource conservation in scenery of increasing cultivation: the case of Mexico

Marcela Betancourt-Olvera · Raúl Nieto-Ángel · Beatriz Urbano  · Fernando González-Andrés 

Received: 19 July 2017 / Accepted: 3 November 2017 / Published online: 23 November 2017
© Springer Science+Business Media B.V., part of Springer Nature 2017

Abstract One hundred and forty Mexican hawthorn (*Crataegus* spp.) accessions from six regions are preserved at the BGT-UACH germplasm bank (Mexico), comprising the most comprehensive living collection of Mexican hawthorns with different degrees of human management. The objective of this study was to assess the biodiversity of this valuable collection from morphological, molecular (microsatellite), and ethnobotanical viewpoints in order to delineate the most adequate strategy for the conservation of the native hawthorn germplasm in the present scenario of incipient establishment of

commercial hawthorn plantations, which is likely to increase. Molecular characterisation revealed that the biodiversity was chiefly (90%) placed within the regions. Morphological characterisation indicated that the group from Chiapas was the most different germplasm pool compared with the other five. This was confirmed by molecular analysis, because in spite of the lack of a phylogeographical pattern, two germplasm pools were detected: one composed mainly by accessions from Chiapas and the other mainly by accessions from the other regions. The only clear differences among the regions in the ethnobotanical study were those derived from putting hawthorns into commercial cultivation, which occurred in just one region in the centre of the country (Mexico–Puebla–Tlaxcala). As a consequence, an *ex situ* conservation programme is necessary for those regions shifting patterns of cultivation from traditional to commercial, regardless of whether other *on-farm* programmes are also implemented. The germplasm collections within each region must be exhaustive due to their high genetic diversity.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10722-017-0583-4>) contains supplementary material, which is available to authorized users.

M. Betancourt-Olvera · R. Nieto-Ángel
Departamento de Fitotecnia, Universidad Autónoma
Chapingo, Km. 38.5 carretera México-Texcoco,
Chapingo C.P 56230, Estado de México, Mexico

B. Urbano
Área de Economía, Sociología y Política Agraria,
Departamento de Ingeniería Agrícola y Forestal,
Universidad de Valladolid, Campus Palencia, Avenida de
Madrid 57, 34004 Palencia, Spain

F. González-Andrés (✉)
Instituto de Medio Ambiente, Recursos Naturales y
Biodiversidad, Universidad de León, Avenida de Portugal
41, 24009 León, Spain
e-mail: fernando.gonzalez@unileon.es

Keywords *Crataegus* · Genetic erosion · Germplasm conservation · SSR · STMs · Tejocote

Introduction

The plants from the genus *Crataegus* L., commonly named hawthorns, are distributed in the temperate regions of the northern hemisphere (Phipps et al. 2003). More than 250 species are reported for this genus (Evans and Campbell 2002), making it one of the largest genera in the Rosaceae tribe Pyreae (Campbell et al. 2007). Many of the species of the genus *Crataegus* are of local interest from the ethnobotanical or environmental viewpoint: Their medicinal use is well documented from ancient times among the native American peoples (Edwards et al. 2012) and in traditional Chinese medicine (Edwards et al. 2012; Chen et al. 2013); the fruits of some species are used as human food in several cultures (Edwards et al. 2012), and some species are of ornamental interest (Campbell et al. 2007). From the environmental viewpoint, *Crataegus* also play an important role in different ecosystems and agrosystems, where they are plants of high ecological value (Rahmani et al. 2015; Brown et al. 2016).

In Mexico, the hawthorn is named ‘tejocote’, a Spanish term derived from the Nahuatl word ‘Texocotl’ (stone fruit) (Cabrera 1992) in reference to the petrous endocarp. Hawthorns have a strong cultural link with Mexicans since the pre-Hispanic time, when the fruits were consumed as food and the trees were used as decoration in Indian villages (Núñez-Colín et al. 2008, 2009; Zizumbo-Villarreal et al. 2016). At the present time, hawthorns are considered an under-utilised crop (Nieto-Ángel 2007; Lozano-Grande et al. 2016), although they are part of the contemporaneous Mexican culture, as they are used not only as food (fresh or processed fruits) or decoration, but also in traditional celebrations, as on All Saints’ Day and Christmas, mainly in the production of fruit-based beverages and as treats inside piñatas (Borys and Leszczyńska-Borys 1994; Núñez-Colín et al. 2008, 2009). Despite its cataloguing as an under-utilised crop, 900 hectares of hawthorn have been established in commercial orchards in Mexico during the last few years, mainly in the states of Puebla and Mexico (SAGARPA 2015), and this surface is currently increasing with new plantations. As a consequence, there is also incipient industrial use of hawthorns for the pectin industry due to the high quality of the pectins extracted from the fruit (Wang et al. 2007; Lozano-Grande et al. 2016). Moreover,

extraction of antioxidants from the plant for industrial purposes is now under research (Banderas-Tarabay et al. 2015). Two varieties of hawthorn belonging to *Crataegus mexicana* Moc. et Sessé have already been developed in Mexico (SNICS 2012; Velasco-Hernández et al. 2017) and used in the new commercial orchards. With the increase of commercial activity, the hawthorn biodiversity is at risk of reduction, as has happened for several crops (Hammer 2004; Stamp et al. 2012). Therefore, germplasm banks need to play an important role in its conservation.

The ‘Banco de germoplasma de tejocote de la Universidad Autónoma Chapingo’, Mexico (BGT-UACH), is a hawthorn living *ex situ* germplasm bank that consists of 140 nuclear accessions representing native Mexican hawthorns (*Crataegus* spp.), which were collected in Central Mexico and Chiapas, the regions with the strongest cultural link with this species (Nieto-Ángel 2007). This germplasm bank is composed by accessions corresponding to the following biological statuses according to the passport data (Alercia et al. 2015): semi-natural (either wild or sown), or traditional cultivar/landrace. The germplasm collection is a sound representation of the Mexican hawthorns belonging to the following management statuses according to Mapes and Basurto (2016): tolerated and favoured plants; favoured plants or cultivated landraces.

The starting hypothesis is that the anthropic use of plant species such as favoured and tolerated plants, affects to the genetic structure of populations, increasing the intrapopulation biodiversity, with implications in the conservation programmes.

The objective of this work was to assess the variability of the Mexican hawthorns conserved at the BGT-UACH from the biological (morphological and molecular) and ethnobotanical viewpoints in order to understand the best strategies to conserve the biodiversity of hawthorn in the present scenario, in which commercial orchards are being introduced, with the risk of losing the existing biodiversity with the introduction of improved varieties.

Materials and methods

Germplasm sampling

The 140 accessions belonging to the BGT-UACH were collected in the following states: Chiapas, Mexico, Puebla, Michoacan, Morelos, and Tlaxcala. These states encompass the geographical areas of Mexico where hawthorns are important from the ethnobotanical viewpoint. The germplasm bank consists of three genetically identical individuals per accession. Figure 1 georeference the geographical origin of the germplasm. The most relevant passport data of the accessions are shown in Online Resource 1, which includes the collection site with coordinates, the putative species according to Phipps (1997) and Eggleston (1909) (Núñez-Colín and Hernández-

Martínez 2011), and the biological status of the accessions, whether they are semi-naturalised, wild, or a traditional cultivar/landrace according to the description of Alercia et al. (2015).

A priori definition of germplasm groups based on a geographical pattern

The 140 accessions were clustered into six geographical groups defined on the basis of a geographical pattern based on the proximity between the collecting places of the accessions, and the isolation from the rest of the accessions (Fig. 1 and Online Resource 1). The altitudinal variable was also considered for the definition of the groups. The nomenclature of the geographical groups corresponds to the official abbreviation of the Mexican states where they were

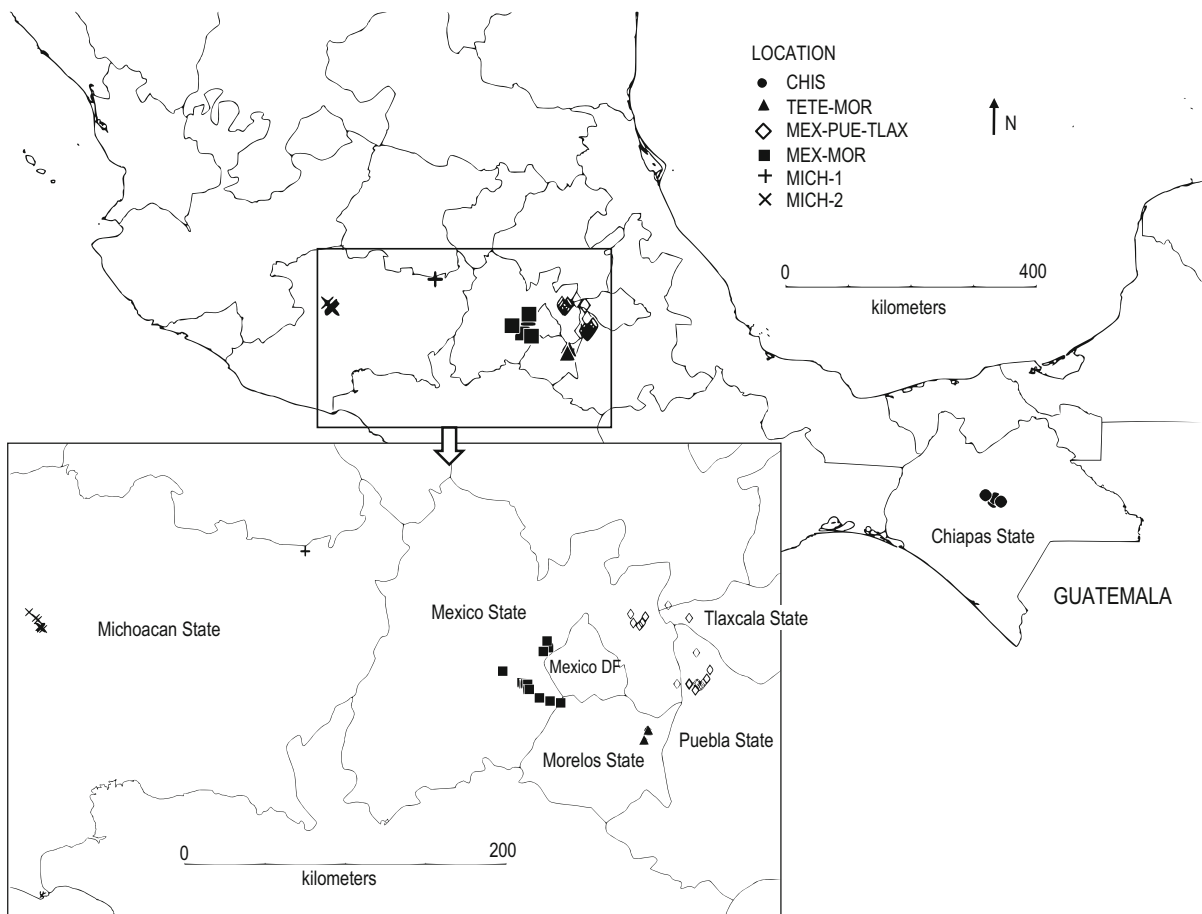


Fig. 1 Origin of the hawthorn (*Crataegus* spp.) accessions used in this work obtained from the ‘Banco de germoplamsa de Tejocote de la Universidad Autónoma Chapingo’ (BGT-

UACH), Mexico. Locations were grouped into six geographical groups, considered a priori as germplasm pools

collected: Pool 1 CHIS (Chiapas), Pool 2 Tete-MOR (Tetela del Volcán in the state of Morelos), Pool 3 MEX-PUE-TLAX (Mexico, Puebla and Tlaxcala), Pool 4 MEX-MOR (Mexico and Morelos), Pool 5 MICH-1 (East Michoacan), and Pool 6 MICH-2 (West Michoacan). All the groups are clearly defined from the visual viewpoint, except for groups 2 (Tete-MOR) and 3 (MEX-PUE-TLAX), which were considered as different because the accessions of group 2 were collected in the lowest altitudes of the germplasm collection. The consistency of such groups defined a priori was tested using the morphological, molecular, and ethnobotanical traits.

