



# Wild papaya shows evidence of gene flow from domesticated Maradol papaya in Mexico

Pedro Jesús Ruiz-Gil · Ana Wegier ·  
Valeria Alavez · Santiago Rosas-Plaza ·  
Juan Núñez-Farfán · Mariana Chávez-Pesqueira 

Received: 13 December 2022 / Accepted: 13 March 2023 / Published online: 27 March 2023  
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

**Abstract** Multiple factors threaten the conservation of crop wild relatives (CWR), such as habitat transformation and fragmentation, rapid environmental changes, introduction of invasive species, urbanization, and introduction of genetically modified organisms. Moreover, in regions where CWR coexist with their domesticated counterparts, reductions of genetic

diversity and local extinction of wild populations due to homogenization in allelic frequencies, are expected. Nowadays, papaya plantations in Mexico extend over large areas including the natural distribution of its wild relative, potentially favoring genetic exchange between wild and domesticated plants. In this study, we evaluated the potential distribution of wild papaya and analyzed the genetic diversity and structure, using microsatellites and cpDNA markers, to identify the evolutionary units and possible events of gene flow within the wild-to-domesticated complex of papaya, including conventional and transgenic domesticated plants in Mexico. The potential distribution showed a good climatic suitability for the tropical regions of the Pacific coast, the Gulf of Mexico and the Yucatan Peninsula. We identified large evolutionary units for wild plants and evidence of gene flow from papaya plantations to wild units. Wild papaya units in the vicinity of papaya plantations showed reduced genetic diversity, owing to the highly reduced genetic diversity in domesticated plants. No evidence was found of the presence of transgenes in wild populations and plantations of papaya throughout Mexico. Our findings raise concerns about the maintenance of CWR gene pools and the need for improved management and conservation strategies for *Carica papaya* in its center of origin, domestication and diversification.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10722-023-01568-x>.

P. J. Ruiz-Gil · M. Chávez-Pesqueira (✉)  
Unidad de Recursos Naturales, Centro de Investigación Científica de Yucatán, 97205 Mérida, Mexico  
e-mail: mariana.chavez@cicy.mx

P. J. Ruiz-Gil  
Posgrado en Ciencias Biológicas, Unidad de Recursos Naturales, Centro de Investigación Científica de Yucatán, 97205 Mérida, Mexico

A. Wegier · V. Alavez · S. Rosas-Plaza  
Laboratorio de Genética de la Conservación, Jardín Botánico, Instituto de Biología, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico

V. Alavez · S. Rosas-Plaza  
Posgrado en Ciencias Biológicas, Instituto de Biología, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico

J. Núñez-Farfán  
Laboratorio de Genética Ecológica y Evolución, Instituto de Ecología, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico

**Keywords** Crop wild relatives · Genetic diversity · Population structure · Conservation genetics · Transgenic monitoring

## Introduction

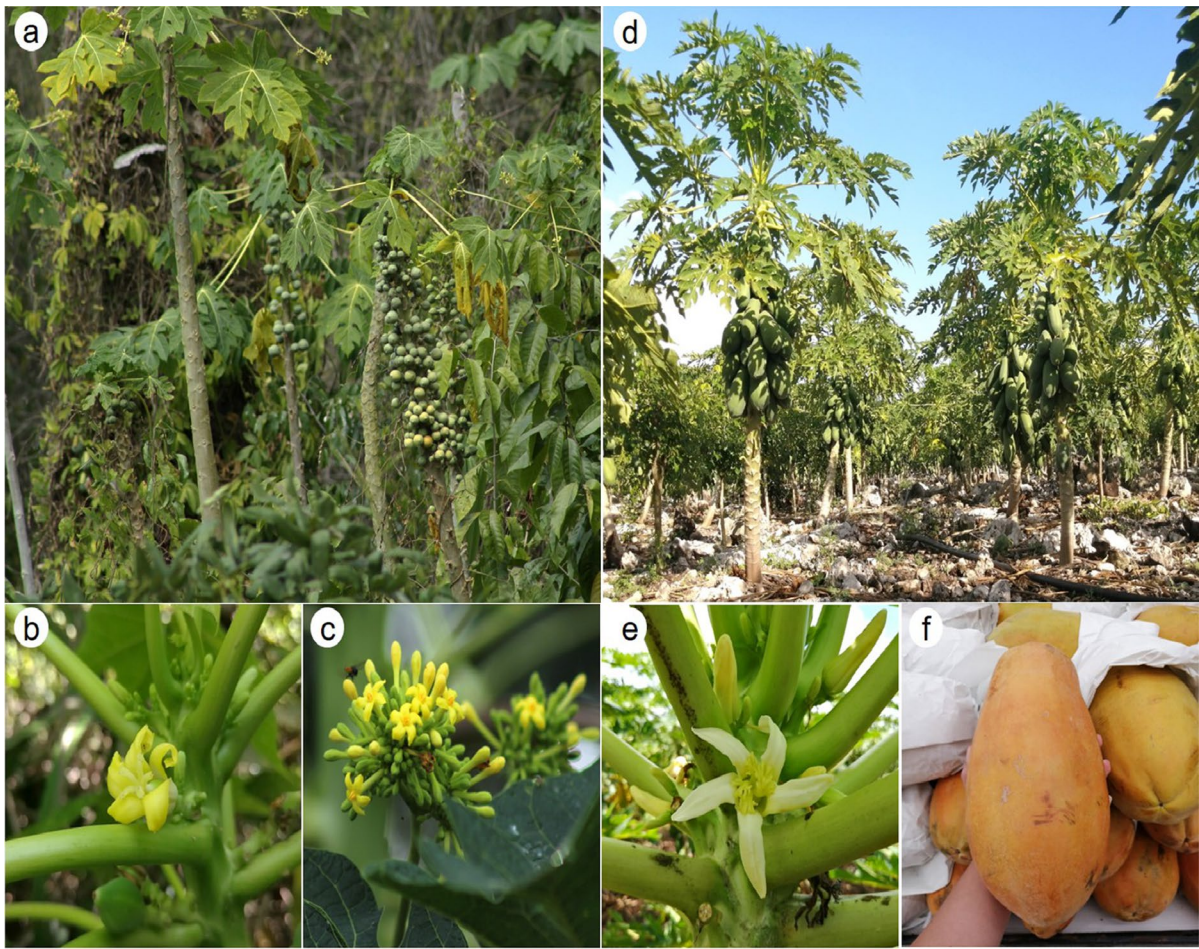
The future of food security depends on in situ conservation of agrobiodiversity and maintaining the evolutionary potential of domesticated species (Thomas et al. 2016). Such potential depends largely on the conservation of crop wild relatives (CWR); however, multiple factors threaten their conservation and genetic diversity, mainly driven by habitat transformation and fragmentation, rapid environmental changes, introduction of invasive species, urbanization, introduction of genetically modified organisms, and widespread use of biological resources, such as logging and overharvesting (Goettsch et al. 2021). In addition, the proximity between wild plants and plantations of their domesticated forms can also cause a loss of genetic diversity in areas where they coexist (Slatkin 1987; Ellstrand et al. 1999). In this study, we analyzed the genetic diversity and structure, to identify the wild evolutionary units and possible gene flow within the wild-domesticated complex of papaya in Mexico, part of the center of origin, domestication and diversification of the species.

