



RESEARCH ARTICLE

Random amplified polymorphic DNA-based assessment of genetic variation in natural and cultivated allspice (*Pimenta dioica* (L.) Merr.) populations in Guatemala

José Alejandro Ruiz-Chután^{1,3*}, Sara Bethsabe Barrios de León^{2,4}, Evelyn Marie Gutiérrez Rico³, Belém González-Onofre⁴, Carlos Enrique Villanueva-González⁵ & Julio Ernesto Berdúo-Sandoval¹

¹Faculty of Agronomy, University of San Carlos of Guatemala, Guatemala

²Faculty of Chemical Sciences and Pharmacy, University of San Carlos of Guatemala, Guatemala

³Faculty of Health Sciences, Universidad del Istmo, Fraijanes, Guatemala

⁴Grupo Crescentia S.A., Antigua Guatemala, Sacatepéquez, Guatemala

⁵Faculty of Environmental and Agricultural Sciences, Rafael Landívar University, San Pedro Claver S.J. de la Verapaz, Guatemala

*Correspondence email - josealejandro.ruiz@icloud.com

Received: 24 April 2026; Accepted: 18 August 2026; Available online: Version 1.0: 28 September 2026

Cite this article: Ruiz-Chután JA, Barrios de León SB, Gutiérrez Rico EM, González-Onofre B, Villanueva-González CE, Berdúo-Sandoval JE. Random amplified polymorphic DNA-based assessment of genetic variation in natural and cultivated allspice (*Pimenta dioica* (L.) Merr.) populations in Guatemala. Plant Science Today (Early Access). <https://doi.org/10.14719/pst.15226>

Abstract

The fruit of the allspice tree (*Pimenta dioica* (L.) Merr.) is a valuable non-timber forest product traded for its flavour, oils and secondary metabolites. Intense selection pressure from traditional harvesting methods has favoured low-height, high-yield phenotypes, potentially impacting genetic diversity. Although phenotypic variations have been observed, no baseline molecular genetic assessments have been conducted in Guatemala to evaluate the impact of these practices. This study assessed the genetic diversity and structure of natural and cultivated populations of allspice in Guatemala using molecular markers. Seven populations from northern Guatemala, representing a range of ecological conditions, were analysed using 15 random amplified polymorphic DNA (RAPD) primers. A total of 126 amplified fragments, with 88.1 % polymorphism, were obtained. The RAPD markers showed high levels of polymorphism (Polymorphic information content (PIC) = 0.331), good discriminatory power ($D = 0.907$) and reasonable resolving power ($R_p = 3.976$). Estimates of genetic diversity, based on Nei's gene diversity (H_e) and Shannon's diversity index (H), were slightly higher in natural populations ($H_e = 0.24–0.28$; $H = 0.35–0.40$) than in cultivated populations ($H_e = 0.17–0.22$; $H = 0.23–0.31$), suggesting that natural populations maintain greater genetic variability, due to lower levels of human intervention. Population structuring methods, unweighted pair group method with arithmetic mean (UPGMA), discriminant analysis of principal components (DAPC), sparse non-negative matrix factorisation (sNMF) revealed genetic similarity among most populations, with San Cristobal being the exception, showing significant differentiation ($p < 0.05$). These findings highlight the importance of implementing integrated conservation strategies (in situ, circa situ and ex situ methods), as well as sustainable management of naturally regenerated and cultivated populations, to preserve the Guatemalan gene pool of *P. dioica*.

Keywords: genetic resources; genetic structure; molecular markers; non-timber forest products

Introduction

The allspice tree (*Pimenta dioica* (L.) Merr.) is a species naturally distributed in the broadleaf forests of northern Guatemala. Its main product is dried fruits, which are used as a condiment. The taste resembles a mixture of nutmeg and cinnamon, accounting for its name “allspice” on the international market (1). This species is native to the Americas (2), its natural distribution includes the eastern, southeastern and southern regions of Mexico, Central America, Panamá, Cuba, Jamaica and northern South America

However, currently it is only possible to find small natural populations, mostly in Peten (2, 3). In the rest of the country its presence is limited to forest remnants or plant succession. In the Franja Transversal del Norte area, allspice (*P. dioica*) is mainly found

in small commercial crops, family crop plots and backyards; reporting a greater abundance of the species in less conserved sites than in sites considered to be their natural habitat.

Allspice yields essential oils from its fruits, seeds and leaves, which are used as flavouring agents in the manufacture of soaps, hair products, medicine, fungicides and insecticides (1). In 1601, it was introduced to Europe as a condiment and for medicinal use and in 1824 to the Asian countries of Sri Lanka, Singapore, Indonesia and India, where it is one of the main spices grown alongside major commercial crops (4).

In Guatemala, natural populations of this species have been heavily exploited throughout their range as an economical alternative for local communities. Recently, efforts have been made

to boost trade in this species in its main production sites. As a result, producers have selected female trees, that exhibit higher fruit production than male trees and shorter trees to facilitate artisanal harvesting made by climbing trees and shaking the branches.

The consequences of exploitation and selection process on the genetic diversity of cultivated and natural populations of the species are unknown. *Pimenta dioica* is considered as vital non timber forest resource, since its fruits and leaves have various uses, that can be medicinal, aromatic, industrial, artisanal, among others, highlighting its importance as a culinary condiment. The allspice fruit has at least nine reported uses in its original distribution area (1) as essential oil, flavouring, tea leaves; fruit as a condiment; essence for cosmetics and as an insecticide and medicinal.

Traditionally, communities have used this species for different purposes, including chewing fresh green leaves to relieve toothaches; preparing it as a tea to relieve stomach pain and using it as remedy for dysentery and cough. It has also been used to hasten childbirth and to stop menstruation (4). In Guatemala this species has an economic importance since it is a key income for families in the region where it is produced. Allspice fruit is distributed across 5 continents and has high commercial value worldwide.

In forest species, genetic diversity is one of the least studied topics, although it is a crucial component of biodiversity, as it determines the adaptive potential of species. To study genetic diversity, molecular markers are a better approach than morphological markers, as they are not influenced by environmental factors (5, 6). In addition, molecular markers establish the differences between individuals and populations of a species at the DNA level, making it possible to estimate its genetic richness and variability.

For species with limited genomic resources, cost-effective molecular markers remain essential tools for initial genetic assessments, particularly in developing countries where research budgets are constrained (7, 8). Random amplified polymorphic DNA (RAPD) have been utilised among the other kinds of molecular markers to study species for which no previously published genomic data are accessible, as is the case of *P. dioica*. These markers enable data to be retrieved fast and somewhat cheaply by designing certain primers without requiring sequence information (9, 10).

Among the advantages of RAPD are high frequency of polymorphism, methodological simplicity and the requirement of a minimal amount of starting DNA. These characteristics make RAPD markers particularly useful for studies of inter- and intraspecific variability in plant species (11–13). The RAPD markers continue to be effectively employed in baseline genetic diversity studies of tropical tree species where no prior genomic information is available (14, 15). Although simple sequence repeat (SSR) markers can sometimes be transferred across species within a plant family, no validated microsatellite primer sets have been published for *Pimenta* species or confirmed to amplify reliably in *P. dioica*. In addition, next-generation sequencing approaches require bioinformatic infrastructure and reference genomic data not yet available for this genus in its native range.

No studies have been conducted in Guatemala to date to assess the genetic diversity of this important forest resource. Since no prior molecular studies on *P. dioica* have been undertaken in Guatemala, the purpose in this work was to assess the efficacy of RAPD markers in collecting genetic variation within this species. With these markers, the genetic diversity and structure of natural and

cultivated populations of *P. dioica* was then assessed as an approximation of the status of the species' genetic pool in Guatemala.

It was hypothesised that natural populations exhibited higher genetic diversity because they are not subjected to selection pressure and/or fruit extraction. The generated data can be used to develop management and exploitation plans aimed at preserving the genetic resources of *P. dioica*. Additionally, this study provides insights into the utility of RAPD markers for future genetic studies on this and related species in the region.

Materials and Methods

Study area characteristics

The study encompassed seven *P. dioica* populations across northern Guatemala, representing diverse ecological conditions. The Peten population is in the Tropical Moist Forest life zone at approximately 400 m elevation, with mean annual temperatures of 24–26 °C and rainfall of 1600–2000 mm. Soils are well-drained and limestone-derived. The remaining populations are situated in the Franja Transversal del Norte (FTN) region, spanning a range of elevations, microclimates and life zones. San Cristobal, found in the Subtropical Lower Montane Moist Forest zone, is at higher elevations (around 1000 m) with cooler temperatures (18–22 °C) and higher rainfall (2000–2500 mm annually). Coban and Coban FTN, both within the Subtropical Moist Forest zone, share similar mid-elevation conditions (400–800 m) with mean temperatures of 20–24 °C and annual rainfall of 2000–2500 mm.

Chisec, Raxruhá and Fray are located at lower elevations (300–500 m) in the FTN, in the transition between Tropical Moist Forest and Subtropical Wet Forest zones, experiencing warmer temperatures (22–26 °C) and variable rainfall (1800–2500 mm). Soils in the FTN populations range from clay-rich in higher elevations to more varied compositions in lower areas, reflecting the diversity of these life zones (16).