Morphological characterisation

Morphological traits

Seventy-two morphological traits from all organs of the plant and throughout its life cycle were used for the morphological characterisation (Online Resource 2). The characteristics were selected on the basis of their relevance for the taxonomy of the genus *Crataegus*, according to the key of Phipps (1997), and the hawthorn descriptors list (UPOV 2008). They were in accordance with the general recommendations for the selection of descriptors concerning objectivity and consistency. For each accession, 20 mature leaves per each type of shoot (reproductive, long vegetative, and short vegetative) were collected for the analysis. Each leaf belonged to a different shoot, located in an intermediate vertical position in the tree and at the four cardinal points. Ten flowers, including a pedicel per accession, were collected, and the different anatomic parts of the flower were separated and scanned for the analysis. The receptacle, calyx, and pedicel were scanned together. Twenty mature fruits were collected from the four cardinal points of the trees and used for the analysis. The endocarps were sampled and prepared for measurements according to Nieto-Angel et al. (2009). The sampling was repeated in 2014 and 2016, and one average value was obtained per descriptor and per accession. In such a way, the average values used for the statistical analysis included the potential inter-annual variation.

The samples were processed immediately after collection and digitalised for measurements. The images were enhanced with image-processing software, and the measurements in the digital photographs

were obtained with the image analysis freeware Image Tool 3.0 (Wilcox et al. 1995). Prior to the multivariate analysis of the results, in order to estimate the overall variation for each trait, the coefficient of variation (%) as the standard deviation \times mean value⁻¹ \times 100 was calculated for each trait and accession.

Multivariate analysis

Three different multivariate analyses were performed for the 140 accessions, based on the 72 morphological traits. Firstly, a data matrix was prepared with mean values of the 2 years for every accession and standardised by subtracting the mean value of each trait and dividing the result by the standard deviation.

For the cluster analysis, a consensus similarity matrix between accessions was calculated using the Euclidean distance after resampling based on 1,000 bootstraps. Clustering was based on the Ward method as it is adequate for clustering of elements which are maximally similar with respect to specified characteristics (Ward, 1963), as it is the case of Mexican *Crataegus* as previously indicated by Núñez-Colín et al. (2008). The obtained dendrogram was used for the selection of the accessions to be characterised by microsatellite markers.

The other two multivariate analyses with the whole set of accessions consisted of a principal components analysis (PCA) and a canonical discriminant analysis (CDA). For the PCA, a similarity matrix among the morphological traits was calculated based on the Pearson product–moment correlation. The PC obtained were rotated using the Varimax orthogonal rotation method, and the resultant PC was used for the accessions projection. For the CDA, the dependent variable was the germplasm groups defined a priori based on the geographical pattern of distribution. The analysis included the calculation of the Mahalanobis distance and a resubstitution test. The cluster and the PCA analysis were carried out with Ntsys-pc (Rohlf 2008), and the CDA was performed with the software IBM SPSS Statistics (IBM Corp. Released 2016).

Molecular characterisation

Selection of accessions for the microsatellite analysis

Fifty-six accessions from five (CHIS, MEX-PUE-TLAX, MEX-MOR, MICH-1, MICH-2) out of the six

geographical groups were selected for the microsatellite analysis based on the dendrogram obtained for the 140 initial accessions and the 72 morphological traits (Online Resource 3). The dendrogram was cut at a dissimilarity level of 5, and in the first three branches 40% of the accession within each of the branches was randomly selected and used for the microsatellite analysis. Conversely, in the fourth branch, due to the higher internal diversity observed in the aggrupation pattern, 60% of the accessions were randomly selected. Three landraces of fruit trees with a previously known microsatellite profile were used as controls in each run to ensure size accuracy and to minimise run-to-run variation and that of outliers: one of apple (*Malus domestica* Borkh.), another of pear (*Pyrus communis* L.), and the third of quince (*Cydonia oblonga* Mill.), all from San Miguel Tulancingo (Oaxaca, Mexico), belonging to the UACH germ-plasm bank. Moreover, an accession of *Crataegus pinnatifida* Bunge of Chinese origin, donated by Dr Leszek S. Jankiewicz (Research Institute of Pomology Skierniewice, Poland), was also used as an outlier.

DNA extraction, amplification, and sequencing of microsatellites

The DNA was extracted from young leaves using a modified version of the CTAB-based protocol of Doyle and Doyle (1990), after discarding commercial kits due to the lack of successful amplification (Betancourt-Olvera et al. in press). Fifteen mg of young-leaf samples were lyophilised with a Labconco freeze dryer Freezone 4.5 and grounded with a TissueLyser II mill from Qiagen. The grounded tissue was mixed with 600 μ L of buffer (10 mM Tris–HCl, 1 mM EDTA 2 M NaCl, 0.05% serum albumin, pH 8.0), vortexed until the mixture was homogenised, and incubated for 15 min at room temperature. Subsequently, it was centrifuged at 5,000 g for 5 min, and the supernatant was discarded. Then, 600 μ L of the extraction buffer (100 mM Tris–HCl, pH 8.0, 20 mM EDTA, 1.4 M NaCl, 2.0% CTAB, pH 8.0) supplemented with 3 μ L of β -mercaptoethanol was added, vortexed, and incubated at 55° C for 60 min. Subsequently, 400 μ L of chloroform:isoamyl alcohol (24: 1) was added, following the protocol of Doyle and Doyle (1990).

Microsatellite markers tested were CH05D04, CH03D08, CH04G04, CH03A02, CH05G07,

CH01F02, and CH04F06, designed for apple tree (*Malus domestica* L.) (Liebhard et al. 2002) and tested in *Crataegus* sp. with positive results by Lo et al. (2009) and Dickinson et al. (2008). PCR amplifications were performed on a Mastercycler Eppendorf thermocycler in a volume of 25 μ L containing 0.8 ng/ μ L (20 ng per reaction) of genomic DNA, 200 mM of each dNTP, 1 X buffer, 1.5 mM MgCl₂, 1.5 U of Taq DNA polymerase, recombinant from Invitrogen™, and 0.3 mM each of the primers from Sigma-Aldrich. The forward primers were end-labelled with the following fluorescent dyes: CH03A02, CH04G04, CH05D04, CH01F02, and CH04F06 with 6-FAM; CH03D08 and CH05G07 with HEX. Two multiplex PCRs were designed as A: CH05D04 + CH03D08; B: CH03A02 + CH05G07—and the other three microsatellite loci were processed in simplex reactions. The PCR temperature profiles were as follows: For the multiplex reactions, an initial denaturation step at 94 °C for 2 min 30 s, followed by 30 cycles at 94 °C for 30 s, 55 °C for 1 min, 72 °C for 1 min, and a final step at 72 °C for 5 min. CH04G04 and CH04F06 were modified, as the annealing temperature for the 30 cycles was 62 °C and for CH01F02 60 °C.

PCR products were analysed on an ABI 3730XL automatic sequencer (Applied Biosystems) from Macrogen USA. The peaks were scored using the GeneMapper version 3.5 programme (Applied Biosystems) and corroborated with Peak Scanner v.1.0.

Microsatellite marker analysis

The allelic profile for each accession was determined for each microsatellite locus. As the hawthorn accessions analysed showed different ploidy levels, SPA-GeDI v1.2 g (Hardy and Vekemans 2002) was used to compute genetic information statistics, as this software supports analysis of datasets containing individuals with different ploidy levels (Ferreira et al. 2016).

First, the genetic diversity indexes were calculated for the whole set of accessions: The number of observed genotypes (Go), the number of alleles per locus (A), the effective number of alleles per locus (Ae), the observed (Ho) and the expected (He) heterozygosity, the polymorphism information content (PIC), the probability of coincidence (C), the discrimination power (D), and the inbreeding coefficient (Fis). Subsequently, the observed and the expected heterozygosity were calculated for each of

the geographical groups defined a priori. Next, we investigated whether there was a genetic structure coherent with the geographical groups defined a priori. With this purpose, several analyses were carried out: the test of a phylogeographic pattern, the analysis of the most likely number of clusters with the software Structure (Pritchard et al. 2000), and the clustering of the geographical groups. Firstly, the presence of a phylogeographic pattern was tested according to Hardy (2003), which recognised a phylogeographic pattern when gene copies sampled at nearby locations carry alleles that are more related on average than gene copies sampled further apart. In our case, the phylogeographic pattern was tested using Rst by permuting allele sizes among alleles. The permutation procedures permit to assess the distribution of Rst under the null hypothesis that there is no phylogeographic pattern. In consequence, testing the global Rst indicates whether there is a phylogeographic signal within populations, whereas the slope (*b-lin* or *b-log* values of pairwise Rst) indicates whether there is a phylogeographic signal among populations (Hardy and Vekemans 2002).

Secondly, the most likely number of genetic pools was estimated with the programme Structure ver. 2.3.3 (Pritchard et al. 2000), which estimates the most likely number of clusters (K) by calculating the log probability of data (LnP(D)) for each value of K (L(K)), which is an estimate of the posterior probability of the data for a given K value. In our case, posterior probability of the data for each K was obtained for K = 1 through K = 8 clusters using the Admixture model. Ten runs were completed for each run, with 100,000 iterations, after a burn-in period of 10,000. Although the most likely number of different groups can be inferred from the highest K value, Evanno et al. (2005) demonstrated that the real number of groups is best detected by the modal value of ΔK , estimated as the mean of the absolute values of $L''(K)$ averaged over 20 runs divided by the standard deviation of $L(K)$. Therefore, we calculated ΔK and represented their values in the Y-axis and the corresponding number of groups (K) in the X-axis; the modal value of ΔK corresponded to the K value that produced a clear peak.

Thirdly, the geographical groups defined a priori were clustered based on Nei's (1978) standard genetic distance between them, using the software SPAGeDI, and the dendrogram was obtained with the NTSYS-pc software using the UPGMA method (Rohlf 2008).