In general, CWR and landraces exhibit high levels of genetic variation in their centers of origin (Gepts and Papa 2003). Domesticated plants and their CWR are the result of distinct evolutionary histories that, in many cases, have increased intraspecific diversity without losing the ability to interbreed (Brozynska et al. 2016; Dempewolf et al. 2017). In many crop species, the occurrence of gene flow that enables natural interbreeding and the generation of fertile offspring is possible (Harlan and de Wet 1971) if domesticated and wild plants are sexually compatible, grow in sympatry, share pollinators (or the same pollination syndromes) and if flowering coincides totally or partially (Papa and Gepts 2003; Papa 2005). Increasing crop plant density, however, changes the scenario, and consequences may include further reduction in genetic diversity, local extinction, and the development of aggressive weedy varieties (Bartsch et al. 1999; Ellstrand et al. 1999; Mardonovich et al. 2019). Modern varieties of many crops have been selected under international standards of genetic homogeneity

and phenotypic stability, therefore, their diversity is even lower, so a higher loss of genetic diversity for wild populations is expected if interbreeding occurs (Ehrlich and Raven 1969; Bartsch et al. 1999; Ellstrand et al. 1999; Papa and Gepts 2003). Moreover, establishing transgenic crops or introducing imported seeds in regions where the CWR occurs, raises the potential genetic and environmental risks of hybridization (Hails 2000; Snow 2002).

Wild populations of papaya (*Carica papaya*) are composed of fast-growing trees, with a simple stem and indeterminate growth (Teixeira Da Silva et al. 2007) that reach 10 m in height and can live up to 20 years (Jiménez et al. 2014) (Fig. 1a). Wild papayas bear either male or female flowers (i.e., dioecious; Fig. 1b, c). Naturally distributed in lowland tropical forests of Mexico and northern Central America, papaya is a pioneer species that colonizes newly created large clearings (Núñez-Farfán and Dirzo 1988; Paz and Vázquez-Yanes 1998; Aradhya et al. 1999; Carvalho and Renner 2012; Chávez-Pesqueira and Núñez-Farfán 2017). Wild papayas are within the pollen/seed dispersal range of papaya plantations in many parts of Mexico (Chávez-Pesqueira and Núñez-Farfán 2017). Due to agricultural practices, domesticated plants complete more generations in fewer years and show higher plant abundance, which translate into a higher availability of pollen and seeds (Papa and Gepts 2003; Papa 2005). There is evidence that wild and domesticated papaya plants share pollinators and floral visitors, such as moths, bees, and ants (Moo-Aldana et al. 2017; Badillo-Montañón et al. 2018; Pacheco-Huh et al. 2021), facilitating interbreeding between these two groups with different evolutionary histories (Badillo-Montañón et al. 2019; Pacheco-Huh et al. 2021). Pollen and seeds of domesticated plants can reach wild populations from different sources such as plantations, feral plants, and even from individuals with intermediate phenotypes, such as those that have been identified as natural populations in Costa Rica (Brown et al. 2012), as there is evidence that they facilitate gene flow from domesticated to wild papaya (Wu et al. 2017). Flowering and fruiting of wild plants occur throughout the year. Fruits are consumed and probably dispersed by birds and small mammals (Chávez-Pesqueira and Núñez-Farfán 2016).