Plant material

A total of 69 *P. dioica* trees from seven populations were sampled (Fig. 1). All samples were positioned geographically using a portable global positioning system (GPS) device. Three young, fresh leaves were taken from each tree, desiccated using silica gel, placed in labeled plastic pouches and transported to the Plant Biotechnology Laboratory at the Faculty of Agronomy, Universidad de San Carlos de Guatemala (USAC), where DNA extraction and genotyping were carried out.

Trees were selected to maximise spatial coverage within each population, maintaining a minimum inter-individual distance of approximately 50 m to reduce the probability of sampling clonally related individuals. Since *P. dioica* is a dioecious species and female trees are the primary targets of anthropogenic selection and fruit harvesting in Guatemala, sampling was restricted to female trees. While this approach captures the genetic diversity most relevant to current management practices, it does not encompass the male component of the gene pool, which is acknowledged as a limitation and should be addressed in future surveys.

A georeferenced map of sampling localities was produced using R v.4.2.0. Geographic data for Guatemala was obtained using the "ne_states" function from the *natnat* package v.1.0.1 (17). The *elevatr* package v.0.99.0 (18) was used to obtain topographical

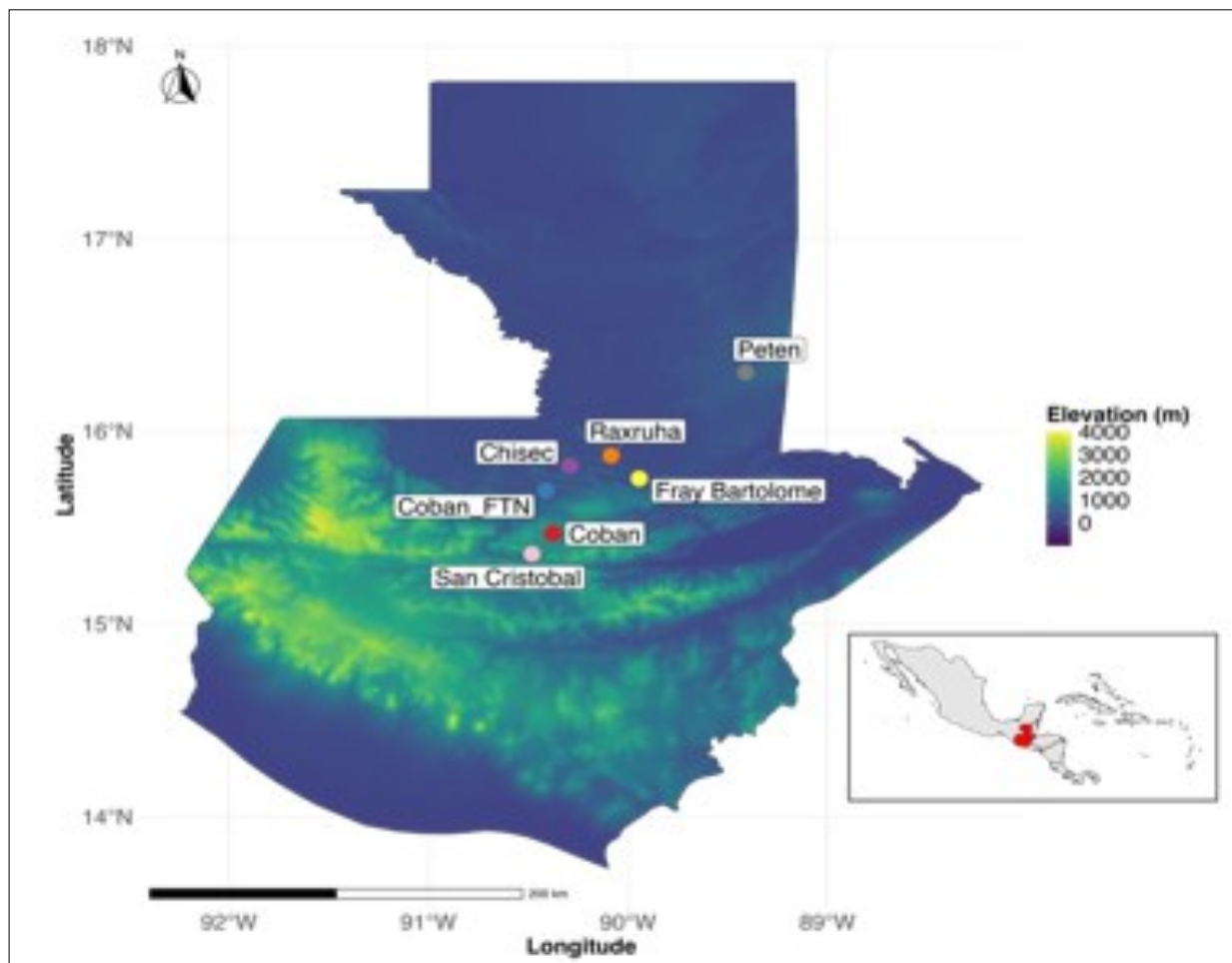


Fig. 1. Sampling locations of 69 *Pimenta dioica* trees across 7 geographic populations in Guatemala, representing four Holdridge life zones: Tropical Moist Forest (Peten), Subtropical Lower Montane Moist Forest (San Cristobal), Subtropical Moist Forest (Coban and Coban FTN) and a transition between Tropical Moist Forest and Subtropical Wet Forest (Chisec, Raxruha and Fray Bartolome).

data at zoom level 7. The raster dataset that was produced was then reduced to the departmental limits of Guatemala and transformed into a data frame structure. The final visualisation (Fig. 1) that shows the sample coordinates for each population was created using ggspatial v.1.1.9 (19) and ggplot2 v.3.5.1.

DNA extraction and amplification

A commercial plant genomic DNA extraction kit was used for genomic DNA extraction following the manufacturer's instructions. The DNA quality and concentration was assessed using a fluorometric quantification system.

Polymerase chain reaction (PCR) amplification was performed in a total reaction volume of 25 μ L containing 25–30 ng of genomic DNA (1 μ L), 2.5 μ L of 10x reaction buffer, 1.5 μ L of 25 mM magnesium chloride ($MgCl_2$), 0.1 μ L of 10 mM deoxyribonucleotide triphosphate (dNTP), 0.5 μ L of primer and 0.2 units of Taq DNA polymerase. Polymerase chain reactions followed the protocol described in previous research (20). The initial denaturation was performed at 94 $^{\circ}$ C for 4 min, followed by 40 cycles of denaturation at 94 $^{\circ}$ C for 1 min, annealing at 37 $^{\circ}$ C for 45 sec and extension at 72 $^{\circ}$ C for 2 min, followed by a final extension at 72 $^{\circ}$ C for 10 min. Fifteen primers (OPB-01, OPB-02, OPB-04, OPB-05, OPB-06, OPB-08, OPB-10, OPB-11, OPB-12, OPB-13, OPB-14, OPB-15, OPB-16, OPB-17 and OPB-18) were selected from standardisation assays.

Amplification products were separated on 1.5 % agarose- tris-borate-EDTA (TBE) 1x gels, stained with ethidium bromide (0.1 μ g/mL) and visualised with a 100 bp DNA size marker. The RAPD fragments were scored manually by two independent

observers as present (1) or absent (0). Only clearly defined, unambiguous bands were retained; faint, smeared or inconsistently sized bands were systematically excluded to minimise scoring errors. To assess reproducibility, PCR amplification was independently repeated for approximately 30 % of the samples; identical banding patterns were obtained in all cases, confirming the reliability of the data generated.

The RAPD markers were chosen due to their efficiency in assessing genetic diversity without prior sequence information, their rapid generation of polymorphic data and their effectiveness in detecting genetic variability. As this represents the first genetic study of this species in the region, the use of RAPD markers was considered appropriate, given limited genomic information available.

Data analysis

Analysis of the effectiveness of random amplified polymorphic DNA markers

With the banding pattern obtained from the agarose gel, a binary matrix of presence (1) absence (0) was generated for each of the amplified loci. From this binary presence/absence matrix, the following metrics were estimated. 1) Expected heterozygosity (H_e): probability that an individual is heterozygous for the locus in the population; 2) Polymorphic information content (PIC): probability that the genotype of a given offspring allows one to deduce, in the absence of intercrossing, which of the two parental alleles it received (21); 3) Marker index (MI): product of the effective multiple ratio and the average expected heterozygosity for polymorphic markers; 4) Discriminatory power (D): probability that two randomly selected

individuals show different banding patterns and are therefore distinguishable from each other; 5) Resolving power (R_p): ability to distinguish between two analysed samples. The above indices were calculated using the R v.4.2.0 program based on the code implemented in the iMEC online program (22).

Genetic diversity

The binary matrix was converted into a *genind* class object using the *poppr* package v.2.9.4 implemented in R v.4.2.0. To determine genetic diversity, H , H_e (Nei's diversity) were calculated with the *poppr* package and the percentage of polymorphic loci, the number of different alleles and the number of effective alleles with the *GenoDive* v.3.05 program (23). Allele frequencies at each RAPD locus were estimated under the assumption of Hardy-Weinberg equilibrium (HWE) using the square-root method, as described by previous researchers (24). The frequency of the null allele (q) was estimated as $q = \sqrt{f_0}$, where f_0 is the observed frequency of the null phenotype (band absence) within each population. This approach, implemented within the *poppr* package v.2.9.4 for dominant marker data, provides diversity estimates consistent with those reported in comparable RAPD-based population genetic studies.