Ethnobotanical characterisation

A total of 63 on-farm, face-to-face surveys were carried out, representing together the same locations where the hawthorn accessions of the germplasm bank were collected. Each person surveyed fulfilled the conditions of knowing hawthorn (*Crataegus* sp.) and either growing it as a crop or as a backyard tree or plot boundary tree, or, in some cases, just harvesting fruits from semi-natural trees. A semi-structured questionnaire was applied in order to obtain information about the ethnobotanical knowledge about hawthorn and about agronomical practices with this species. From an initial number of variables, 24 showed variation within the sample and were considered for the analysis (Online Resource 4). All the variables were transformed into categorical variables. The data were subjected to multiple correspondence analysis (MCA), which quantifies nominal (categorical) data by assigning numerical values to the cases and categories so that objects within the same category are close together and objects in different categories are far apart. The software IBM SPSS Statistics was used for the analysis (IBM Corp. Released 2016).

Results

Morphological characterisation

Variability of the morphological traits of hawthorn in the whole set of accessions

The coefficient of variation of each characteristic was calculated in order to know the usefulness of each descriptor to detect variability among the whole set of 140 accessions (Online Resource 2). The characteristics that showed less variability between accessions (variation coefficient under 9%) included the number of petals (Npet) and several different characteristics related to the shape of the petals and the shape of the maximum projected area of the endocarp (ICPF, CE, IAPF, EE, IRPF). On the contrary, the highest variation coefficient reached 85% for the percentage of endocarps with abortive seeds (PESA20F) and exceeded 60% for other characteristics related to fruits, like weight (PF), the ratio between the endocarp and seed weight (RSPSF), the percentage of endocarps without seeds (PESS20F), the stamen's maximum

projected area (AEF), and the area of the leaves in reproductive shoots (AFBVR).

The dendrogram obtained for the 140 accessions and the 72 morphological traits revealed four main branches at a dissimilarity level of 5 (Online Resource 3). The internal variability within each branch was higher for the fourth one. The clustering pattern did not separate accessions by putative species or by location, and thus it was only used to select representatives of the most divergent morphological groups for the microsatellite marker analysis.

The PCA (Table 1) showed that the first three components accounted for 52.2, 12.25, and 8.13% of the variance, respectively, with their cumulative variance being 64.4% for two and 72.6% for three. On the basis of the Eigenvector values for traits along the first three components (Fig. 2 and Online Resource 5), the first component was strongly and positively associated with the roundness and compactness of the organs where leaves (IRBR, IRBVC, IRBVL, ICBVC), different parts of the flower (IRRF, IRPF, IREF, ICRF, ICPF, ICEF), the equatorial plane of the fruit (RF), and the maximum projected area of the endocarp (RE, CE) were measured. The ratio length of the main axis/length of the secondary axis of the petals (IAPF) and of the endocarp (EE) were also strongly and positively correlated with the first component. Finally, the same happened for the ratio length of the petiole/length of the main axis of the leaves in long vegetative shoots (LongPBVL), for the weight of the endocarps (PS), and for the maximum projected area of the stamens (AEF, LEMenEF). In all of those cases, the values of the Eigenvectors in the first PC were over 0.95 (Fig. 2). The second component was strongly and positively associated with the rest of the endocarp characteristics, which means those not associated with the first PC (AE, PE, LEMaE, LEMeE, DFE), with values over 0.50, and strongly although negatively (values in parentheses) with other parameters of the leaves related to their size, specifically the length of the secondary axis of the leaves in the reproductive shoots LEMenBR (− 0.507) and the feret diameter of the leaves in vegetative shoots DFBVC (−0.515), DFBVL (−0.504). Figure 3 shows the dispersion of the accessions across the first two PCs, labelled with their origin according to the germplasm groups made a priori.

Table 1 Eigenvalues and cumulative variance for the first 27 components out of 72 in the principal components analysis (PCA) of the morphological descriptors evaluated in the accessions of hawthorn from Mexico

Component	Eigenvalue	Percent	Cumulative
1	37.579	52.19	52.2
2	8.822	12.25	64.4
3	5.856	8.13	72.6
4	3.690	5.13	77.7
5	3.198	4.44	82.1
6	2.099	2.91	85.1
7	1.605	2.23	87.3
8	1.490	2.07	89.4
9	1.060	1.47	90.8
10	0.868	1.21	92.0
11	0.824	1.14	93.2
12	0.790	1.10	94.3
13	0.586	0.81	95.1
14	0.495	0.69	95.8
15	0.491	0.68	96.5
16	0.416	0.58	97.0
17	0.396	0.55	97.6
18	0.320	0.44	98.0
19	0.287	0.40	98.4
20	0.242	0.34	98.8
21	0.209	0.29	99.1
22	0.174	0.24	99.3
23	0.128	0.18	99.5
24	0.123	0.17	99.7
25	0.111	0.15	99.8
26	0.093	0.13	99.9
27	0.077	0.11	100.0

Morphological variability between the germplasm groups defined a priori

In the scatter diagram of accessions in the first two PCs (Fig. 3), the accessions from the different origins roughly tended to be mostly grouped in certain areas of the space formed by the first two principal components, but with the exception of the Chiapas accessions (CHIS), the accessions of the remaining sources do not form well-defined groups in the two-dimensional diagram. In order to determine the variability between the geographic groups established a priori, a CDA was carried out with the geographic group as the

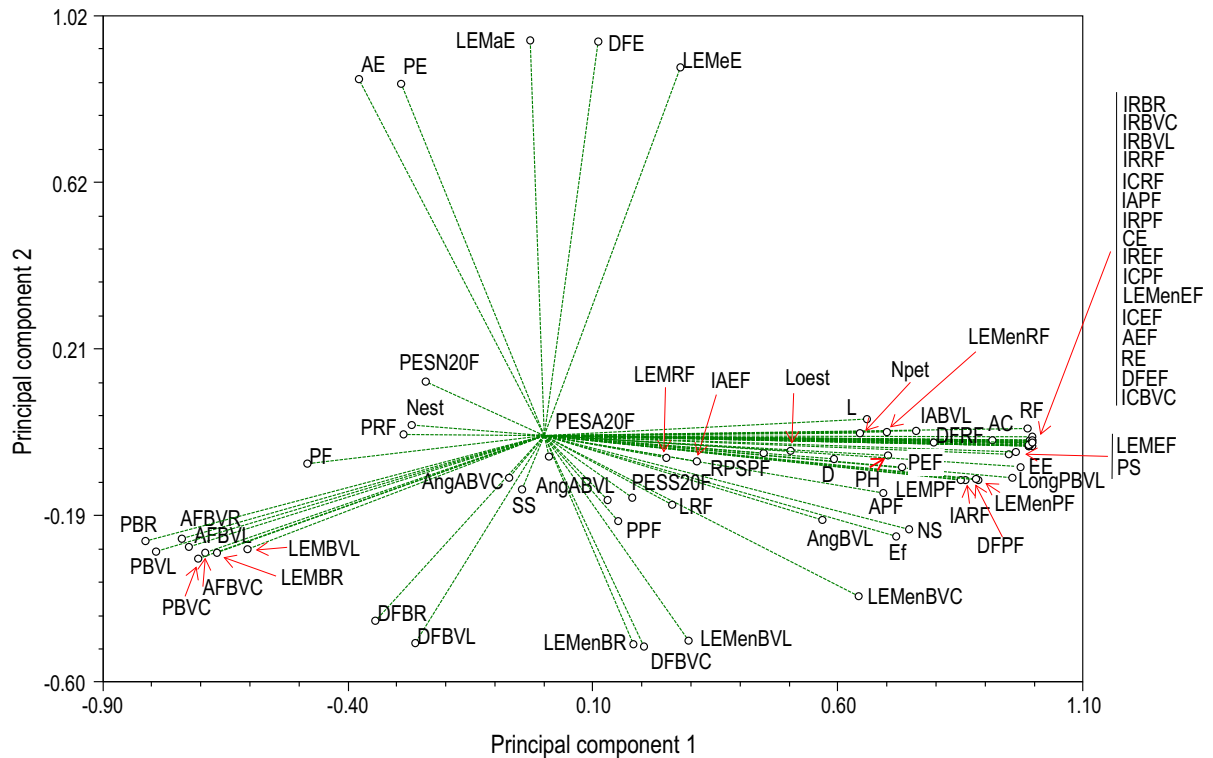


Fig. 2 Plot obtained after the projection on the first two principal components (PC) of the 72 morphological characteristics analysed in the 140 accessions of Mexican hawthorns

classification variable. The variability was partitioned into five canonical roots (CRs) (Table 2). The first CR accumulated 62.1% of the variance and was highly significant. From the morphological traits included in the analysis, the following were more strongly associated with it (with values obtained from the canonical structure in parenthesis): apex and basal angle of the leaves in long vegetative shoots, AngABVL and AngBVL (− 0.392 and − 0.315, respectively); ratio endocarps weight/fruit weight, RPSPF (− 0.325); and percentage of endocarps with normal seeds, PESN20F (0.440) (Online Resource 6). The second CR accumulated 19.6% of the variance and was also highly significant, and it was associated with the roundness of the maximum projected area of the stamens, IREF (− 0.330), and of the leaves in the reproductive shoots, IRBVC (0.490) (Online Resource 6). The third and fourth CRs accounted for a very low variability (10.5 and 5.8%, respectively), although they were also significant (Table 2).

The grouping obtained after the projection of the accession and their centroids in the first two CRs (Fig. 4) showed a separate group with the accessions from CHIS and the rest of the geographical groups not clearly defined. However, the F values associated to the Mahalanobis distances between the germplasm groups (Table 3) showed that all the groups were significantly different ($P < 0.001$) between them, although the lowest F values were for the distance between MICH-1 and MICH-2, and the highest F values were for the bilateral distances between CHIS, on the one hand, and MEX-PUE-TLAX and MEX-MOR, on the other.