Characteristics associated with the papaya domestication syndrome are fruit gigantism (Carvalho and



**Fig. 1** **a** Wild population of *Carica papaya* showing female individuals with immature fruits, at Yucatan, Mexico. **b** Female flower of wild *C. papaya*. **c** Male inflorescence of wild

*C. papaya*. **d** Plantation of *C. papaya* showing immature fruits of the Maradol variety, at Yucatan, Mexico. **e** Hermaphrodite flower of a Maradol papaya. **f** Mature fruit of Maradol papaya

Renner 2012; Fig. 1d, f), loss of dormancy (Paz and Vázquez-Yanes 1998), reduced defense strategies against antagonists (Pacheco-Huh et al. 2021), and changes in floral morphology and reproductive system, such as a tendency to hermaphroditism in domesticated varieties (Fig. 1e) (Chávez-Pesqueira and Núñez-Farfán 2017), which in turn can lead to changes in sex ratio, thus altering the effective population size (Chávez-Pesqueira et al. 2014). The genetic diversity of domesticated papaya is considerably lower compared to wild populations (Chávez-Pesqueira and Núñez-Farfán 2017). Mexico is the world's leader exporter of papaya (FAO 2022). The Maradol papaya is the predominant improved variety grown in Mexico (Soriano-Melgar et al. 2016) and

has displaced local varieties since its introduction ca. 1990s, currently representing 95% of the country's production (Vázquez et al. 2010; SIAP 2018). However, despite the fact that local varieties have lost their large-scale commercial importance, some are still grown in rural communities at low densities in home gardens and milpas, where they are associated with diverse uses (Moo-Aldana et al. 2017).

In the case of wild papaya plants, much of its natural distribution in Mexico is fragmented due to changes in land use, thus producing a decreased intrapopulation genetic diversity in populations of forest fragments, and an increased interpopulation genetic differentiation when compared to populations in the primeval forest (Chávez-Pesqueira et al. 2014) In



addition, being strictly dioecious, isolated populations of wild papaya experience biased sex ratios and high levels of biparental inbreeding (Brown et al. 2012; Chávez-Pesqueira et al. 2014). Regions where wild populations coexist with their domesticated relatives, such as Mexico, are of particular interest for conservation matters. Here, we aim to determine the distribution of wild papaya, to assess the genetic diversity and structure of wild papaya populations to identify the species' evolutionary units, and to analyze the possible effect of gene and transgene flow between domesticated and wild papaya plants in Mexico.

## Materials and methods

### Potential distribution map

#### *Occurrence data*

We collected occurrence records for wild *C. papaya* across their geographic range in Mexico. We examined, gathered, and cleaned the data from the following sources: (1) through direct observations during several surveys across Mexico (Chávez-Pesqueira et al. 2014; Chávez-Pesqueira and Núñez-Farfán 2016; unpublished field observations); and (2) by carefully reviewing specimens from seven herbariums (six from Mexico and one from the United States): Herbario-Fibroteca U Najil Tikin Xiw of Centro de Investigación Científica de Yucatán, A.C. (CICY), Herbario Nacional of the Universidad Nacional Autónoma de México (MEXU), Herbario Alfredo Barrera Marín of Universidad Autónoma de Yucatán (UADY), New York Botanical Garden (NYBG), Herbario of Escuela Nacional de Ciencias Biológicas (ENCB), Instituto de Ecología, A.C. (XAL), and Estación de Biología Los Tuxtlas, UNAM (Universidad Autónoma de México), distinguishing specimens with wild characteristics (small fruits with scarce pulp). From the 816 papaya specimens we examined in herbariums, only 69 were validated as wild plants. From these, only 26 had precise and reliable geographic information and were used, along with locations of wild populations collected in this and previous studies (Chávez-Pesqueira et al. 2014; Chávez-Pesqueira and Núñez-Farfán 2016), for a total of 106 records (Table S1). After cleaning the geographic inconsistencies, we thinned the

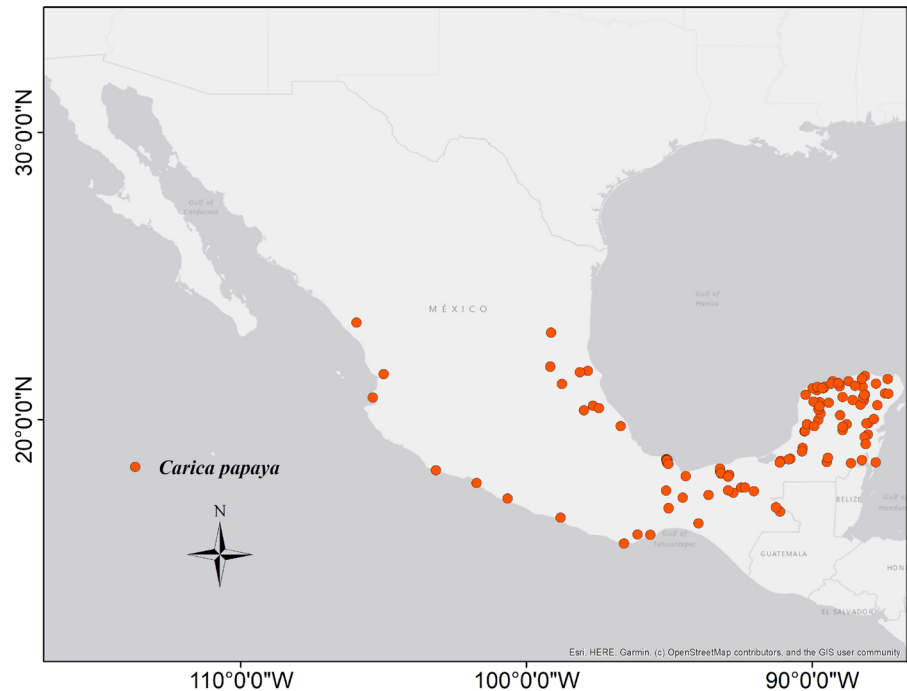
dataset following a distance-based approach (excluding duplicated records within a grid of  $1 \times 1$  km) with the NTBOX package version 0.4.6.1 (Osorio-Olvera et al. 2020) in R version 3.5.0 (R Core Team 2021). Finally, for modeling and evaluation purposes, we split the thinned database into two sets (i.e., training and testing) of nearly identical sizes following the checkerboard partition method implemented in the package ENMeval (version 0.3.0; Muscarella et al. 2018). We show the localities used in this study in Fig. 2 ( $n = 106$ ).