Population structure

To study genetic structure, samples with genetic traits were grouped using hierarchical cluster analysis. Nei's genetic distance was estimated with the *poppr* program and the bootstrap statistical support value using 1000 replicas using the unweighted pair group method with arithmetic mean (UPGMA) hierarchical clustering approach. For last customising and visualisation, the web program Interactive Tree of Life (iTOL) v4 (25) was employed. This operation was carried out both at the level of individuals and populations.

The "find.cluster" function included in the *ade4* package v.2.1.6 was used to explore the population structure of the sampled trees in-depth by applying discriminant analysis of principal components (DAPC). The acquired clusters were first shown using a scatter plot over the first two linear discriminants then further shown with a bar chart displaying the degree of genetic mixing. The "optim.a.score" tool confirmed the proper number of discriminant functions and main components to be kept.

The sparse non-negative matrix factorisation (sNMF) technique accessible through the *sNMF* function in *LEA* package v.1.6.0 within the R environment was used to analyse individual genetic heritage in further addition to the DAPC findings. Not requiring HWE assumptions (26), this non-parametric alternative to the *STRUCTURE* technique (27) has advantages over others. In the implementation, the ancestry coefficients were estimated for many plausible ancestral populations (K). To consider several random initialisations, the sNMF method was implemented 1000 times for every K value (from 1–10). The cross-entropy criteria were applied to evaluate the predictive capacity of the model by hiding some genotypes and quantifying imputation errors using the estimated ancestry coefficients to find the ideal number of ancestral populations.

Employing a variety of analytical methods is a thorough approach to population structure characterisation. Through molecular marker research, these approaches have been well confirmed for their ability to reliably show genetic links between populations (28).

Population differentiation

Using the *poppr* program, an additional analysis of molecular variance (AMOVA) was performed to explore population genetic structure. To clarify patterns of gene flow and population divergence, fixation indices Φ were calculated from the covariance components (29). Statistical significance was established with 1000 permutations and randomisation testing. The formula $N_m = 0.5 (1 - \Phi_{ST}) / \Phi_{ST}$ was used to calculate the amount of gene flow (N_m) between populations. The Mantel test was used to assess isolation by distance, applying 10000 permutations with the *ade4* package v.1.7 (30). Pairwise population differentiation was examined using the coefficient of genetic differentiation (G_{ST}) index with 1000 permutations utilising the *mmod* package v.1.3.3 and Nei's standard genetic distance using *poppr* package.

Results

Analysis of the effectiveness of random amplified polymorphic DNA markers

The detailed polymorphic indices of the evaluated RAPD primers are given in Table 1. A total of 126 amplified bands were obtained with the 15 primers evaluated. For each primer, the PIC is an indicator of the diversity and frequency of alleles generated among the samples. The average PIC was 0.3319 and the highest value was recorded for OPB-02 (PIC=0.3756) followed by OPB-08 (0.3533), while the lowest PIC was presented by OPB-13 (0.303). Heterozygosity ranged from 0.2896 (OPB-14) to 0.4786 for OPB-02 with an average of 0.4106 per primer.

The MI was calculated to recognise the utility of each RAPD primer in *allspice*, which was highest for OPB-02 (MI = 0.0027) followed by OPB-08 (MI = 0.0026), while the minimum was presented by OPB-06 (MI = 0.00011) with an average of 0.0267 per primer. To determine the depth accuracy of the primer, the D was determined with an average index of $D = 0.907$ and ranging from 0.843 (OPB-06) to 0.969 (OPB-08). The highest value of R_p was 6.927 (OPB-02), while the lowest was $R_p = 1.536$ (OPB-16) with an average of 3.976 per primer.

Genetic diversity

A total of 126 amplified bands were obtained with the 15 primers evaluated. The minimum, maximum and average percentage of polymorphic loci was 62.79, 80.16 and 71.77 % respectively. The genetic diversity indices, including measures of genotypic richness and abundance, showed high genetic diversity in all populations. The number of different alleles ranged from 1.31 (Chisec) to 1.61 (San Cristobal), the number of effective alleles from 1.17 (Coban) to 1.44 (San Cristobal), the H from 0.23 (Coban) to 0.40 (Peten) and H_e (Nei's diversity) between 0.17 (Chisec) and 0.28 (San Cristobal) (Table 2).

Population differentiation

Analysis of molecular variance was conducted to assess the genetic structure of *P. dioica* populations (Table 3). After grouping according to geographical populations, most genetic variation was found within populations (88.69 %), with smaller proportions attributed to differences between populations within regions (9.69 %) and between regions (1.62 %). The overall population differentiation (Φ_{ST}) was 0.103, indicating moderate genetic structuring.

Table 1. Polymorphism statistics for random amplified polymorphic DNA primers used in allspice (*Pimenta dioica*)

Primer	TNB	NPB	H_e	PIC	MI	D	Rp
OPB-01	8	7	0.4176	0.3304	0.0017	0.911	4.753
OPB-02	13	13	0.4786	0.3756	0.0027	0.946	6.927
OPB-04	12	11	0.4116	0.3328	0.0017	0.916	5.304
OPB-05	8	7	0.4369	0.3221	0.0020	0.896	2.811
OPB-06	8	8	0.3562	0.3541	0.0011	0.843	2.811
OPB-08	11	10	0.3638	0.3553	0.0026	0.969	6.376
OPB-10	7	6	0.4037	0.3361	0.0016	0.921	3.072
OPB-11	8	6	0.4185	0.3300	0.0018	0.911	3.246
OPB-12	12	9	0.4205	0.3291	0.0018	0.909	4.579
OPB-13	6	5	0.3821	0.3030	0.0014	0.934	2.144
OPB-14	6	4	0.2896	0.3514	0.0012	0.943	3.681
OPB-15	7	7	0.4696	0.3073	0.0025	0.858	3.130
OPB-16	3	3	0.3584	0.3446	0.0012	0.945	1.536
OPB-17	6	5	0.4742	0.3051	0.0023	0.851	2.985
OPB-18	11	10	0.4774	0.3036	0.0027	0.845	6.289
Mean	8.4	7.4	0.4106	0.3319	0.0267	0.907	3.976

TNB: total number of bands; NPB: number of polymorphic bands; H_e : expected heterozygosity; PIC: polymorphism information content; MI: marker index; D: discriminant power; Rp: resolving power.

Table 2. Geographical location and summary statistics on genetic diversity of *Pimenta dioica* populations in Guatemala

Population	Status	Region	Latitude	Longitude	N	PLP	Na	Ne	H	H_e
Peten	Natural	North	16.30	-89.40	10	73.81	1.50	1.39	0.40	0.24
San Cristobal			15.36	-90.48	10	80.16	1.61	1.44	0.35	0.28
Coban			15.47	-90.37	10	69.84	1.39	1.17	0.23	0.18
Chisec	Cultivated	FTN	15.82	-90.29	10	62.70	1.31	1.37	0.30	0.17
Raxruha			15.87	-90.08	10	73.81	1.46	1.35	0.26	0.19
Coban FTN			15.69	-90.41	9	71.43	1.42	1.33	0.29	0.20
Fray			15.75	-89.94	10	70.63	1.47	1.32	0.28	0.20
Mean						71.77	1.45	1.34	0.30	0.21

FTN: Franja Transversal del Norte; N: number of samples; PLP: percentage of polymorphic loci; Na: number of different alleles; Ne: number of effective alleles; H: Shannon-Wiener diversity index; H_e : Nei 1978 genetic diversity or expected heterozygosity.

When samples were grouped by population status, a similar pattern emerged, with most variation occurring within population types (89.96 %). The proportion of variation among population status (10.03 %) was higher than that observed between geographical regions. The Φ_{CT} value for population status (0.114) was also higher and highly significant ($p = 0.001$), suggesting that the type of population accounts for more genetic differentiation than geographic location alone.

Gene flow and pairwise genetic differentiation analyses revealed a complex pattern of population relationships (Fig. 2). The number of migrants (N_m) value between populations was 4.35. The coefficient of genetic differentiation (G_{ST}) values ranged from 0.105–0.577, with the highest differentiation observed between Raxruha and San Cristobal ($G_{ST} = 0.577$) and the lowest between Peten and San Cristobal ($G_{ST} = 0.105$). Nei's genetic distances showed a similar pattern, spanning from 0.050–0.598. The greatest genetic distance was found between San Cristobal and Fray (0.598), while Raxruha and Coban FTN exhibited the closest genetic relationship (0.050).

The San Cristobal population consistently displayed high G_{ST} values and Nei's genetic distances when compared to other populations, suggesting it may be the most genetically distinct group. In contrast, geographically proximate populations such as Coban and Coban FTN generally showed lower genetic differentiation and distances. The Peten population exhibited moderate levels of differentiation and distance from other populations, with G_{ST} values ranging from 0.105–0.326 and Nei's genetic distances from 0.181–0.326. All G_{ST} values and Nei's genetic distances were statistically significant ($p < 0.05$).

These results indicate a complex pattern of genetic structure among *P. dioica* populations, with both geographic proximity and potential isolation playing roles in shaping genetic relationships. The Mantel test determined that populations did not exhibit isolation by distance (IBD) ($r = 0.047$, $p > 0.05$) (Fig. 3).