The resubstitution test carried out to know the consistency of the grouping obtained (Table 4) indicated that 100% of the accessions from Tete-MOR and MICH-1 were clustered in the same group after the resubstitution test; less than 5% of the accessions from CHIS, 8% of the accessions from MEX PUE TLAX, and less than 15% of the accessions from MEX MOR

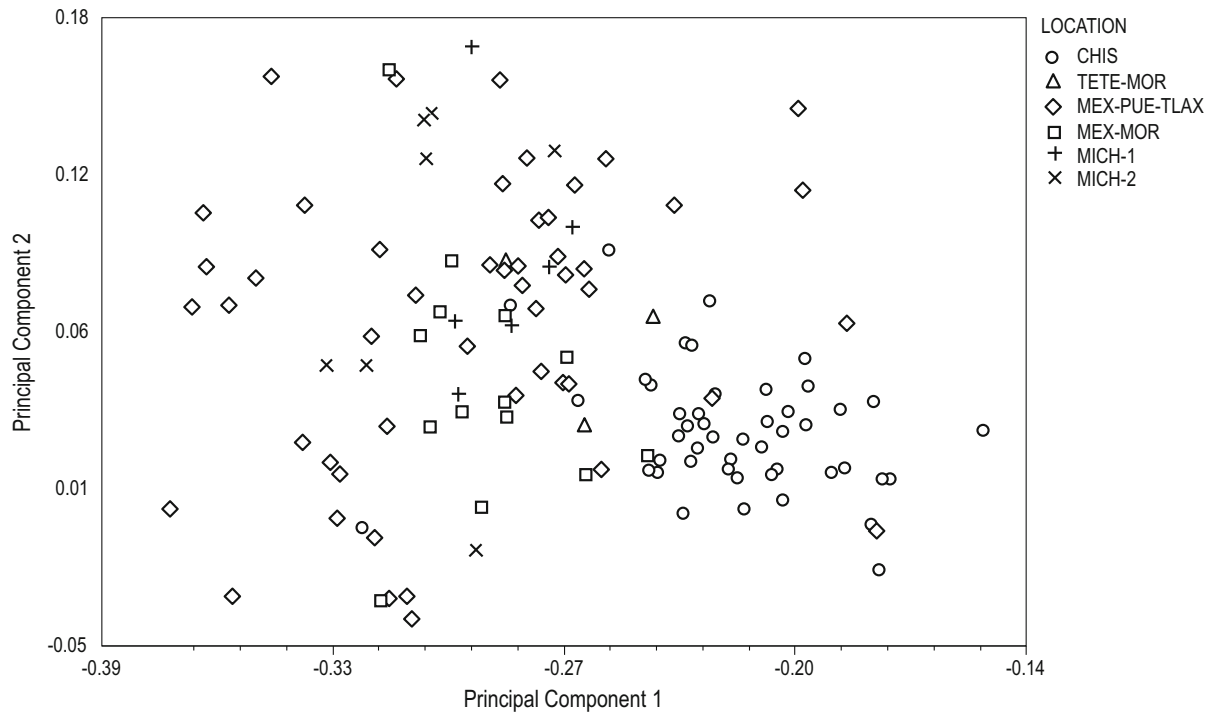


Fig. 3 Plot obtained after the projection in the first two principal components of the accessions of Mexican hawthorns in the PC analysis based on 72 morphological characteristics

Table 2 Eigenvalues from canonical discriminant analysis (CDA) of the morphological descriptors evaluated in the accessions of hawthorn from Mexico

Canonical root	Eigenvalue	Percentage of variance	Cumulative variance (%)	Canonic correlation	$P > F$
1	5.014	62.1	62.1	0.913	0.000
2	1.582	19.6	81.6	0.783	0.000
3	0.865	10.7	92.3	0.681	0.000
4	0.465	5.8	98.1	0.564	0.000
5	0.153	1.9	100.0	0.365	0.085

were clustered in a different group. Conversely, the accessions from MICH-2 were distributed after the resubstitution test in three different groups, and only 57% of them were clustered in the same group (Table 4).

Molecular characterisation

Genetic diversity indexes

The seven microsatellite loci analysed were polymorphic, and each of the 56 accessions studied produced a

unique genotype considering the combination of the seven microsatellite loci (Online Resource 7). The six loci amplified a total of 82 alleles. The number of alleles per locus ranged from 7 (CH04F06) to 18 (CH05G07), with an average number of 12 (Table 5). The mean of the effective number of alleles was 4.6, ranging from 2.5 (CH03D08) to 8.4 (CH05G07) (Table 5). Fifty out of the 82 alleles were rare alleles (with a frequency lower than 0.05), which is consistent with the difference between the mean effective number of alleles per locus (4.6) and the mean number of alleles per locus (12). Considering the seven loci

Fig. 4 Graph obtained after the projection in the two first canonical root of the accessions of Mexican hawthorns in the canonical discrimination analysis based on 72 morphological characteristics as independent variables and the geographical region of origin as the dependent variable

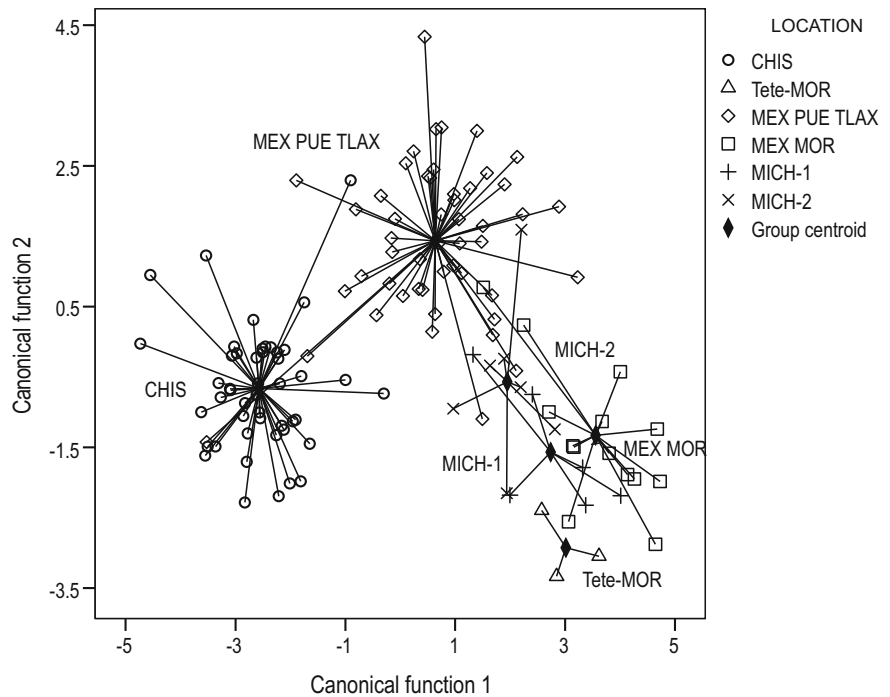


Table 3 F values (above) and corresponding P values (below) for the Mahalanobis distances among the group centroids (Fig. 4) of the different regional groups of accessions of hawthorn from Mexico

	CHIS	Tete-MOR	MEX PUE TLAX	MEX MOR	MICH-1	MICH-2
CHIS		8.972	22.793	27.902	12.539	10.959
Tete-MOR	< 0.0001		6.657	3.770	3.846	3.154
MEX PUE TLAX	< 0.0001	< 0.0001		13.429	7.290	5.175
MEX MOR	< 0.0001	< 0.0001	< 0.0001		4.617	6.283
MICH-1	< 0.0001	< 0.0001	< 0.0001	< 0.0001		2.165
MICH-2	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.014	

Table 4 Resubstitution test to a posterior probability of membership in each location from CDA of the morphological descriptors measured in the accessions of hawthorn from Mexico

	CHIS	Tete-MOR	MEX PUE TLAX	MEX MOR	MICH-1	MICH-2	Total
CHIS	45 (95.7)	0 (0.0)	2 (4.3)	0 (0.0)	0 (0.0)	0 (0.0)	47 (100)
Tete-MOR	0 (0.0)	3 (100)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (100)
MEX PUE TLAX	2 (4.0)	0 (0.0)	46 (92)	1 (2.0)	0 (0.0)	1 (2.0)	50 (100)
MEX MOR	0 (0.0)	0 (0.0)	1 (7.1)	12 (85.7)	1 (7.1)	0 (0.0)	14 (100)
MICH-1	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	6 (100)	0 (0.0)	6 (100)
MICH-2	0 (0.0)	0 (0.0)	1 (14.3)	0 (0.0)	2 (28.6)	4 (57.1)	7 (100)

Parentheses indicate percentages of the total number of accessions

Table 5 Measures of genetic diversity for the seven STMS loci analysed in the whole set of accessions of hawthorn in Mexico: number of observed genotypes (Go), number of alleles per locus (A), effective number of alleles (Ae), observed (Ho) and expected (He) heterozygosity, polymorphism information

content (PIC), probability of coincidence (C), discrimination power (D), and inbreeding coefficient (Fis). The table includes the values of Ho and He for each of the five geographical groups defined a priori

Locus	Go	A	Ae	Ho	He	PIC	C	D	Fis	Population	Ho	He
CH05D04	13	10	4.3	0.846	0.769	0.730	0.176	0.824	– 0.058	CHIS	0.645	0.747
CH03D08	16	10	2.5	0.358	0.600	0.564	0.242	0.758	0.406***	MEX-PUE-TLAX	0.521	0.705
CH04G04	16	10	3.8	0.503	0.739	0.694	0.120	0.880	0.377***	MEX-MOR	0.556	0.550
CH03A02	32	17	7.9	0.538	0.873	0.858	0.044	0.956	0.362***	MICH-2	0.619	0.576
CH05G07	30	18	8.4	0.830	0.881	0.862	0.049	0.951	0.072*	MICH-1	0.460	0.704
CH01F02	19	12	3.4	0.685	0.702	0.660	0.211	0.789	0.042			
CH04F06	11	7	2.2	0.255	0.540	0.495	0.311	0.689	0.505***			
Mean	–	12	4.6	0.573	0.729	0.695	–	–	0.228***	Mean	0.560	0.656
Cumulative	–	–	–	–	–	–	7.2×10^{-7}	> 0.999	–	–	–	–

Significance level: *0.01 $\leq P < 0.05$; ** 0.001 $\leq P < 0.01$; *** $P < 0.001$

and the whole set of accessions, the observed heterozygosity ranged from 0.255 (CH04F06) to 0.846 (CH05D04), with a mean of 0.573 for all loci, whereas the expected heterozygosity ranged from 0.540 (CH04F06) to 0.881 (CH05G07), with a mean value of 0.729 (Table 5).

The polymorphism information content ranged from 0.495 (CH04F06) to 0.858 (CH03A02), with a mean of 0.695, and the discrimination power ranged from 0.689 to 0.956 in the same loci. The cumulative discrimination power exceeded 0.999. A broad variation was observed depending on the locus for the inbreeding coefficient values (Fis), with five out of the seven being significantly different from 0, with a mean value of 0.228.

Considering the five geographical groups represented in the molecular analysis, the observed heterozygosity of the five geographical groups considered a priori ranged from 0.460 (MICH-1) to 0.645 (CHIS), whereas the expected heterozygosity ranged from 0.550 (MEX-MOR) to 0.747 (CHIS) (Table 5). The mean values of the expected heterozygosity in the five geographical groups was 0.656 (Table 5), and the total expected heterozygosity of the whole group of accessions was 0.729 (Table 5). Considering the expected heterozygosity as an estimator of biodiversity, 90.0% of the total biodiversity was within the geographical groups and 10.0% was between them.

Genetic structure

The analysis of the distribution of the biodiversity based on the expected heterozygosity explained above revealed that most of the variability was within geographical groups, and only 10% was between them. In order to explore more deeply the genetic structure, three more analyses were performed. In the first place, the presence of a phylogeographic pattern was analysed according to Hardy (2003) (Table 6). The P values of the global R_{st} were not significant for the whole set of loci analysed nor for each single locus, except for CH05D04 (P value < 0.01). This indicates that there is no phylogeographic signal within populations, as it is not possible to discard the null hypothesis that distinct alleles are equally related within populations as among them. Moreover, the P values of the slope (b-lin and b-log values) of pairwise R_{st} were not significant for any of the loci nor for the whole set of loci, indicating there is no phylogeographic signal among populations, as it is not possible to discard the null hypothesis that distinct alleles are equally related between nearby populations as between distant populations.