#### *Climatic variables*

The environmental layers used in this study correspond to bioclimatic variables summarizing annual, seasonal, and extreme tendencies derived from monthly temperature and precipitation data for Mexico between 1910 and 2009 at 30" resolution ( $\sim 1$  km) (Cuervo-Robayo et al. 2014). We excluded four variables (i.e., mean temperature of the most humid quarter, mean temperature of the least humid quarter, precipitation of the warmest quarter, and precipitation of the coldest quarter) due to artifacts resulting from the combination of temperature and precipitation data (Escobar et al. 2014). To avoid model overfitting due to multicollinearity among variables, we selected a sub-set of uncorrelated and biologically meaningful variables. For this, we first assessed variables' contribution in exploratory runs of a Maxent model (Phillips et al. 2006) (Table S2) by measuring the variable contribution percentage, permutation importance, and model gain through jackknife tests. Then, we calculated pairwise Pearson's correlation coefficients from the occurrences. From these analyses, we excluded highly correlated variables ( $|r| > 0.8$ ) and retained important variables with the greatest biological relevance for further modeling (Wegier 2013; Simoes et al. 2020). Thus, we used five variables: annual mean temperature (Bio1), temperature seasonality (Bio4), minimum temperature of the coldest month (Bio6), temperature annual range (Bio7), and precipitation of the driest month (Bio14).

The calibration area, or the 'M' element of the BAM diagram, refers to areas that have been accessible to the species via dispersal over relevant time periods (Barve et al. 2011; Peterson and Soberón 2012). We considered the geographic extent of Mexico as the calibration area.

**Fig. 2** Distribution of 106 records (orange dots) of wild *Carica papaya* in Mexico. (Color figure online)



### Ecological niche modeling

We modeled the ecological niche of wild *C. papaya* using the Maxent algorithm version 3.3.3 k (Phillips et al. 2006). Maxent is a machine-learning algorithm that uses the maximum entropy principle to identify a target probability distribution subject to a set of constraints related to the environmental characteristics of occurrences and a sample of the calibration area (Phillips et al. 2004, 2006). To select an optimal combination of Maxent parameters (regularization multipliers [RM] and feature classes [FC], see below), we developed a series of candidate models with the R package ENMeval (Muscarella et al. 2018); thus, we assessed five FC combinations: L, LQ, LQH, LQHP, and LQHPT (where, L: linear; Q: quadratic; H: hinge; P: product, and T: threshold) and tested RM values ranging from 0.5 to 4 in increments of 0.5. We selected the best combination of model parameters according to performance and complexity criteria, namely: area under the ROC curve (AUC) > 0.9, omission rates < 0.10, and the Akaike information criterion (AICc) where  $\Delta AIC \leq 2$ . Then, a model with specific RM and FC settings was run in Maxent (i.e., tenfold cross-validation models, 20,000 background points, and 1000 maximum iterations). We

selected the average Cloglog output for environmental continuous suitability visualization and reclassified it into a binary map (i.e., presence or absence of suitable conditions) in ArcGIS version 10.2.1 (ESRI 2015) by applying a threshold value balancing a low omission error and the proportional predicted area. Finally, to assess model performance and significance, we evaluated the AUC ratio of the partial receiver operating characteristic curve (pROC; 1,000 replicates, and  $E=0.05$ ), calculated the omission rate, and performed binomial tests with the testing datasets obtained previously (see the occurrence data section above) in NTBOX (Osorio-Olvera et al. 2020).

### Sampling for genetic analyses

Samples from 556 papaya individuals (447 wild and 109 domesticated) were obtained from 33 collection sites. For wild papaya, samples were obtained from 24 collection sites throughout the natural distribution of the species in Mexico. Samples for domesticated papaya were obtained from nine plantations of Maradol papaya from two different regions of Mexico (four plantations of the Pacific coast from the states of Colima, Jalisco and Guerrero, and five plantations from the state of Yucatan) (Fig. 4, Table S3).

Genetic data were obtained from six microsatellite primers (Ocampo et al. 2006, Table S4) for all 556 individuals, and one cpDNA region, *psbA-trnH*, for 411 individuals (368 wild and 42 domesticated). For wild individuals, genetic data were obtained from field collections in the western coast of Mexico, and integrated with previous data from southeastern Mexico (Chávez-Pesqueira and Núñez-Farfán 2016) (Table S3) to represent the entire natural distribution of papaya in Mexico. DNA was extracted from leaf tissue using the CTAB protocol (Doyle and Doyle 1987) and the amplification of both markers (microsatellites and the *psbA-trnH* region) was following Chávez-Pesqueira and Núñez-Farfán (2016).

#### Data analyses for nuclear microsatellites

The genetic structure of wild individuals was inferred with two approximations, using Geneland v.4.0.9 (Guillot et al. 2008) and STRUCTURE v. 2.3.4 (Pritchard et al. 2000). With Geneland, we carried out ten independent runs, each run consisted of 2,000,000 Markov chain Monte Carlo (MCMC) iterations with a thinning value of 100, and a coordinate uncertainty of 0.5 (which takes into account the presence of individuals with identical coordinates) (Salamon et al. 2020). Uncorrelated and null allele model options were chosen. Runs were recalculated with a burn-in of 1,000 iterations, and the average posterior probability was used to choose the best-suited run. For post-processing of the output files, the spatial domain was defined as  $X=300$  and  $Y=300$  with a burn-in of 1,000. The results of three individual runs were compared to assess spatial group correspondence. With STRUCTURE, ten iterations were run under an admixture model assuming uncorrelated frequencies with a range of K values from one to eight (structure proposed by Geneland). Each run had a burn in period of 500,000 followed by 1,000,000 Markov chain replicates (MCMC). To obtain the optimum K value, the method proposed by Evanno et al. (2005) was used.