Population structure

The UPGMA analysis to determine the genetic relationship between populations showed the formation of two clades with a bootstrap support value of 100 after 1000 iterations. The first clade grouped the populations of Chisec, Coban, Fray, Peten, Coban FTN

Table 3. Analysis of molecular variance (AMOVA) for seven *Pimenta dioica* populations. Sigma values represent within-population variance and demonstrate the relative contribution of each variance component to the total. Population differentiation is measured using Phi (Φ) statistics, with greater values indicating more distinct population structuring

Source of variation	Sigma	%	Φ Phi statistics	p-value
A: when the samples trees were grouped according to geographical populations				
Between region	0.38	1.62	$\Phi_{CT} = 0.012$	0.01
Between populations within regions	2.28	9.71	$\Phi_{CS} = 0.089$	0.001
Within populations	20.82	88.67	$\Phi_{ST} = 0.103$	0.001
Total	23.48	100		
B: when the samples trees were grouped according to type of status				
Among type population	2.16	10.06		
Within type populations	19.40	89.94	$\Phi_{CT} = 0.114$	0.001
Total	21.57	100		

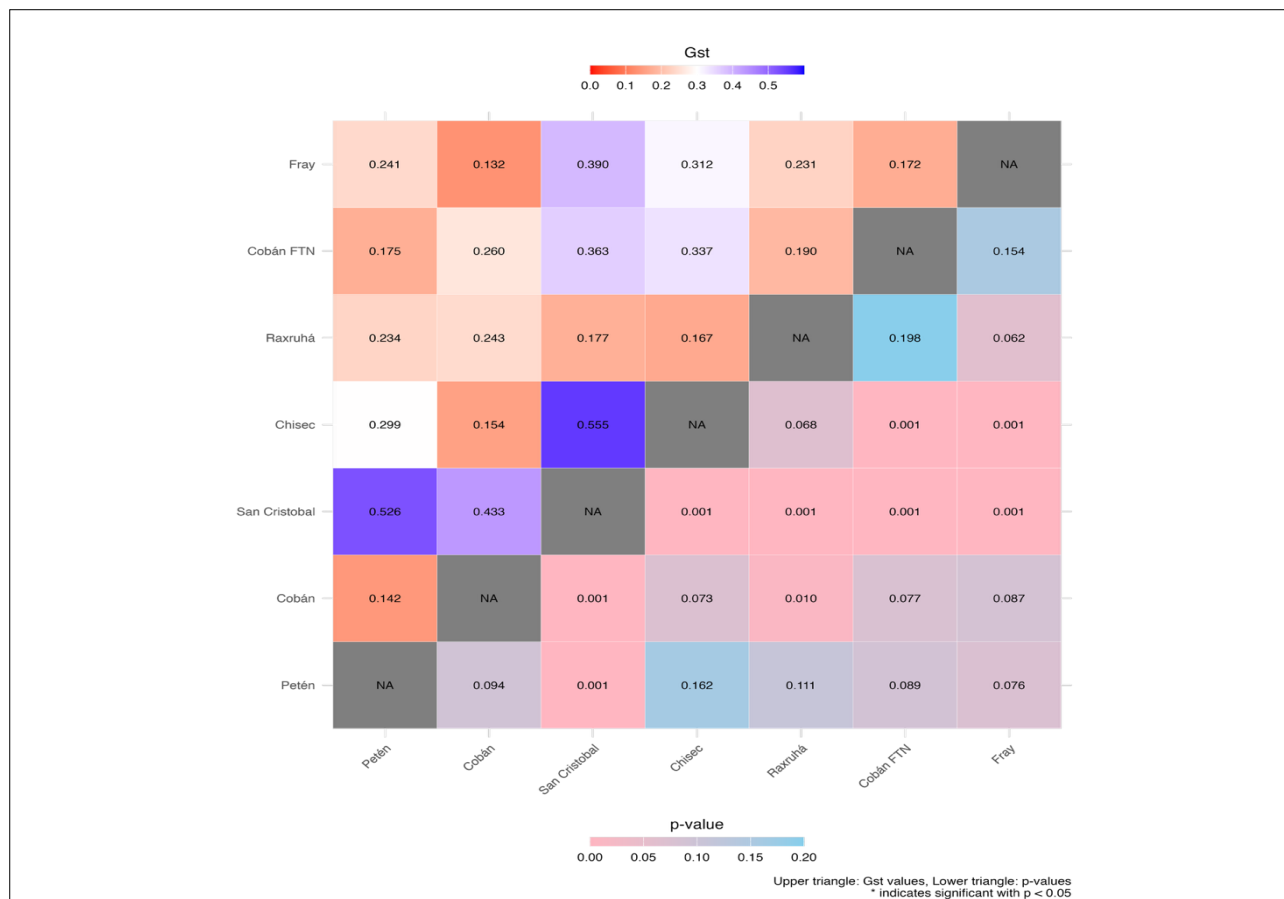


Fig. 2. Heatmap of genetic differentiation and distance among *Pimenta dioica* populations. G_{ST} values are shown in the upper triangle and Nei's standard genetic distances are shown in the lower triangle. The colour gradient represents the magnitude of differentiation and distance, with darker colours indicating greater genetic variation between populations. Diagonal values are NA, representing within-population comparisons.

and Raxruhá, while the second clade included only the population of San Cristobal (Fig. 4A). At the individual level, the analysis of genetic relationships reveals the formation of two clades. The dendrogram showed that one of the clades included mainly all samples from the San Cristobal population. In contrast, the second clade exhibited a mixture of individuals from the remaining 6 populations (Fig. 4B).

The principal component analysis (PCA) plot illustrates the genetic structure of *P. dioica* populations, with Axis 1 and Axis 2 explaining 18.4 % and 8.7 % of the total genetic variation respectively. The plot reveals distinct clustering patterns among the seven studied populations (Fig. 5). San Cristobal population forms a tight, well-defined cluster in the upper right quadrant, indicating its genetic distinctiveness from other populations. This suggests that San Cristobal may have experienced limited gene flow with other populations or has undergone unique selective pressures.

Chisec and Petén populations are widely dispersed across the left side of the plot, suggesting higher intra-population genetic diversity. Their overlapping distribution indicates a degree of genetic similarity between these two populations. Cobán, Raxruhá, Cobán FTN and Fray populations occupy intermediate positions in the plot. While they show some level of overlap, indicating genetic similarities, they also maintain distinct centroids, suggesting some degree of genetic differentiation. Cobán FTN appears slightly more isolated, positioned towards the bottom of the plot.

The genetic structure of *P. dioica* populations was evaluated using DAPC (Fig. 6). The scatter plot (A) depicts the distribution of individuals along the first two discriminant functions, which collectively explain 78 % of the genetic variation (DF1: 54 %, DF2: 24 %). This analysis reveals distinct genetic structuring among the populations, with San Cristobal (purple) forming a clearly separate cluster on the far right, indicating its genetic uniqueness. Other populations exhibit varying degrees of overlap, particularly Chisec (green) and Cobán (light blue), which show the widest dispersion, suggesting higher intra-population diversity.

The circular plot (B) provides a detailed visualisation of individual genetic assignments, where each segment represents an individual and colours correspond to their genetic composition. This representation further emphasises San Cristobal's genetic distinctiveness through its uniform purple colouration. In contrast, other populations display varying levels of admixture, most notably observed in Petén, Fray and Cobán FTN. Collectively, these results indicate a complex genetic landscape within *P. dioica* populations, characterised by a pronounced differentiation of San Cristobal and varying degrees of genetic similarity and admixture among the remaining populations, likely reflecting the interplay of isolation and gene flow in shaping the species' genetic structure in Guatemala.

Genetic structure was also assessed using sNMF analysis. The cross-entropy plot (Fig. 7A) suggests the optimal number of ancestral populations to be three ($K=3$), as indicated by the lowest cross-entropy value. This finding is represented in the circular bar

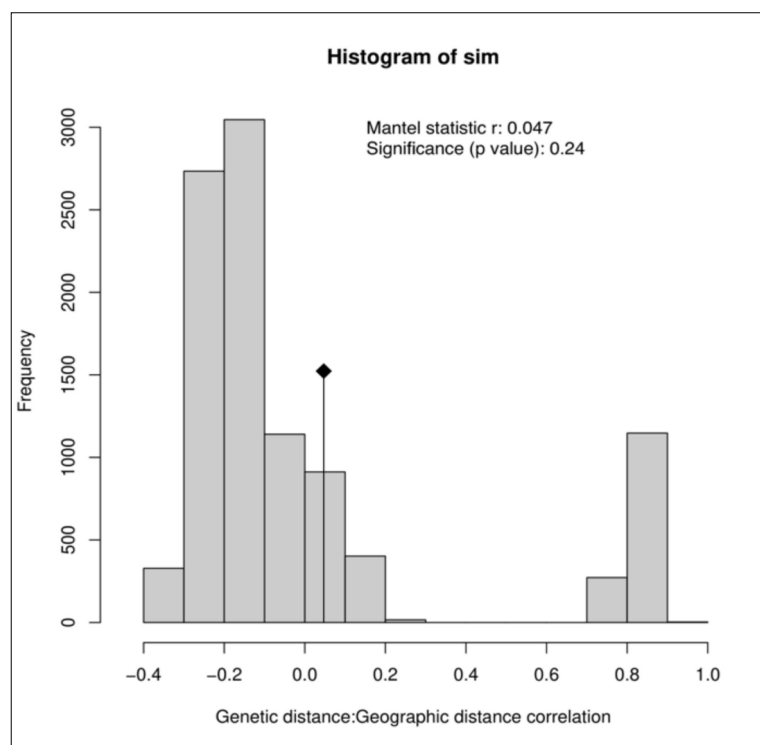


Fig. 3. Isolation-by-distance analyses (IBD) with Mantel test results showing the correlation between empirical genetic distance and geographic distance, presented alongside a histogram of Monte Carlo simulations under the assumption of no spatial structure.