The second analysis consisted of the test of the most likely number of genetic pools based on the modal value of ΔK (Fig. 5a), which indicated that the most likely number is two. Moreover, Bayesian clustering analysis with structure software (Fig. 5b) confirmed the presence of two groups. The first consisted mainly of the accessions from CHIS, although some from

Table 6 *P* values corresponding to the test of the phylogeographic pattern (under the null hypothesis that there is no phylogeographic pattern) based on *R*_{st} by permuting allele sizes (10,000 random permutations) among alleles. Initial seed = 1125439161

Locus	P-values for <i>R</i> _{st}	b-lin (slope linear dist)	b-log (slope ln(dist))
CH05D04	0.0011**	0.4149	0.4918
CH03D08	0.8716	0.5104	0.6201
CH04G04	0.8502	0.5612	0.4697
CH03A02	0.2851	0.8707	0.7995
CH05G07	0.0275	0.1767	0.1113
CH01F02	0.1845	0.5652	0.4283
CH04F06	0.2453	0.4988	0.5106
ALL LOCI	0.3885	0.9165	0.8588

Significance level: *0.01 ≤ *P* < 0.05; ** 0.001 ≤ *P* < 0.01; *** *P* < 0.001

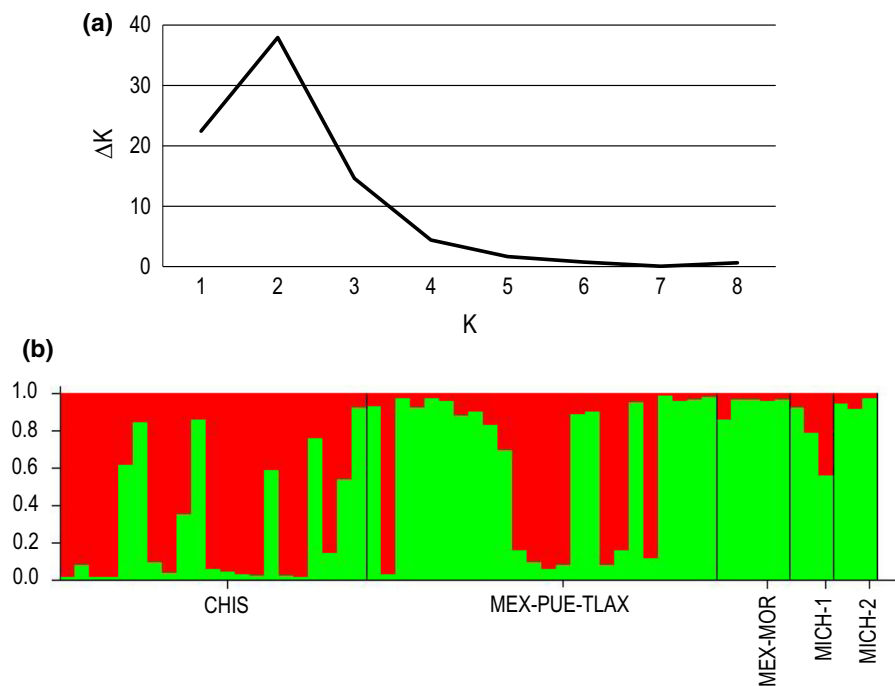


Fig. 5 **a** Graphic representation of the optimum number of groups (*K* value) after the analysis of microsatellites from Mexican hawthorns with the programme structure. ΔK was

calculated as $\Delta K = m(L''(K))/s[L(K)]$ (Evanno et al. 2005); **b** genetic divergence among the Mexican hawthorn populations inferred from the Bayesian clustering method

other regions were also included in this pool. The second pool consisted mainly of the accessions from regions other than CHIS, although some from CHIS were also clustered in this second pool, mainly in the MEX-PUE-TLAX group.

The third analysis consisted of the clustering based on Nei's genetic distance of the five geographical groups (Fig. 6). A second dendrogram was obtained including the five hawthorn groups plus the outliers

(Fig. 6). The five geographical groups were grouped in the dendrogram's two main pools, although the first consisted of the CHIS and MEX-PUE-TLAX groups mixed together, and the second of the rest of the accessions.

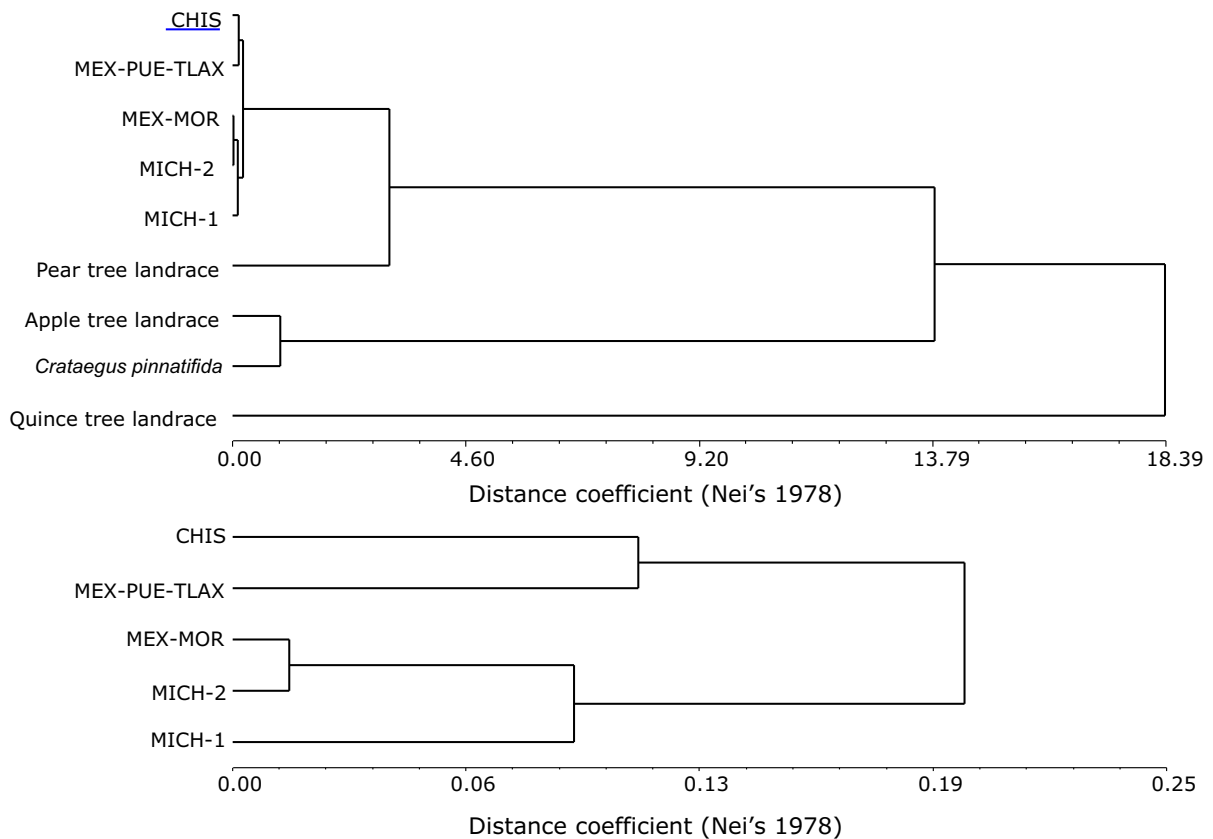


Fig. 6 Genetic divergence among the Mexican hawthorn populations inferred from UPGMA cluster analysis based on Nei's (1978) genetic distance obtained with the allelic analysis

of the seven microsatellite loci analysed. The upper figure includes the controls and outliers besides the hawthorn populations, and the lower figure only the hawthorn populations

Ethnobotanical study

The global results of the survey, with the frequencies of each state of the variable, in the six geographical regions together are shown in the supplementary material (Online Resource 4). Regarding the profile of the sample, the age of the surveyed individuals ranged from young to elderly, with most (57.4%) from 36 to 55 years old. Most of the families (67.7%) had five or fewer members, and very few of them more than 10. Most of the surveyed persons were farmers (46.0%), and the rest were mainly small retailers. In terms of cultural level, the surveyed individuals completed primary or secondary studies, with an irrelevant percentage of illiterate individuals and university graduates. Regarding the knowledge of hawthorn, most knew about this species their whole life, and only 6.4% discovered this species in the last few years. This knowledge is acquired mostly from parents, although

20.6% recognised that they do not transfer their knowledge about this species to anybody. Regarding the relationships with this species, only 28.6% of the hawthorn growers were professional farmers of this crop, and the rest of them cultivated fewer than eight trees just as a hobby. Only seven out of the 63 growers did not use the hawthorn in any way, although they kept and preserved the trees. They belonged to CHIS (4), TETE-MOR (2), and MEX-MOR (1). Of the others, most used the fruits as fresh food or to prepare a broad range of traditional cakes. Half of the surveyed growers sell at least part of the fruits harvested. Regarding agronomic practices, the most common were pruning (32.3% of the growers) and fertilisation (27.4%), which indicates limited cultural practices, even for the professional hawthorn farmers.

MCA was carried out to obtain a grouping of the cases (farmers surveyed) based on the ethnobotanical features, with the objective of analysing whether the

obtained ethnobotanical grouping is consistent with the geographical groups considered a priori. The first two dimensions obtained after the MCA accounted for 31.9% and 11.6% of the total variance, respectively, with a cumulative variance of 43.5% (data not shown). This level of variance is usually accepted to approve the model. Figure 7 and Online Resource 8 show the discrimination measures for the ethnobotanical features analysed in the two first dimensions obtained after the MCA. The variables with a higher influence in dimension 1 were related to the professionalism in the exploitation of hawthorn, specifically the following (with discrimination measures in parentheses): hawthorn cultivation as a professional activity (0.867), industrial use (0.852), fertilisation practice (0.833), pruning practice (0.704), and number of trees (0.744).

Regarding dimension 2, the features with higher influence (with discriminant measures in parentheses) were the occupation of the hawthorn grower (0.479) and the age of the personal hawthorn knowledge (0.356). Figure 8 shows the projection of all the cases in the space defined by the first two dimensions, labelling them by the geographic group of origin. Dimension 1 distinguishes between the cases from MEX-PUE-TLAX, located in the left side of the graphic, and the rest of the cases. As the most relevant variables for dimension 1 are those related to the professionalism of the exploitation of hawthorn, the main difference between the MEX-PUE-TLAX

geographical group and the others corresponds to this feature, because professional growers are concentrated in the MEX-PUE-TLAX region. Conversely, dimension 2 does not distinguish between any of the geographical groups, and all cases are distributed along the component, indicating that none of the corresponding features are linked to any of the geographical groups.