We evaluated pairwise genetic differentiation of populations by assuming a step-wise mutational model with  $R_{ST}$  statistic (Slatkin 1995), then partitioning of genetic variability within and among populations groups was tested by analysis of molecular variance (AMOVA; Excoffier et al. 1992) among the genetic cluster derived from Geneland using GeneAlex v.6.5 (Peakall and Smouse 2012).

Standard genetic diversity indices as the percentage of polymorphic loci (%P), number of alleles (Na), observed heterozygosity ( $H_O$ ), and expected heterozygosity ( $H_E$ ), were estimated for each wild genetic cluster derived from Geneland and for the 109 domesticated individuals (which formed a single genetic group) using the Arlequin 3.5 software (Excoffier and Lischer 2010).

#### Gene flow estimators

Gene flow between domesticated and wild populations was evaluated using two methods. (1)  $R_{ST}$  was used to indirectly calculate the number of migrants per generation ( $Nm$ ) between pairs of populations:  $Nm = [(1/R_{ST}) - 1]/4$ . (2) BayesAss v.3.0.4 (Wilson and Rannala 2003) was used to obtain recent migration rates ( $m$ ). This software uses genetic assignments to estimate short-term dispersal rates (over the past two generations). We performed five runs (each with different starting seed value) of 10,000,000 generations, with a burn-in of 1,000,000 generations, and sampled the chain every 2,000 generations.

#### Data analyses of cpDNA

368 cpDNA sequences were used to assess the genetic structure of wild individuals using Geneland v.4.0.9 (Guillot et al. 2008), considering the polymorphic sites detected with DnaSP v.6.12.03 (Rozas et al. 2017). To make the structure consistent between different runs, we removed from the analysis those individuals that had identical sequences to the domesticated group, resulting in 296 individuals. Options were as described for the microsatellite markers but using 4,000,000 Markov chain Monte Carlo (MCMC) iterations.

We used Arlequin v.3.5 (Excoffier and Lischer 2010) to obtain the number of polymorphic sites and number of haplotypes, and haplotype and nucleotide diversity for 368 wild and 42 domesticated papaya individuals.

We evaluated pairwise genetic differentiation of populations using the  $N_{ST}$  estimator.  $N_{ST}$  is analogous to  $F_{ST}$ , but considers the average number of differences between pairs of sequences, either of the same or different subpopulations (Lynch and Crease 1990). We also calculated the number of migrants per generation ( $Nm$ ) between pairs of populations. Estimators

of genetic differentiation and *Nm* were obtained using DNAsp v.6.12.03 (Rozas et al. 2017).

Haplotype networks were constructed to examine genealogical relationships between haplotypes using statistical parsimony with TCS 1.2.1 (Clement et al. 2000). Gaps in the sequences generated by insertions and deletions were considered as a fifth state. For this analysis, the domesticated populations of papaya were not considered.

### Transgene monitoring

For the transgene monitoring, 935 individuals of papaya were used (Table S5). For this analysis we included samples from the wild-domesticated complex of papaya, including local varieties from home gardens to cover the broadest representation of the species in Mexico. From the 935 individuals, 320 belonged to wild individuals obtained from 22 collection sites of wild papaya used in the genetic analyses described above (it was not possible to use the 24 collection sites, given that individuals from two sites did not amplify the primers), and from six new collections sites from the states of Tabasco and Guerrero, for a total of 320 wild individuals. We used 615 domesticated samples obtained from different sources (plantations, markets and commercial seeds): 240 obtained from plantations (mostly from the Maradol variety, but one from the Tainung variety), from markets (Maradol variety), and from commercial seeds (from the Amarilla, Criolla and Maradol varieties), and 135 belonged to samples collected in home gardens (mainly Maradol variety, but some local varieties like Amarilla, Amameyada and Criolla) (Table S5). DNA extraction was following the CTAB protocol with some modifications (Doyle and Doyle 1987). Three primers obtained from JRC GMO MATRIX (European Commission 2014) were used to detect possible transgenic events in Mexico (Table S6). Primers P35s and Tnos, allow the detection of all transgenic papaya lines registered in the European Commission database, and the first PRSV-CP, identifies the protein coat of Papaya Ringspot Virus (PRSV). For the detection of possible transgenic events, PCR reactions were performed for 187 pools of DNA containing 1 µl of DNA (10 ng/µl) from five individuals. In total, 64 pools from wild individuals, 48 from plantation individuals (domesticated), 27 pools from individuals from home gardens

(domesticated), and 48 from individuals from commercially sold seeds (domesticated) were analyzed. PCR reactions were composed of: 1 × buffer, 1.5 mM MgCl<sub>2</sub>, 0.6 mM dNTPs, 1 µM primer (F and R), 1U taq polymerase (Promega Gotaq® Flexi DNA) and 5 µl of DNA. The reaction conditions for primers P35S and Tnos were: initial denaturation at 94 °C for 3 min, 35 cycles of 30 s at 94 °C, 40 s at 55 °C–40 s and 40 s at 72 °C, and a final extension of 72 °C for 10 min. The reaction conditions for the PRSV-CP and CHY primers were: initial denaturation at 94 °C for 10 min, 35 cycles of 30 s at 94 °C, 30 s at 60 °C, 30 s at 72 °C, and a final extension of 72 °C for 3 min.