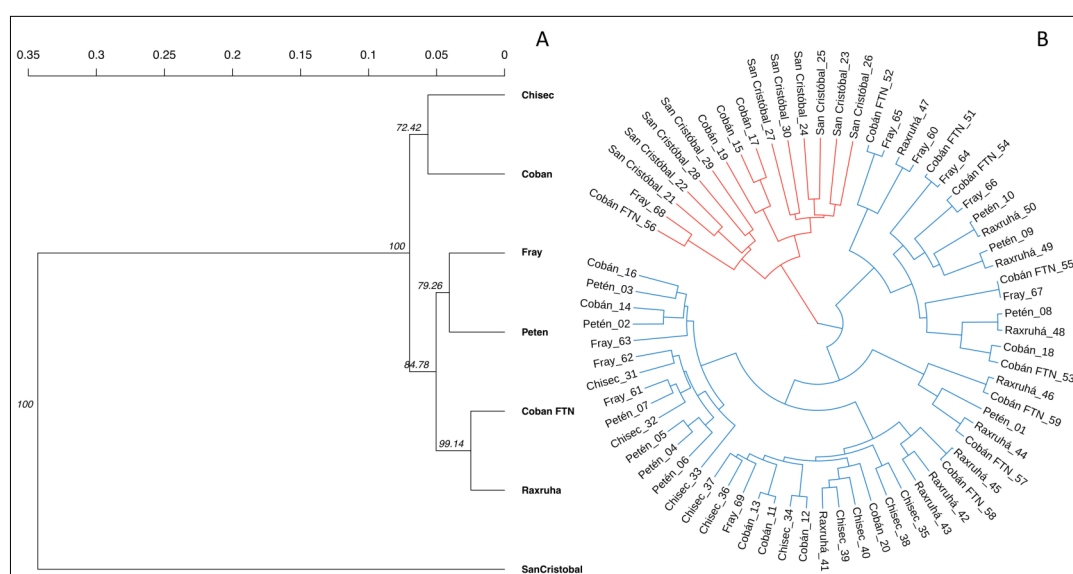


Fig. 4. Unweighted pair-group method with arithmetic mean dendrogram based on Nei's genetic distance at the hierarchical level of (A) populations; (B) individuals of *Pimenta dioica*. Values at branch nodes represent the bootstrap statistical support of 1000 iterations.

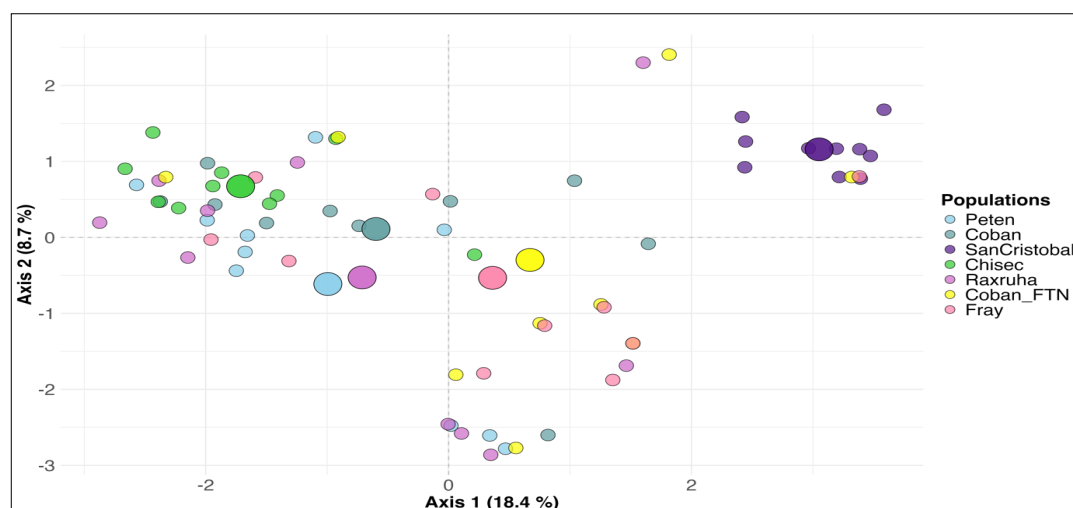


Fig. 5. Genetic structure of *Pimenta dioica* populations visualised by principal component analysis.

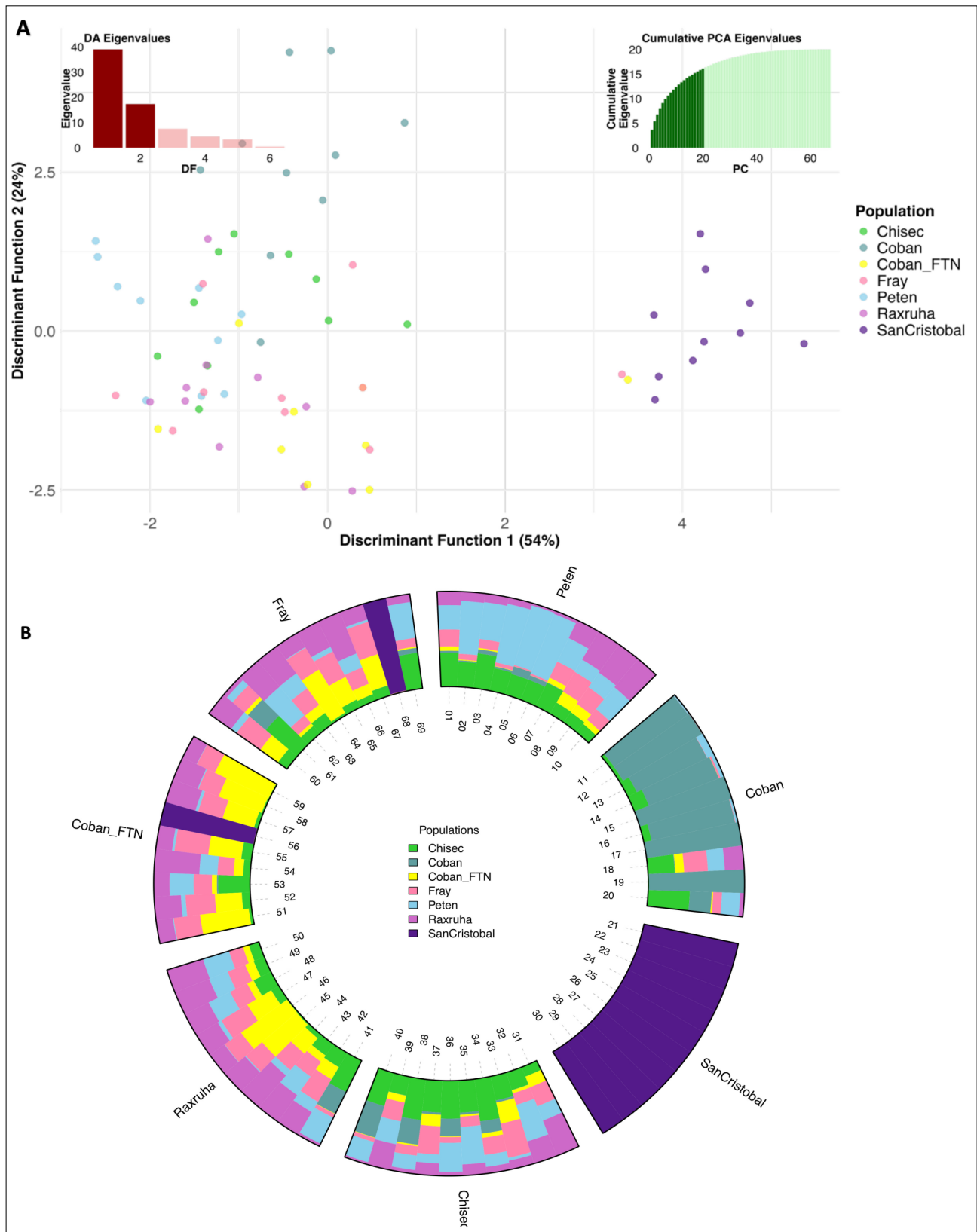


Fig. 6. Population genetic structure, obtained from a discriminant analysis of principal components analysis of 69 *Pimenta dioica* individuals from 7 different populations. A: scatter plot of the first two discriminant functions. Each colour represents a population shown as ellipses; B: the bar chart of the discriminant analysis of principal components analysis shows a coloured vertical line representing each individual. The same colour in different individuals indicates that they belong to the same group.

plot (Fig. 7B), which displays the assignment probabilities of individuals to each of the three inferred genetic clusters. The bar plot reveals a complex pattern of genetic structure across the seven sampled populations. Most notably, the San Cristobal population exhibits a highly distinct genetic composition, dominated almost entirely by a single cluster (purple).