Discussion

The *tejocote* (hawthorn) tree is a well-known plant species for food and other uses in Mexican culture since pre-Hispanic times, although it has usually been considered an underutilised crop (Nieto-Ángel 2007; Lozano-Grande et al. 2016). However, worldwide there is a movement for the revalorisation of traditional crops (Bessi ere 2013), and Mexico is part of this movement (Castellanos and Bergstresser 2016). Mexican hawthorns are a good example of revalorisation of a traditionally underutilised crop, and different actions for the promotion of this crop (Nieto- ngel 2013) have resulted in the establishment of commercial orchards in the centre of the country, with a likely increase (SAGARPA 2013). The fruits from such plantations are sold in the local market for food or for the emerging pectin-extraction industry; they are even exported to the USA for the nostalgia market of the

Fig. 7 Plot obtained after the projection of the ethnobotanical variables analysed in the first two dimensions of the multiple correspondence analysis (MCA) carried out with the ethnobotanical data obtained in the 63 face-to-face surveys in six geographical areas of Mexico where hawthorn accessions were collected

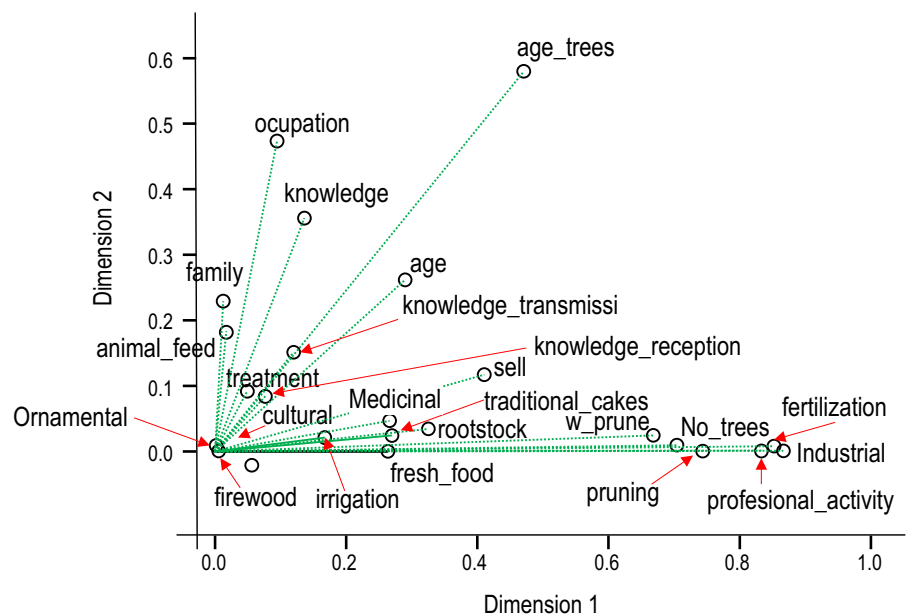
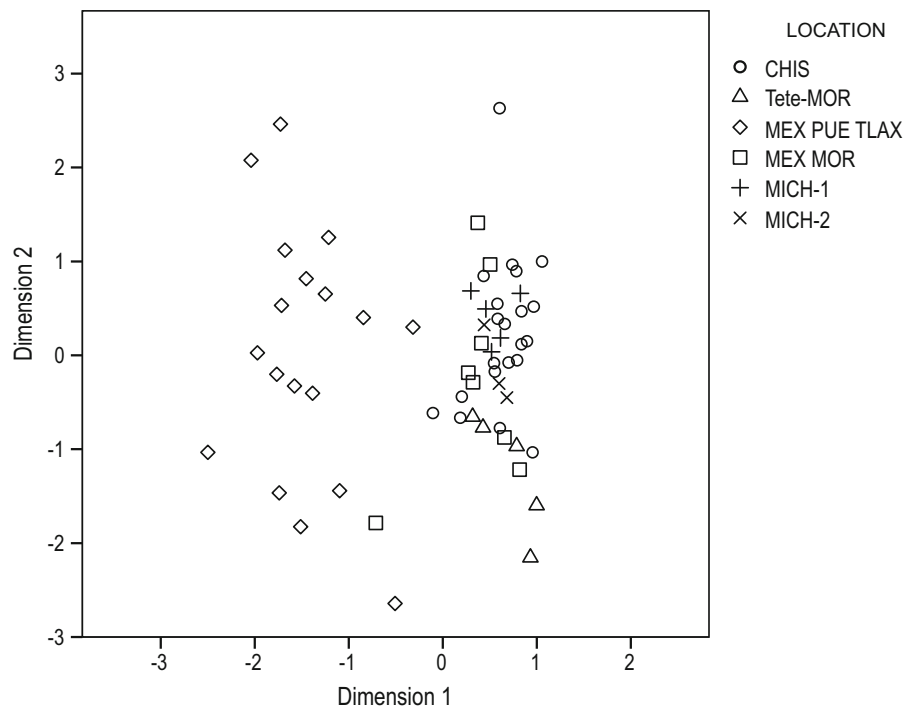


Fig. 8 Plot obtained after the projection of the cases (farmers surveyed in the six geographical areas of Mexico where hawthorns were collected) in the first two dimensions of the MCA carried out with the ethnobotanical data obtained in the 63 face-to-face surveys



Mexican migrant population (Mena and García 2014). As a consequence, commercial varieties based on the improvement of local genotypes are being registered and released (SNICS 2012; Velasco-Hernández et al. 2017).

Despite the apparent benefit, this kind of evolution from an underutilised crop to a commercial crop carries the intrinsic threat of decreasing the biodiversity of the species (Hammer 2004; Stamp et al. 2012). Besides the risks of hybridisation, in such processes the tolerated and favoured wild plants (Mapes and Basurto 2016), as well as the cultivated landraces, are especially vulnerable because they are at serious risk of being abandoned or replaced by improved varieties with nicer, bigger, and tastier fruits. This neglect of genotypes not only affects professional hawthorn farmers but could also affect familiar backyard trees, trees inside crops, or trees used as boundaries in cultivated lands; thus, the conservation of this germplasm must become a priority (Ford-Lloyd et al. 2011).

This is, to our knowledge, the most comprehensive study of the biodiversity of Mexican hawthorns, including plants with anthropic management (cultivated landraces, favoured plants, and tolerated plants), in a geographical region spanning from the south

(Chiapas) to central Mexico, stretching from Michoacan in the west to Puebla in the east. The results allow the development of a strategy for future Mexican hawthorn conservation programmes.

Germplasm biodiversity distribution

One important issue with Mexican hawthorns when attempting to assess the existing biodiversity is that under the name ‘tejocote’ there is a great range of putative species. In the BGT-UACH, the collected species belong to the following series (Fig. 9): *Baroussanae* (*C. baroussana* Eggl., 1909; *C. cuprina* J.B. Phipps, 1997.), *Crus-galli* (*C. gracilior* J.B. Phipps, 1997; *C. rosei* Eggl., 1909), *Greggianae* (*C. greggiana* Eggl., 1909; *C. sulfurea* J.B. Phipps, 1997), *Madrenses* (*C. aurescens* J.B. Phipps, 1997; *C. tracyi* Ashe ex Eggl., 1909), and *Mexicanae* (*C. mexicana* Moc. et Sessé ex DC., 1825; *C. nelsoni* Eggl., 1909; *C. stipulosa* [Kunth] Steud., 1840). Most of the biodiversity studies using morphological and molecular characteristics carried out with *Crateagus* have focused on the study of the taxonomy of a species or a group of related species, such as *C. douglasii* Lindley and *C. suksdorfii* (Sargent) Kruschke (Lo et al. 2009) and the *C. rosei* complex (Piedra-Malagón et al. 2016)

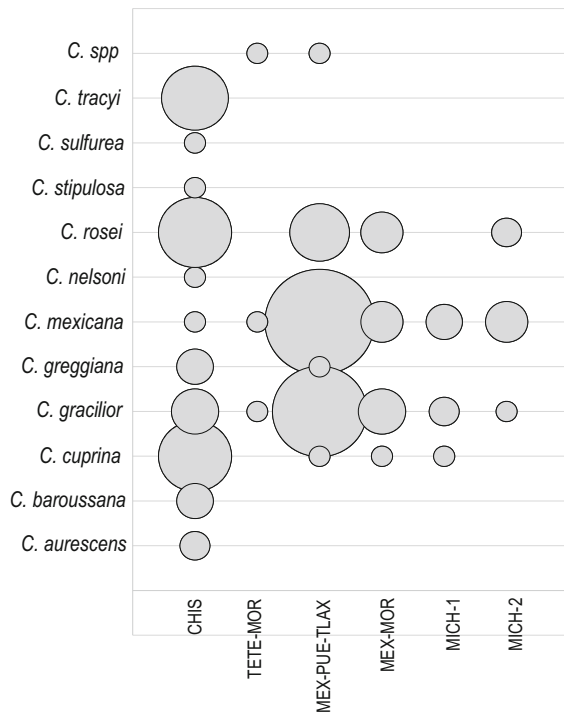


Fig. 9 Relative abundance, expressed by the dot's size, of the different putative species in the six geographical areas where Mexican hawthorns were collected

in America, or other studies with species from Europe and Western Asia (Khiari et al. 2015; Rahmani et al. 2015). Conversely, in our approach, we aimed to study the diversity of the Mexican hawthorns with anthropic use as a whole in different geographical regions regardless of the taxonomy, which is problematic for the American *Crataegus* (Piedra-Malagón et al. 2016).

As already observed by Piedra-Malagón et al. (2016) with the *C. rosei* complex, in our work the individuals from the different putative species were not grouped into monophyletic groups after cluster analysis. In the case of both morphological (Online Resource 3) and molecular (data not shown) characterisation, some accessions were immersed in groups that include accessions from other species. A possible explanation for this is the hybridisation and introgression events (Seehausen 2004) that have been reported for Mexican *Crataegus* species (Piedra-Malagón et al. 2016). The diffuse boundary among the *Crataegus* species supports the advantage of analysing the biodiversity of hawthorns by regional pools, regardless of the species to which the accessions are

attributed, when the purpose is the assessment of the global biodiversity of the crop, as in the present study.