## Results

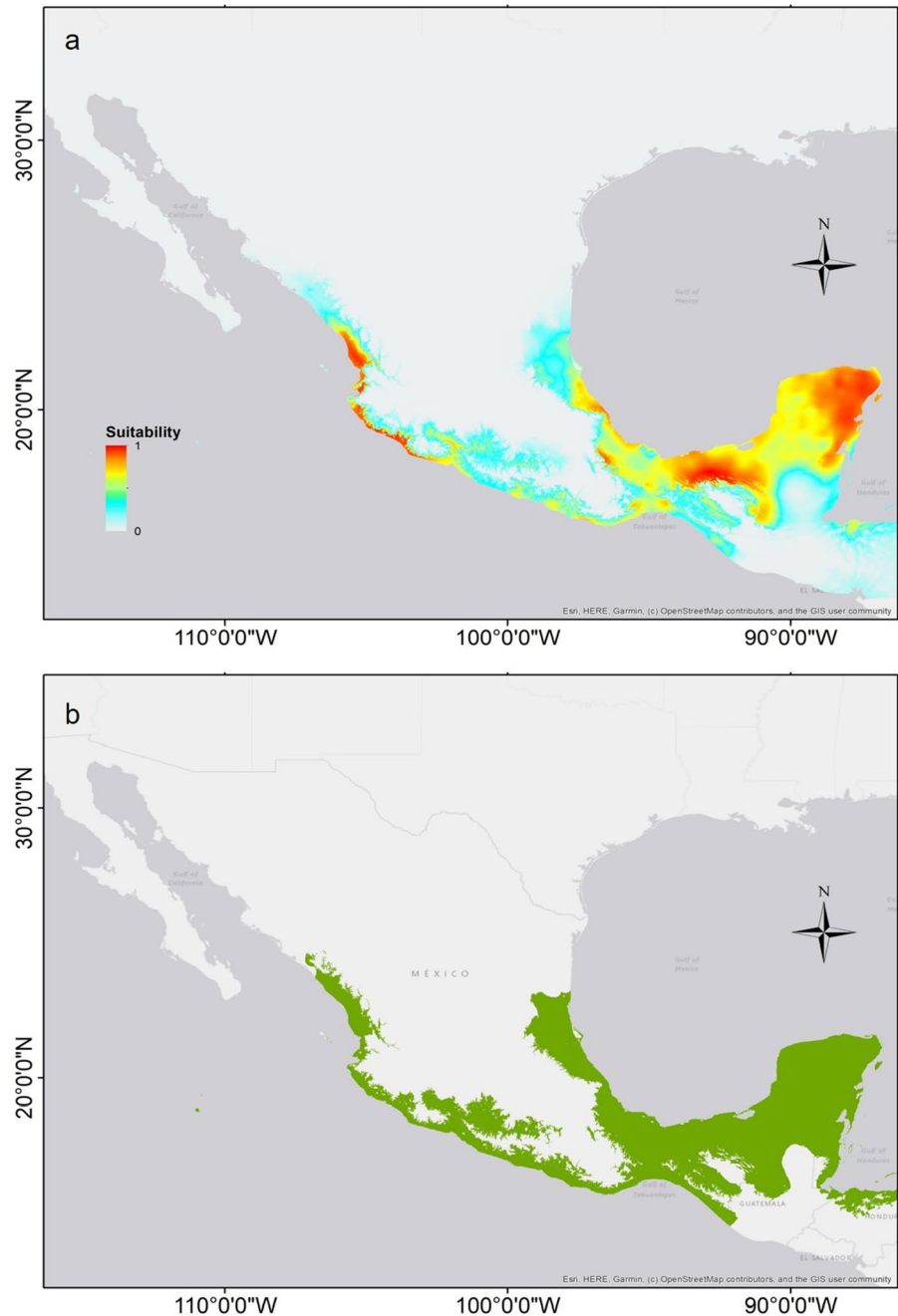
### Potential distribution map

We generated 40 candidate models and selected the parameter combinations that obtained the best performance and complexity estimates (i.e., FC=LQ, RM=0.5; Supplementary Materials 2). With these settings, the resulting niche model was statistically significant ( $p < 0.0001^{***}$ ) according to both significance evaluations (binomial test and partial ROC analysis; AUC ratio=1.79). The model exhibited good predictive performance (omission error=0.018) and indicates climatic suitability in the coasts of the Pacific, the Gulf of Mexico, and the Yucatán Peninsula (Fig. 3). The most important variables were annual mean temperature (Bio1) and minimum temperature of the coldest month (Bio6), as the former decreases the model gain the most when it is omitted and the latter results in the highest gain when used in isolation.