The genetic isolation of the San Cristobal population appears to reflect a unique evolutionary history distinct from all other sampled localities. In contrast, other populations show varying degrees of admixture between the three genetic clusters. Peten and Coban display a predominance of the light blue cluster, with some admixture from the other two clusters. Chisec and Raxruha are characterised by a more balanced mixture of the light blue and teal clusters, with occasional contributions from the purple cluster. Interestingly, Coban FTN and Fray show a notable presence of all three genetic clusters, indicating a higher level of admixture.

Discussion

Several criteria demonstrated the effectiveness of RAPD markers in evaluating the genetic diversity of *P. dioica* in Guatemala. Comparable to research in other Neotropical Forest species (6), the great degree of polymorphism (88 %) demonstrated that RAPD markers efficiently identify genetic variants in this species. The mean predicted H_e of 0.4106, obtained across many marker systems, indicated modest genetic diversity in Guatemalan *P. dioica* populations, in line with values recorded in native woody plants from Central America (8).

The findings match those of research on other native tree species, like *Cordia alliodora* (Ruiz & Pav.) Oken, which revealed comparable degrees of genetic variation around Central America (31). While moderate, the average PIC of 0.3319 combined with the high D of 0.907 and Rp of 3.976 showed these markers' efficacy in separating individuals or populations (6, 21). For evaluating genetic diversity in *P. dioica*, primers OPB-02, OPB-08 and OPB-18 demonstrated especially high values across many criteria.

Within the Myrtaceae family, RAPD-based studies of related genera have reported comparable patterns. Previous researchers (32) documented substantial inter-population genetic differentiation in *Syzygium cumini* (L.) Skeels (Myrtaceae) correlated with ecological zone and population size, with smaller and more isolated populations showing reduced diversity—a pattern analogous to the contrast between San Cristobal and the FTN cultivated populations observed in *P. dioica*. Similarly, some earlier studies (33) demonstrated that RAPD markers effectively discriminated wild and cultivated genotypes of *Psidium guajava* L. with wild accessions retaining broader allelic diversity than commercial cultivars. These findings corroborate the broader trend observed in outcrossing tropical trees subject to anthropogenic selection, whereby the repeated use of a narrow set of preferred genotypes tends to homogenise allele frequencies in managed populations while unmanaged natural stands retain greater allelic richness.

The RAPD markers have proven helpful in initial genetic diversity studies of various forest species, including *Pinus roxburghii* Sarg. (34), *Calyculophyllum spruceanum* (Benth.) Hook.f. ex K.Schum. (35) and *Cedrela odorata* L. (10). The RAPD proved to be a practical

choice for investigating genetic diversity for *P. dioica*, when genetic sequence information is absent. Although the relatively low MI of 0.0267 suggests the potential for developing more informative marker systems in future research, the overall performance indicates that RAPD markers are suitable for preliminary genetic diversity studies in *P. dioica*, particularly when resources for more advanced molecular technologies are limited.

While more informative marker systems such as SSRs and single nucleotide polymorphisms (SNPs) would provide higher resolution for population genetic analyses, the absence of genomic resources for *P. dioica* and budget constraints common to research institutions in the species' native range made RAPD markers a practical choice for this initial assessment. The results demonstrated that RAPD markers successfully captured biologically meaningful patterns of genetic structure, providing essential baseline data for conservation planning. Future studies with co-dominant markers would complement these findings by enabling more precise estimates of heterozygosity and inbreeding coefficients.

An additional methodological limitation concerns the estimation of expected H_e from dominant RAPD markers. Since dominant markers cannot distinguish heterozygotes from dominant homozygotes, H_e estimation requires the assumption of HWE within populations (24, 36). Deviations from HWE—which may occur in small, fragmented or intensively managed populations—could bias H_e estimates (36, 37). This limitation is inherent to dominant marker systems (24, 37) and further underscores the value of future studies employing codominant markers to provide empirically validated diversity estimates for *P. dioica*.

The present work is a first attempt to assess population structure and genetic diversity of *P. dioica* over a large natural distribution area in Guatemala using molecular markers. The RAPD molecular markers revealed notable levels of polymorphism (71.77 %) across the investigated populations. Peten and San Cristobal had the highest Shannon and Nei's index. These results reflect what is expected of wild populations minimally influenced by artificial selection (35).

Peten and San Cristobal's natural regeneration populations, which face little fruit extraction pressure, showed somewhat greater genetic diversity when diversity metrics (percentage of polymorphic loci (PLP), Nei's gene diversity (H_e) and expected heterozygosity (H_e)) were compared against cultivated populations. This trend can be explained by ecological as well as human-related aspects. From an ecological standpoint, Peten and San Cristobal respectively occupy separate Holdridge life zones: Tropical Moist Forest and Subtropical Lower Montane Moist Forest. These different environmental variables probably help to maintain different genetic pools fit for their specific situation.

The elevated genetic diversity in pristine ecosystems likely reflects adaptations to a wider spectrum of ecological niches (37). Since they are not under selection pressure or fruit extraction, further research on forest species has demonstrated that natural populations usually preserve more genetic diversity levels than their farmed counterparts (38, 39). The preservation of genetic variety in natural populations is clearly essential for their ongoing existence since it reduces local extinction risk under natural conditions.

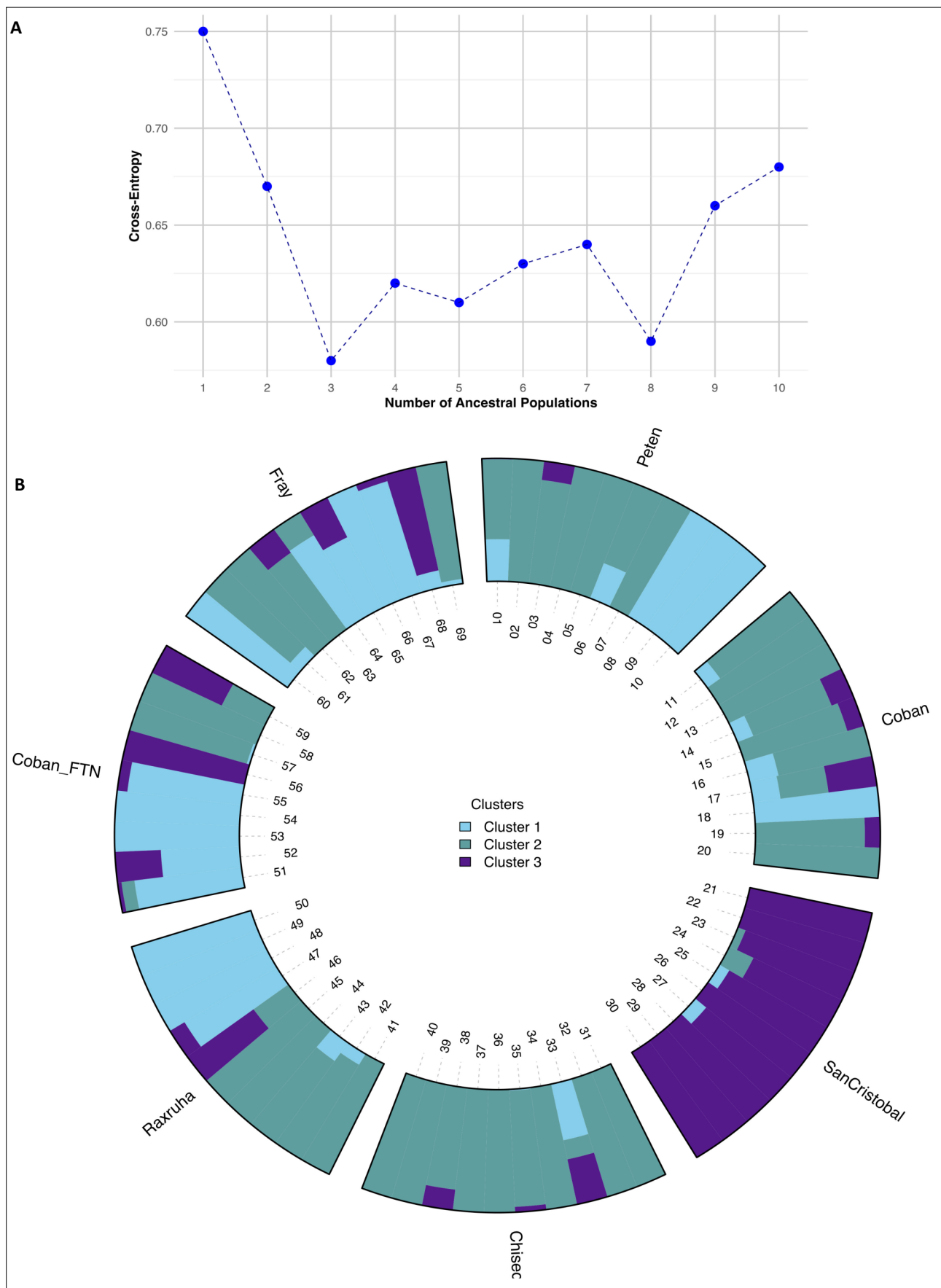


Fig. 7. Assessment of the genetic structure of *Pimenta dioica* populations. A: cross-entropy plot indicating three ancestral populations; B: barplot displaying the assignment probability of each individual into one of the inferred genetic clusters using sNMF.

Cultivated populations in the Subtropical Moist Forest zone (Coban and Coban FTN) and the transition zone between Tropical Moist Forest and Subtropical Wet Forest (Chisec, Raxruha and Fray) show rather less genetic diversity. This scenario implies that increasingly homogenous ecological conditions and artificial selection for harvesting reasons have combined to lower allele richness and homogenise populations, therefore affecting the genetic diversity levels of the species (40, 41).