According to the obtained values of expected heterozygosity, most of the genetic diversity is located within the geographical regions (90%), and therefore only 10% is between them. For one *Crataegus* species or a small group of related species, previous studies have demonstrated that most of the biodiversity was also located within geographic sites, i.e. the *C. rosei* complex (Pedra-Malagón et al. 2016) and *C. pontica* C. Koch (Rahmani et al. 2015). Interestingly, in our case the biodiversity inside a geographical region was not related to the number of putative species present in the region. Certainly the expected heterozygosity (H_e) reached the highest value (0.747) in Chiapas (CHIS), where the number of species was also the highest, as 11 species were present in that region, and five were exclusive of it (*C. tracyi*, *C. sulfurea*, *C. stipulosa*, *C. baroussana*, and *C. aurescens*). However, the expected heterozygosity of the area corresponding to MEX-PUE-TLAX in central Mexico, was very similar (0.705) to that of Chiapas, with only six species and H_e of Eastern Michoacan (MICH-1) was 0.704, with only three species. Moreover, since Chiapas is richer in putative species, Núñez-Colín et al. (2008) found higher biodiversity in a pool from Puebla than in a pool from Chiapas, although the biodiversity was measured only with morphological features, and the number of accessions per geographical site was smaller. This could be due to the anthropic genotypes flow with destination in central Mexico, which according to Núñez-Colín et al. (2008) and Nieto-Ángel et al. (2009), has been due to the immigration in this area. The exchange of genetic material can lead to an increase of biodiversity and, eventually, to a possible development of new taxa, as hybridisation is very common in *Crataegus* and has led to a rapid radiation in this genus (Lo and Donoghue 2012; Zheng et al. 2014). The high biodiversity in central Mexico, with a lower number of putative species than in Chiapas, could be due to the indicated genotypes flow, which could have led to an increase of the biodiversity by hybridization, although it has not still led to the development of new taxa. This could validate and explain the starting hypothesis, by which the anthropic influence may increase the intrapopulational biodiversity.

Possible presence of different genetic pools

The genetic variation of the Mexican hawthorns analysed was weakly geographically structured. The phylogeographical test of microsatellite data (Hardy 2003) revealed the absence of a phylogeographical pattern. The Mexican hawthorns from the BGT-UACH showed different ploidy levels with a mix of sexual, apomictic, and vegetative reproduction, but they were predominantly polyploid, with apomictic and vegetative reproduction (unpublished data from R. Nieto-Ángel, BGT-UACH), as most of the American *Crataegus* species (Talent and Dickinson 2007; Dickinson et al. 2008). Lo et al. (2009, 2013) corroborated with two American species of *Crataegus*, one polyploid apomictic and the other diploid sexual, that in apomictic parthenogenetic plants the variation is not geographically structured, but distributed in a broader geographical scale, which could explain the lack of a geographical pattern of biodiversity observed. Moreover, Lo et al. (2009) demonstrated that the polyploid apomixes is compatible with the maintenance of high genetic variability, as well as with the spread of genotypes across wide geographical distances, which could explain the high levels of biodiversity found in the Mexican hawthorns analysed.

However, despite the lack of a geographically structured pattern of biodiversity, there are signs of differentiation of the pool from Chiapas from both the morphological and molecular viewpoints. In the PCA of the morphological data, the accessions from Chiapas tended to be grouped in the lower right side of the space defined by the first two PCs, whereas the accessions from the rest of the geographical origins were more spread in the plot. Taking into account the contribution of the different characteristics to the corresponding PC, the accessions from Chiapas tended to have a more rounded shape in different organs like leaves, some flower parts, fruits, and the endocarp projected area, as well as a smaller endocarp. Moreover, the CDA of the morphological data showed that the most significant distance among the centroids of the geographical groups corresponded to the distance between the accessions from Chiapas and other geographical origins. This result agrees with the conclusions achieved by Núñez-Colín et al. (2008) and Nieto-Ángel et al. (2009) with morphological data for a smaller sample of accessions from the BGT-UACH from a smaller number of geographical regions.

However, as the CDA selects the morphological features that best discriminate between the different states of the independent variable (the geographical locations, in this case), it forces the discrimination, and therefore the distances were highly significant for all possible pairs, although less significant ($P < 0.05$) for the two pools from Michoacan. The CDA also indicated that the morphological features that best discriminate among the different geographical groups were the roundness, apex, and basal angles of the leaves, and the relative size of the endocarp compared with the full fruit.

The microsatellite characterisation mostly agreed with the result of the morphological analysis: The most likely number of clusters (K), according to Evanno et al. (2005), was two, and the Bayesian cluster defined two main groups; in both cases, the first group consisted mainly of the accessions from Chiapas, but with some accessions from the other groups, whereas the second group consisted mainly of accessions from the rest of the groups, but with some accessions from Chiapas. This could be due to incursions of materials from central Mexico in Chiapas and vice versa, as already reported, based on morphological data, by Núñez-Colín et al. (2008) and Nieto-Ángel et al. (2009), who demonstrated a genotype flow from Mexico and Puebla to Chiapas and vice versa. According to these authors, the accessions from Chiapas with the largest and tastier fruits, belonging to *C. mexicana* and *C. gracilior*, were probably brought from other parts of Mexico; conversely, some plants from Chiapas were carried to central Mexico by Chiapas immigrants. However, the molecular characterisation showed more incursions of the germplasm from one pool to the other than the morphological characterisation, although this kind of partial discordance between the results obtained with morphological and molecular markers is common and could be due to the homoplasy of morphological characteristics (Piedra-Malagón et al. 2016).

Ethnobotany of Mexican hawthorns
and the conservation of hawthorn resources

Ethnobotanical analysis is not common in studies of biodiversity, yet it provides information of key importance for the conservation of the genetic resources of a given taxa (Lira et al. 2016). The MCA of the ethnobotanical data showed differences

between the hawthorn growers from MEX-PUE-TLAX and those from the rest of the geographical groups in the first dimension of the MCA. This was because all the professional growers were in MEX-PUE-TLAX, and therefore some variables showed characteristic values for this region (i.e. the number of trees per grower was higher, and agronomic practices such as pruning and fertilisation were more common, as deduced from the characteristics contributing to the first dimension of the MCA). The use of hawthorns is not affected by the shift to commercial activity, except, of course, industrial use, which was exclusive to MEX-PUE-TLAX. Therefore, MEX-PUE-TLAX is the region with the highest risk of losing the traditional germplasm, being substituted by the germplasm of improved varieties more accepted commercially. In this case, by word of mouth (Ahuja et al. 2016), both professional and amateur hawthorn growers will know about the better organoleptic characteristics of the commercial varieties, and they could change the existing trees by the new ones.

Interestingly, Chiapas, which from the biological viewpoint, shows the highest differences from the other regions, from the ethnobotanical one was grouped together with the other regions, except for MEX-PUE-TLAX, which was alone, as indicated above. However, a more detailed analysis of the ethnobotanical characteristics shows a subtle difference of Chiapas from the rest of the geographical locations, in that even if fresh fruit is consumed, it is consumed only occasionally, because the fruits are not especially appreciated due to a significant presence of hawthorns from the species *C. tracyi*, *C. rosei*, and *C. baroussana* in Chiapas, which have small and acidic fruits.

It is likely that the commercial plantations could extend to other areas of central Mexico and other regions of northern Mexico, which also show a high probability of occurrence of hawthorns with potentially high biodiversity (Núñez-Colín 2009), so the risk of losing biodiversity can be a threat in such areas. Therefore, a politics of germplasm conservation must be a priority in those regions where commercial plantations are being established. Chiapas is also a sensible area, because it can be considered a different gene pool, although taking into account the present social conditions, installation of commercial plantations is unlikely.

Even if *on-farm* conservation is the most adequate strategy (Hammer 2004), *ex situ* conservation is also necessary to ensure successful preservation (Arndorfer et al. 2009), as proposed by Vera-Castillo et al. (2014) with another underutilised Mexican crop that is changing to commercial status (*Jatropha curcas* L.). In the case of hawthorns, *ex situ* conservation is necessary by other means if an allochthonous germplasm is being introduced, because *Crataegus* is prone to produce hybridisation followed by rapid radiation (Lo and Donoghue 2012; Zheng et al. 2014), which can lead to genetic swamping and extinction through genetic replacement of small populations of a species by more numerous populations (Bupp et al. 2017). Due to the high internal biodiversity of the geographical locations, the collection of a germplasm for *ex situ* conservation must be exhaustive in each location. Due to their high PIC values and high discriminant value (D), microsatellites are useful to eliminate possible duplicates in the germplasm bank.

In conclusion, the three sequential steps of analysis carried out (morphological, molecular, and ethnobotanical) have proved to be useful for assessment of the biodiversity of hawthorn and to delineate strategies for germplasm conservation, which can be extended to other regions for the same crop or for other underutilised crops. The preservation strategy must be focused chiefly on those regions where commercial plantations are being established, as well as those with genetically different germplasms, and it will consist in an exhaustive collection of plant resources within each location selected for conservation, because most of the hawthorns' biodiversity was inside regional pools. In order to ensure efficient conservation of genetic resources, any governmental policies to motivate commercial plantations of hawthorn should be accompanied by measures to achieve an efficient collection of material for *ex situ* conservation.

Acknowledgements M. Betancourt-Olvera has been granted by the “Consejo Nacional de Ciencia y Tecnología” from Mexico (Grant Number 103701).

Compliance with ethical standard

Conflict of interest The corresponding author declares on behalf of all the authors: The authors have no conflict of interest.

Human and animal rights The research does not involve human participants and/or animals.