### Population structure and nuclear microsatellite diversity

Six genetic clusters (hereafter named evolutionary units) were detected by posterior cluster membership for the wild plants (Geneland  $K=6$ , Fig. 4a): North-Gulf unit included the following collection sites: Cielo, Huasteca, Tamazunchale; Isthmus-Gulf unit included: Matias Romero, Acayucan, Los Tuxtlas, Palenque and Villa Guadalupe; South-Pacific unit included: Poza Rica, Ventanilla, Santiago Astata and Marquelia; Yucatan Peninsula unit included:

**Fig. 3** **a** Climatic suitability map of wild *Carica papaya* in Mexico. The red color shows the areas with the highest climatic suitability for the species, while the light blue shows the areas with low climatic suitability; white color shows the areas with zero climatic suitability. **b** Binary map of the presence (green)/absence (white) of wild *C. papaya* in Mexico. (Color figure online)

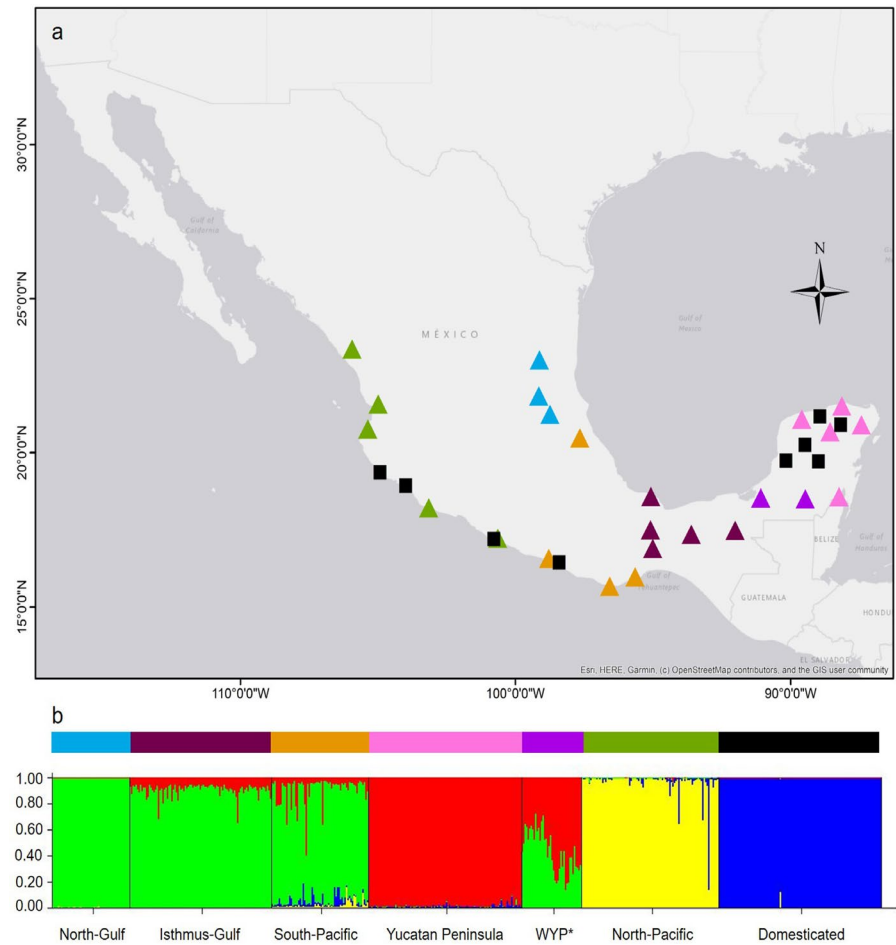


Oxtankah, Chichen Itzá, Dzibilchaltún, Río Lagartos and Cancún; Western Yucatan Peninsula unit included: Mamantel and Caoba; and North-Pacific unit included: Coast of Michoacán, Tecpan, Copala, Tepic and Bucerías. With STRUCTURE, a lower number of genetics clusters was recognized ( $K=4$ , Fig. 4b). All domesticated papaya individuals were

grouped as a single group (hereafter named domesticated group). For the six evolutionary units inferred by Geneland we obtained an overall  $R_{ST}$  value of 0.057 ( $p=0.001$ ), ranging from 0.006 to 0.092 (Fig. 5a). When including the domesticated group, the  $R_{ST}$  had a value of 0.103 ( $p=0.001$ ), and between pairs of wild populations and the domesticated group



**Fig. 4** The six and four evolutionary units inferred from Geneland (a) and STRUCTURE (b) using nuclear microsatellites from wild *Carica papaya* in Mexico. In (a), blue triangles: North-Gulf unit; purple triangles: Isthmus-Gulf unit; orange triangles: South-Pacific unit; pink triangles: Yucatan Peninsula unit; lilac triangles: Western Yucatan Peninsula unit; green triangles: North-Pacific unit. Black squares correspond to the cultivated. In (b), STRUCTURE plot showing the four evolutionary units ( $\Delta K=4$ ) for wild and domesticated *C. papaya*. The colored bar at the top corresponds to the evolutionary units found in Geneland. (Color figure online)

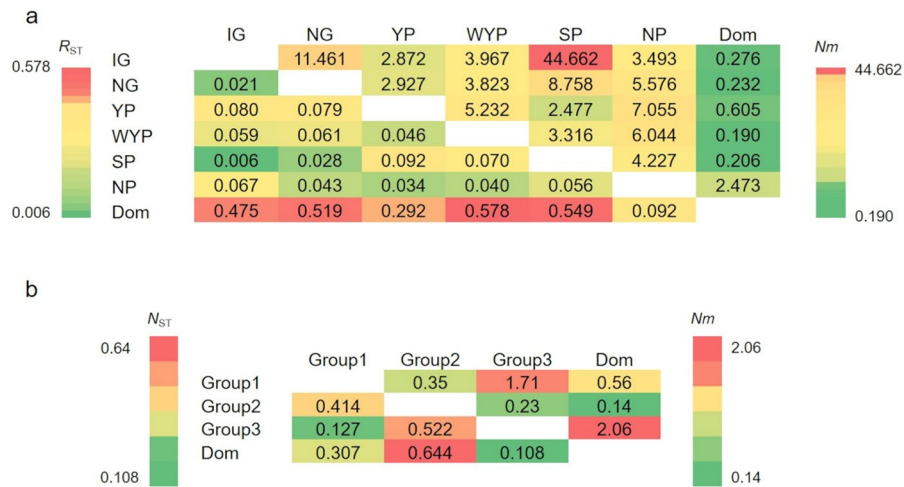


$R_{ST}$  values ranged from 0.006 to 0.568. Hierarchical partition of molecular variance (AMOVA) for the six wild units revealed that the highest proportion of variance was among individuals (72%) and the lowest proportions were within individuals (23%) or among units (6%).