Based on field observations, selected genotypes are usually highly productive and, in many cases, are shorter than natural trees, which facilitates the collection of fruit. The more uniform environmental pressures in these cultivated areas, coupled with human selection for desirable traits, may have led to a narrowing of the genetic base in these populations (39, 41). However, maintaining this diversity in cultivated populations or populations with some degree of management, is also essential. These can help keep the species' genetic pool healthy, especially if it has been severely reduced from its natural range, as is the case of *P. dioica*.

The UPGMA and DAPC, revealed that the 69 samples of *P. dioica* form two well-defined groups, showing a clear separation of the San Cristobal population from the rest of the populations, which were concentrated in a second group. The results obtained by the sNMF were similar. This structure is consistent with the fact that San Cristobal is a remnant population with little or no management, which is why it essentially represented the native genotypes of the FTN.

The genetic distinctiveness of San Cristobal can also be understood in terms of ecological and geographic factors. At approximately 1000 m above sea level (a.s.l.), this population occupies the Subtropical Lower Montane Moist Forest life zone, characterised by cooler mean temperatures (18–22 °C) and higher annual rainfall (2000–2500 mm) than the remaining FTN localities, situated at 300–500 m with temperatures of 22–26 °C and annual rainfall of 1800–2000 mm. These contrasting environmental conditions may have promoted divergent selective pressures over time, favouring locally adapted genotypes suited to cooler and more humid conditions. Additionally, the topographic relief surrounding San Cristobal may have historically restricted natural seed dispersal and pollinator movement between this site and lower-elevation populations, reinforcing genetic drift as a contributing mechanism to the differentiation observed in *P. dioica*.

Field observations and interviews with local producers revealed that the rest of the localities in the FTN contain germplasm introduced from Peten. This introduced germplasm shows favorable phenotypic characteristics, specifically shorter trees that are highly productive. This situation suggested that intentional selection and introduction affect population structure (35) and explains the weak pattern of population structuring observed in the present study. The low level of variation between the regions detected by AMOVA can be explained by the high value of gene flow ($N_m = 4.35$) generated by the absence of population structure (42). While *P. dioica* has limited natural seed dispersal (1), the field documentation indicated that producers actively mobilise its germplasm across longer distances.

It has been noted that when producers find a specimen with beneficial agronomic traits, they often cut scions for grafting

or move young trees from one locality to another, facilitating gene flow between populations. Additionally, the high gene flow observed despite the populations spanning different Holdridge life zones suggests that *P. dioica* possessed considerable adaptive plasticity. This ability to thrive across various ecological conditions, from Tropical Moist Forest to Subtropical Lower Montane Moist Forest, may contribute to maintaining genetic connectivity across populations, even in the face of diverse environmental pressures.

The paired comparison analysis between FTN populations and the Peten population did not show significant differences ($p < 0.05$), except for the San Cristobal population, which is a remnant of the original FTN gene pool. For this reason, excluding the San Cristobal population, the FTN localities could be considered subpopulations of the Peten population; however, this cannot be guaranteed without a more exhaustive sampling of the latter department to detail its genetic reservoir.

The genetic similarity among FTN populations, despite spanning a transition between Tropical Moist Forest and Subtropical Wet Forest zones, indicated that recent human-mediated gene flow may be overriding historical patterns of ecological adaptation. However, the long-term consequences of translocating genotypes across these ecological boundaries should be monitored, as it may affect local adaptation processes.

Although FTN producers have intentionally reduced natural populations to introduce shorter phenotypes that facilitate fruit harvest, the studied populations of *P. dioica*, both natural and cultivated, exhibit moderately high levels of diversity (PLP, H, H_e). It suggested that both types of populations can be used as genetic reservoirs. However, selection pressure due to favoring the introduction of some genotypes with better phenotypic characteristics may decrease genetic diversity in the future, especially in cultivated populations.

The findings on the genetic diversity and structure of *P. dioica* populations in Guatemala have important implications for conservation strategies. The slightly higher genetic diversity observed in natural populations, coupled with the distinct genetic structure of certain populations like San Cristobal, underscores the need for a multi-faceted conservation approach. In situ conservation should be prioritised for populations showing high genetic diversity and unique genetic structures. This method preserves populations in their natural environments, therefore enabling ongoing adaptation to local environmental conditions.

In situ conservation is especially important for the Petén and San Cristobal populations of *P. dioica*, which represent contrasting Holdridge life zones and harbour the greatest genetic variation of the populations studied. These should be prioritised as primary conservation units, with a minimum of 25–30 female reproductive trees per population actively protected from harvest pressure to maintain adequate effective population sizes.

Circa situ conservation is directly relevant for the cultivated populations of the FTN, where *P. dioica* farming is integrated into agroforestry systems (43). This approach supports local livelihoods while preserving genetic diversity in agricultural environments. Local communities should be encouraged to maintain at least 10–15 genotypically diverse female genotypes per farm or community plot, avoiding exclusive selection of morphologically uniform, short-statured phenotypes that risk narrowing the genetic base.

A vital complement to these efforts is ex situ conservation through seed banks and living collections. Seed collections should target a minimum of 15–20 maternal lines per priority population with precedence given to San Cristobal and Petén to capture the allelic diversity identified in this study and serve as a genetic safeguard against local population loss. The effective implementation of these integrated strategies requires cooperative efforts among local producers, researchers and conservation organisations, balancing the economic value of *P. dioica* with its long-term genetic sustainability in the face of evolving social and environmental conditions.

Conclusion

Random amplified polymorphic DNA markers proved a useful and efficient tool for the initial molecular characterisation of *Pimenta dioica* genetic resources in Guatemala, providing baseline data where none previously existed and confirming that natural populations maintain greater genetic diversity than cultivated ones. The genetic uniqueness of the San Cristobal population underscores the conservation value of remnant natural stands as a pool of original Franja Transversal del Norte (FTN) genotypes, highlighting the need for a balanced management approach that preserves genetic diversity across both natural and cultivated populations. These findings support prioritising San Cristobal and Petén as in situ conservation units, promoting genotypically diverse *in situ* management in FTN agroforestry systems and establishing ex situ seed collections from at least 15–20 maternal lines per priority population. Future research should employ codominant molecular markers (SSRs or SNPs) to obtain precise estimates of heterozygosity and inbreeding coefficients, include male trees to characterise pollen-mediated gene flow and monitor the San Cristobal population over the long term. Common garden experiments combining molecular and phenotypic data will clarify the adaptive significance of the genetic structure documented here, ultimately informing domestication and breeding programmes for *P. dioica* in Guatemala.

Acknowledgements

The authors sincerely thank the local farmers of Guatemala who kindly allowed access to their allspice farms and participated in interviews. Their contributions provided valuable perspectives on *Pimenta dioica* management practices and improved the understanding of the species. Financial support for this research was provided by the Dirección General de Investigación (DIGI), Universidad de San Carlos de Guatemala, through project B4CU-2021.

Authors' contributions

JARC conceived the study, led the conceptualisation and conducted the data analysis. Field sampling was carried out by JARC, JEBS, CEVG and SBBL. Laboratory work, including DNA extraction and random amplified polymorphic DNA amplification, was performed by SBBL and BGO. Preliminary writing of the manuscript was undertaken by JARC, EMGR and SBBL. All authors have read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interest to declare.

Ethical issues: None

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Claude (Anthropic) for language editing and manuscript formatting assistance. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

1. El-Nashar HAS, Ali AM, Salem YH. Genus *Pimenta*: an updated comprehensive review on botany, distribution, ethnopharmacology, phytochemistry and biological approaches. *Chem Biodivers*. 2023;20(12):e202300855. <https://doi.org/10.1002/cbdv.202300855>
2. Chinchilla-García JJ. Inventory and management proposal for populations of allspice (*Pimenta dioica* (L.) Merrill) and sapodilla (*Manilkara achras* (Mill.) Fosberg) in the special use zone of Yaxhá National Park, within the Maya Biosphere Reserve. [thesis]. Guatemala: Universidad de San Carlos de Guatemala; 2004.
3. Parker T. Trees of Guatemala. Austin (TX): The Tree Press; 2008.
4. Machuca P, Pulido-Salas MT, Trabanino F. Past and present of allspice (*Pimenta dioica*) in Mexico and Guatemala. *Rev Ethnocol*. 2020;18:1–18. <https://doi.org/10.4000/ethnocol.6261>
5. Fremout T, Thomas E, Taedoum H, Briers S, Gutiérrez Miranda CE, Alcázar Caicedo C, et al. Diversity for Restoration (D4R): guiding the selection of tree species and seed sources for climate resilient restoration of tropical forest landscapes. *J Appl Ecol*. 2022;59(3):664–79. <https://doi.org/10.1111/1365-2664.14079>
6. Bidyananda N, Jamir I, Nowakowska K, Varte V, Vendrame WA, Devi RS, et al. Plant genetic diversity studies: insights from DNA marker analyses. *Int J Plant Biol*. 2024;15(3):607–40. <https://doi.org/10.3390/ijpb15030046>
7. Goswami BO, Rankawat R, Gadi BR. Morphological and molecular markers based assessment of genetic diversity of a valuable endemic plant *Lasiurus sindicus* Henr. in the arid region of Rajasthan, India. *Arid Land Res Manag*. 2022;36(2):1–16. <https://doi.org/10.1080/15324982.2021.2012728>
8. Özbek Ö. Molecular markers used to reveal genetic diversity and phylogenetic relationships in crop plants. *OBM Genet*. 2024;8(4):274. <https://doi.org/10.21926/obm.genet.2404274>
9. Cruz W, Saldaña C, Ramos H, Baselly Villanueva JR, Cancán Loli J, Cuellar-Bautista JE. Genetic structure of natural populations of *Cedrelina cateniformis* "tornillo" from the oriental region of Peru. *Sci Agropecu*. 2020;11(4):521–8. <https://doi.org/10.17268/sci.agropecu.2020.04.07>
10. Tijerino A, Callejas L, Cerda-Granados DA. Assessment of genetic diversity in five Nicaraguan populations of *Cedrela odorata* L. (Meliaceae) using RAPD markers. *Encuentro*. 2016;103:28–39. <https://doi.org/10.5377/encuentro.v0i103.2690>
11. Hamdan YAS, Hawamda AI, Badran L, Daraghme E, Marí L, Khalaf A, et al. Exploring genetic variations among Palestinian watermelon (*Citrullus lanatus*) germplasm using RAPD molecular markers. *Proc Natl Acad Sci India Sect B Biol Sci*. 2024;94:829–34. <https://doi.org/10.1007/s40011-024-01612-5>
12. Hromadová Z, Gálová Z, Mikolášová L, Balázová Z, Vivodík M, Chnapek M. Efficiency of RAPD and SCoT markers in the genetic diversity assessment of the common bean. *Plants (Basel)*. 2023;12(15):2763. <https://doi.org/10.3390/plants12152763>