References

- Ahuja LR, Ma L, Howel TA (2016) Whole system integration and modelling-essential to agricultural science and technology in the 21st century. In: Ahuja LR, Ma L, Howel TA (eds) *Agricultural system models in field research and technology transfer*. Lewis Publishers, New York, pp 1–9
- Alercia A, Diulgheroff S, Mackay M (2015) FAO bioersity multi-crop passport descriptors V.2.1. Food and Agriculture Organization of the United Nations (FAO), Rome
- Arndorfer M, Kajtna B, Vorderwülbecke B (2009) Integrating ex situ and on-farm conservation approaches in the management of local vegetable diversity in Austria. In: 1st international symposium on horticulture in Europe. G.R Dixon, Wien, pp 333–340
- Banderas-Tarabay JA, Cervantes-Rodríguez M, Méndez-Iturbide D (2015) Biological properties and antioxidant activity of hawthorn *Crataegus mexicana*. *J Pharmacogenomics Pharmacoproteomics*. <https://doi.org/10.4172/2153-0645.1000153>
- Betancourt-Olvera B, Perez-Lainez MD, Nieto-Angel R, Barrientos-Priego AF, García-Mateos MR, Corona-Torres T (in press) Comparison of six methods of DNA extraction in tejocote (*Crataegus mexicana* Moc.&Sessé). *Rev Fitotec Mex*
- Borys MW, Leszczyńska-Borys H (1994) Tejocote (*Crataegus* spp.): planta para solares, macetas e interiores. *Rev Chapingo Ser Hortic* 1:95–107
- Brown JA, Beatty GE, Finlay CMV, Montgomery WI, Tosh DG, Provan J (2016) Genetic analyses reveal high levels of seed and pollen flow in hawthorn (*Crataegus monogyna* Jacq.), a key component of hedgerows. *Tree Genet Genomes*. <https://doi.org/10.1007/s11295-016-1020-0>
- Bupp G, Ricono A, Peterson CL, Pruett CL (2017) Conservation implications of small population size and habitat fragmentation in an endangered lupine. *Conserv Genet*. <https://doi.org/10.1007/s10592-016-0883-9>
- Cabrera LG (1992) *Diccionario de Aztequismos*. Colofón, México Distrito Federal
- Campbell CS, Evans RC, Morgan, Dickinson TA, Arsenault MP (2007) Phylogeny of subtribe Pyrinae (formerly the Maloideae, Rosaceae): limited resolution of a complex evolutionary history. *Plant Syst Evol*. <https://doi.org/10.1007/s00606-007-0545-y>
- Castellanos E, Bergstresser S (2016) The Mexican and transnational lives of corn: technological, political, edible object. In: Brulotte RL, Di Giovine MA (eds) *Edible identities: food as cultural heritage*. Taylor and Francis, New York, pp 201–219
- Chen CY, Li H, Yuan YN, Dai HQ, Yang B (2013) Antioxidant activity and components of a traditional chinese medicine formula consisting of *Crataegus pinnatifida* and *Salvia miltiorrhiza*. *BMC Complement Altern Med*. <https://doi.org/10.1186/1472-6882-13-99>
- Dickinson TA, Lo EY, Talent N, Love RM (2008) Black-fruited hawthorns of western North America one or more agamic complexes? *J Bot*. <https://doi.org/10.1139/B08-072>
- Doyle JJ, Doyle JL (1990) A rapid total ADN preparation procedure for fresh plant tissue. *Focus*. https://doi.org/10.1007/978-3-642-83962-7_18
- Edwards JE, Brown PN, Talent N, Dickinson TA, Shipley PR (2012) A review of the chemistry of the genus *Crataegus*. *Phytochem*. <https://doi.org/10.1016/j.phytochem.2012.04.006>
- Eggleston WW (1909) The Crataegi of Mexico and Central America. *Bull Torrey Bot Club* 36:501–514
- Evanno G, Gagnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Evans RC, Campbell CS (2002) The origin of the apple sub-family (*Maloideae*; *Rosaceae*) is clarified by DNA sequence data from duplicated GBSSI genes. *Am J Bot*. <https://doi.org/10.3732/ajb.89.9.1478>
- Ferreira V, Ramos-Cabrera AM, Carnide V, Pinto-Carnide O, Assunção A, Marreiros A, Rodrigues R, Pereira-Lorenzo S, Castro I (2016) Genetic pool structure of local apple cultivars from Portugal assessed by microsatellites. *Tree Genet Genomes*. <https://doi.org/10.1007/s11295-016-0997-8>
- Ford-Lloyd BV, Schmidt M, Armstrong SL, Barazani O, Engels J, Hadas R, Hammer K, Kell SP, Kang D, Khoshbakht K, Li Y, Long Ch, Lu BR, Ma K, Tung Nguyen V, Qiu L, Ge S, Wei W, Zhang Z, Maxted N (2011) Crop wild relatives—undervalued, underutilized and under threat? *Bioscience*. <https://doi.org/10.1525/bio.2011.61.7.10>
- Hammer K (2004) Resolving the challenge posed by agrobiodiversity and plant genetic resources—an attempt. *J Agr Rural Dev Trop* 76:1–184
- Hardy OJ (2003) Estimation of pairwise relatedness between individuals and characterisation of isolation by distance processes using dominant genetic markers. *Mol Ecol* 12:1577–1588
- Hardy OJ, Vekemans X (2002) Spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes*. <https://doi.org/10.1046/j.1471-8286.2002.00305.x>
- IBM Corp. Released (2016) IBM SPSS statistics for windows, Version 24.0. IBM Corp, Armonk
- Khiari S, Boussaid M, Messaoud C (2015) Genetic diversity and population structure in natural populations of Tunisian Azarole (*Crataegus azarolus* L. var. *aronia* L.) assessed by microsatellite markers. *Biochem Syst Ecol* 59:264–270
- Liebharr R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van de Weg E, Gessler C (2002) Development and characterisation of 140 new microsatellites in apple (*Malus x domestica* Borkh.). *Mol Breed*. <https://doi.org/10.1023/A:1020525906332>
- Lira R, Casas A, Blancas J (2016) *Ethnobotany of Mexico: interactions of people and plants in Mesoamerica*. Springer, New York
- Lo EY, Donoghue MJ (2012) Expanded phylogenetic and dating analyses of the apples and their relatives (Pyreae, Rosaceae). *Mol Phylogenet Evol*. <https://doi.org/10.1016/j.ympev.2011.10.005>
- Lo EY, Stefanović S, Dickinson TA (2009) Population genetic structure of diploid sexual and polyploid apomictic

- hawthorns (*Crataegus*; *Rosaceae*) in the Pacific Northwest. *Mol Ecol*. <https://doi.org/10.1111/j.1365-294X.2009.04091.x>
- Lo EYY, Stefanovic S, Dickinson TA (2013) Geographical parthenogenesis in Pacific Northwest hawthorns (*Crataegus*; *Rosaceae*). *Botany*. <https://doi.org/10.1139/cjb-2012-0073>
- Lozano-Grande MA, Valle-Guadarrama S, Aguirre-Mandujano E, Lobato-Calleros CSO, Huelitl-Palacios F (2016) Films based on hawthorn (*Crataegus* spp.) fruit pectin and candelilla wax emulsions: characterization and application on *Pleurotus ostreatus*. *Agrociencia* 50(7):849–866
- Mapes C, Basurto F (2016) Biodiversity and edible plants of Mexico. In: Lira R, Casas A, Blancas J (eds) *Ethnobotany of Mexico interactions of people and plants in Mesoamerica*. Springer, New York, pp 83–133
- Nieto-Ángel R (2007) Colección, conservación y caracterización del tejocote (*Crataegus* spp.). In: Nieto-Ángel R (ed) *Frutales Nativos, un Recurso Fitogenético de México*. Universidad Autónoma Chapingo, Estado de México, pp 25–107
- Nieto-Ángel R (2013) Un análisis de propuestas de políticas públicas para el desarrollo rural en México. Universidad Politécnica de Francisco I Madero, Hidalgo
- Nieto-Ángel R, Pérez-Ortega SA, Núñez-Colín CA, Martínez-Solís J, González-Andrés F (2009) Seed and endocarp traits as markers of the biodiversity of regional sources of germplasm of tejocote (*Crataegus* spp.) from Central and Southern Mexico. *Sci Hortic*. <https://doi.org/10.1016/j.scienta.2009.01.034>
- Núñez-Colín CA (2009) Priority areas to collect (*Crataegus* L.) germplasm in Mexico on the basis of diversity and species richness. *Agric Téc Méx* 35:333–338
- Núñez-Colín CA, Hernández-Martínez MA (2011) The problems in the taxonomy of the genetic resources of tejocote (*Crataegus* spp.) in Mexico. *Rev Mex Cienc Agríc* 2:141–153
- Núñez-Colín CA, Nieto-Ángel R, Barrientos-Priego AF, Saha-gún-Castellanos J, Segura S, González-Andrés F (2008) Variability of three regional sources of germplasm of tejocote (*Crataegus* spp.) from central and southern Mexico. *Genet Resour Crop Evol*. <https://doi.org/10.1007/s10722-008-9316-z>
- Phipps JB (1997) Monograph of Northern Mexican *Crataegus* (*Rosaceae* subfam. *Maloideae*). *SIDA Bot Misc* 15:1–94
- Phipps JB, O'Kennon RJ, Lance RW (2003) Hawthorns and medlars. Royal horticultural society plant collector guide. Timber Press, Portland
- Piedra-Malagón EM, Albarrán-Lara AL, Rull J, Piñero D, Sosa V (2016) Using multiple sources of characters to delimit species in the genus *Crataegus* (*Rosaceae*): the case of the *Crataegus rosei*. *Syst Biodivers*. <https://doi.org/10.1080/14772000.2015.1117027>
- Pritchard JK, Matthew S, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Rahmani MS, Shabaniyan N, Khadivi-Khub A, Woeste KE, Badakhshan H, Alikhani L (2015) Population structure and genotypic variation of *Crataegus pontica* inferred by molecular markers. *Gene*. <https://doi.org/10.1016/j.gene.2015.07.001>
- Rohlf FJ (2008) NTSYSpc: numerical taxonomy system, ver.2.20. Exeter Publishing, Ltd, Setauket
- SAGARPA (2015) Autoriza EUA la importación de higos y tejocotes mexicanos. SAGARPA. http://comunicacion.social.guanajuato.blogspot.com.es/2015_04_06_archive.html. Accessed 19 June 2017
- Seehausen O (2004) Hybridization and adaptive radiation. *Trends Ecol Evol*. <https://doi.org/10.1016/j.tree.2004.01.003>
- SNICS (2012) Catálogo 29 nacional de variedades vegetales del Servicio Nacional de Inspección y Certificación de Semillas. SAGARPA, México
- Stamp P, Messmer R, Walter A (2012) Competitive underutilized crops will depend on the state funding of breeding programmes: an opinion on the example of Europe. *Plant Breed*. <https://doi.org/10.1111/j.1439-0523.2012.01990>
- Talent N, Dickinson TA (2007) Ploidy level increase and decrease in seeds from crosses between sexual diploids and asexual triploids and tetraploids in *Crataegus* L. (*Rosaceae*, *Spiraeoideae*, *Pyreae*). *Can J Bot*. <https://doi.org/10.1139/B07-028>
- UPOV (2008) Hawthorn UPOV Code: CRATA. *Crataegus* L. guidelines for the conduct of tests for distinctness, uniformity and stability. International Union for the Protection of new Varieties of plants, Geneva
- Velasco-Hernández A, Saucedo-Veloz C, Nieto-Ángel R, Ramírez-Guzman ME, Saucedo-Reyes D, Nieto-Loez EH (2017) Fruit growth, maturation, and shelf life of two cultivars of tejocote (*Crataegus mexicana* Moc. et Sesse). *Fruits*. <https://doi.org/10.17660/th.2017/72.2.2>
- Vera-Castillo YB, Cuevas JA, Valenzuela-Zapata AG, Urbano B, González-Andrés F (2014) Biodiversity and indigenous management of the endangered non-toxic germplasm of *Jatropha curcas* L. in the Totonacapan (Mexico), and the implications for its conservation. *Genet Resour Crop Evol*. <https://doi.org/10.1007/s10722-014-0109-2>
- Wang N, Zhang C, Qi Y, Li T (2007) Extraction and food chemical characterizations of haw pectins. *Sci Technol Food Ind* 11:87–92
- Ward JH (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58:236–244
- Wilcox D, Dove B, McDavid D, Greer D (1995) Image tool Version 3, Users guide. Univ. Tex. Health Sci. Cent, San Antonio
- Zheng X, Cai D, Potter D, Postman J, Liu J, Teng Y (2014) Phylogeny and evolutionary histories of *Pyrus* L. revealed by phylogenetic trees and networks based on data from multiple DNA sequences. *Mol Phylogenet Evol*. <https://doi.org/10.1016/j.ympev.2014.07.009>
- Zizumbo-Villarreal D, Colunga-García Marín P, Flores-Silva A (2016) Pre-Columbian food system in West Mesoamerica. In: Lira R, Casas A, Blancas J (eds) *Ethnobotany of Mexico interactions of people and plants in Mesoamerica*. Springer, New York, pp 67–83