13. Niklas A, Olszewska D. Application of the RAPD technique to identify genetic diversity in cultivated forms of *Capsicum annuum* L. *Biotechnologia*. 2021;102(2):209–23. <https://doi.org/10.5114/bta.2021.106523>
14. Mathpal S, Joshi R, Bhatt B, Sahoo B, Joshi S. Randomly amplified polymorphic DNA (RAPD) molecular markers as a valuable tool to confirm the genetic diversity in soybean germplasms (*Glycine max*). *Indian J Agric Res*. 2024;58:1011–7. <https://doi.org/10.18805/IJARE-A-6176>
15. Sudhir S, Velmani Vel S, Selvam H, Arunprasath A. Using RAPD and ISSR markers to assess genetic diversity and conservation of *Bruguiera gymnorrhiza* (L.) Lam. for sustainable management on Kerala's West Coast. *J Appl Biol Biotechnol*. 2024;12(6):105–11. <https://doi.org/10.7324/JABB.2024.184346>
16. Holdridge LR. Life zone ecology. Rev ed. San José: Tropical Science Center; 1967.
17. Massicotte P, South A. r.naturalearth: world map data from Natural Earth. R package version 1.0.1. 2023.
18. Hollister JW. elevatr: access elevation data from various APIs. R package version 0.99.0. 2023.
19. Dunnington D. ggspatial: spatial data framework for ggplot2. R package version 1.1.9. 2023.
20. Goyal P, Jain R, Kachhwaha S, Kothari SL. Assessment of genetic diversity in *Pithecellobium dulce* (Roxb.) Benth. germplasm using RAPD and ISSR markers. *Trees (Berl)*. 2015;29(3):637–53. <https://doi.org/10.1007/s00468-014-1141-8>
21. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*. 1980;32(3):314–31.
22. Amiryousefi A, Hyvönen J, Poczar P. iMEC: online marker efficiency calculator. *Appl Plant Sci*. 2018;6(6):e1159. <https://doi.org/10.1002/aps.3.1159>
23. Meirmans PG. Genodive version 3.0: easy-to-use software for the analysis of genetic data of diploids and polyploids. *Mol Ecol Resour*. 2020;20(4):1126–31. <https://doi.org/10.1111/1755-0998.13145>
24. Lynch M, Milligan BG. Analysis of population genetic structure with RAPD markers. *Mol Ecol*. 1994;3(2):91–9. <https://doi.org/10.1111/j.1365-294X.1994.tb00109.x>
25. Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. 2019;47(W1):W256–9. <https://doi.org/10.1093/nar/gkz2239>
26. Frichot E, Mathieu F, Trouillon T, Bouchard G, Francois O. Fast and efficient estimation of individual ancestry coefficients. *Genetics*. 2014;196(4):973–83. <https://doi.org/10.1534/genetics.113.160572>
27. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155(2):945–59. <https://doi.org/10.1093/genetics/155.2.945>
28. Ruiz-Chután JA, Kalousová M, Lojka B, Colacho-Hernández S, Prado-Córdova JP, Montes L, et al. Impacts of habitat fragmentation on the genetic diversity of the endangered Guatemalan fir (*Abies guatemalensis* Rehder). *Genetica*. 2024;153(1). <https://doi.org/10.1007/s10709-024-00225-0>
29. Grünwald NJ, Hoheisel GA. Hierarchical analysis of diversity, selfing and genetic differentiation in populations of the oomycete *Aphanomyces euteiches*. *Phytopathology*. 2006;96(10):1134–41. <https://doi.org/10.1094/PHYTO-96-1134>
30. Bougeard S, Dray S. Supervised multiblock analysis in R with the ade4 package. *J Stat Softw*. 2018;86(1):1–7. <https://doi.org/10.18637/jss.v086.i01>
31. Boshier DH, Chase MR, Bawa KS. Population genetics of *Cordia alliodora* (Boraginaceae), a neotropical tree. 3. Gene flow, neighborhood and population substructure. *Am J Bot*. 1995;82(4):484–90. <https://doi.org/10.1002/j.1537-2197.1995.tb15668.x>
32. Khan S, Vaishali, Sharma V. Genetic differentiation and diversity analysis of medicinal tree *Syzygium cumini* (Myrtaceae) from ecologically different regions of India. *Physiol Mol Biol Plants*. 2010;16(2):149–58. <https://doi.org/10.1007/s12298-010-0016-8>
33. Coser SM, Ferreira MFS, Ferreira A, MitreLK, CarvalhoCR, Clarindo WR. Assessment of genetic diversity in *Psidium guajava* L. using different approaches. *Sci Hortic*. 2012;148:223–9. <https://doi.org/10.1016/j.jsienta.2012.09.030>
34. Sinha D, Singh J, Tandon PK, Kakkar P. Genetic diversity of *Pinus roxburghii* Sarg. collected from different Himalayan regions of India assessed by random amplified polymorphic DNA analysis. *Toxicol Int*. 2013;20(3):208–13. <https://doi.org/10.4103/0971-6580.121667>
35. Saldaña CL, Cancan JD, Cruz W, Correa MY, Ramos M, CuellarE, et al. Genetic diversity and population structure of capirona (*Calycophyllum spruceanum* Benth.) from the Peruvian Amazon revealed by RAPD markers. *Forests*. 2021;12(8):1125. <https://doi.org/10.3390/f12081125>
36. Bonin A, Ehrich D, Manel S. Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Mol Ecol*. 2007;16(18):3737–58. <https://doi.org/10.1111/j.1365-294X.2007.03435.x>
37. Nybom H. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol Ecol*. 2004;13(5):1143–55. <https://doi.org/10.1111/j.1365-294X.2004.02141.x>
38. Salgotra RK, Chauhan BS. Genetic diversity, conservation and utilization of plant genetic resources. *Genes (Basel)*. 2023;14(1):174. <https://doi.org/10.3390/genes14010174>
39. Tuisima-Coral LL, Cepková PH, Weber JC, Lojka B. Preliminary evidence for domestication effects on the genetic diversity of *Guazuma crinita* in the Peruvian Amazon. *Forests*. 2020;11(8):795. <https://doi.org/10.3390/f11080795>
40. Li YR, Fritsch PW, Zhao GG, Cheng XJ, Ding ZL, Lu L. Population differentiation and dynamics of five pioneer species of *Gaultheria* from the secondary forests in subtropical China. *BMC Plant Biol*. 2024;24(1):516. <https://doi.org/10.1186/s12870-024-05189-z>
41. Carvalho C, Ballesteros-Mejia L, Ribeiro MC, Correa Cortez M, Souza A, Garcia Collevati R. Climatic stability and contemporary human impacts affect the genetic diversity and conservation status of a tropical palm in the Atlantic Forest of Brazil. *Conserv Genet*. 2017;18(2):467–78. <https://doi.org/10.1007/s10592-016-0921-7>
42. Martin SL, Benedict L, Wei W, Sauder CA, Beckie HJ, Hall LM. High gene flow maintains genetic diversity following selection for high EPSPS copy number in the weed *kochia* (Amaranthaceae). *Sci Rep*. 2020;10:18864. <https://doi.org/10.1038/s41598-020-75345-6>
43. Dawson IK, Guariguata MR, Loo J, Weber JC, LengkeerA, BushD, et al. What is the relevance of smallholders' agroforestry systems for conserving tropical tree species and genetic diversity in circa situ, in situ and ex situ settings? A review. *Biodivers Conserv*. 2013;22(2):301–24. <https://doi.org/10.1007/s10531-012-0429-5>

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonpublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc. See https://horizonpublishing.com/journals/index.php/PST/indexing_abstracting

Copyright: © The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited (<https://creativecommons.org/licenses/by/4.0/>)

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.