The number of migrants per generation ( $Nm$ ) among pairs of wild units was in all cases above 1 (Fig. 5a). However, this estimator was lower when comparing between the wild populations and the domesticated population ( $Nm < 1$ ), except for the North-Pacific unit ( $Nm = 2.473$ ). Recent migration rates among units ( $m$ ), as estimated with BayesAss, showed low recent migration rates (Fig. 6), with higher rates among geographically close evolutionary units (Yucatan Peninsula—Western Yucatan Peninsula), and lower rates between geographically distant units (North-Pacific with any evolutionary units). However, this estimator showed low levels

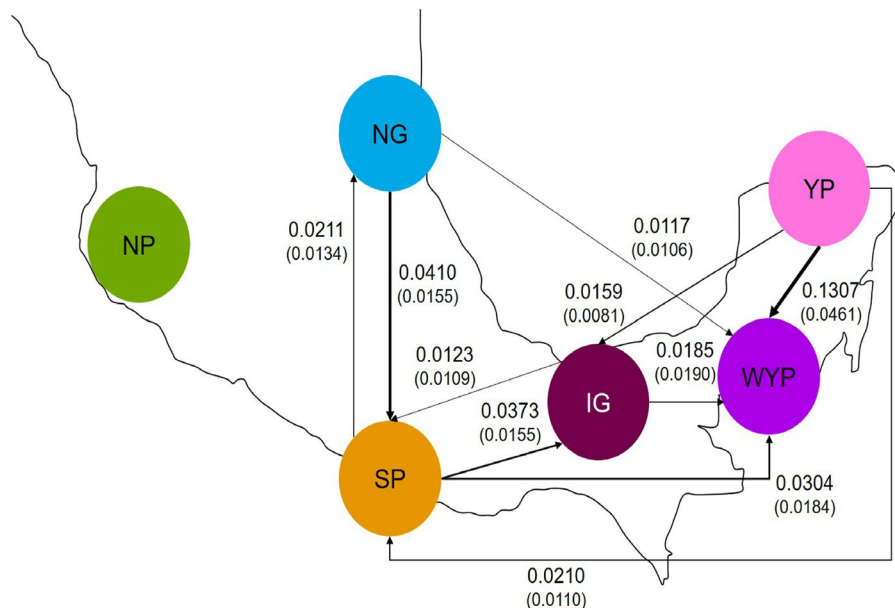
between the domesticated group and the wild evolutionary units ( $m < 0.01$ , values =  $0.0033 \pm 0.0033$  to  $0.0084 \pm 0.0056$ ), in contrast to the wild units which presented higher recent migration rates ( $m$  values =  $0.0032 \pm 0.0031$  to  $0.1307 \pm 0.0461$ ). However, the direction of gene flow was higher from the domesticated group (source population) to the wild units (recipient population) (Fig. 4b).

All loci were polymorphic and moderately high values of neutral genetic diversity were obtained for wild papaya in Mexico ( $H_O = 0.645$  and  $H_E = 0.836$ ; ranges for  $H_O$  values = 0.485 to 0.771 and  $H_E$  values = 0.624 to 0.816) (Table 1). Most of the wild evolutionary units showed high values of genetic diversity ( $H_E > 0.7$ ), except the North-Pacific unit. Contrastingly, the domesticated group presented extremely low values of genetic diversity ( $H_O = 0.018$  and  $H_E = 0.106$ ).



**Fig. 5** Heatmap of pairwise genetic differentiation ( $R_{ST}$  and  $N_{ST}$ ) values (lower part) and number of migrants per generation ( $Nm$ ) (upper part) among wild evolutionary units and domesticated papaya using two molecular markers: **(a)** genetic differentiation ( $R_{ST}$ ) and number of migrants per generation ( $Nm$ ), using six nuclear microsatellites for six wild evolutionary units and domesticated papaya; **(b)** genetic differentiation ( $N_{ST}$ ), and number of migrants per generation ( $Nm$ ), for the

*psbA-trnH* marker in three genetic groups and domesticated papaya. The colors indicate  $R_{ST}$ ,  $N_{ST}$  or  $Nm$  values ranging from lower values (green) to higher values (red). IG, Isthmus-Gulf unit; NG, North-Gulf unit; YP, Yucatan Peninsula; WYP, Western Yucatan Peninsula unit; SP, South-Pacific unit; NP, North-Pacific unit; Dom, domesticated group. (Color figure online)



**Fig. 6** Recent migration rates ( $m \pm s.d.$ ) between source and recipient populations of wild papaya in Mexico. The arrows show the direction of gene flow between evolutionary units. Only  $m$  values higher than 0.01 are shown; the thicker arrows indicate higher values of  $m$ . The NP unit does not show connections due to low values of  $m$  ( $m < 0.01$ ). The color of each

circle represents the inferred evolutionary unit of Geneland using nuclear microsatellites. Letters inside the circles represent the abbreviation of the evolutionary units: IG, Isthmus-Gulf unit; NG, North-Gulf unit; YP, Yucatan Peninsula; WYP, Western Yucatan Peninsula unit; SP, South-Pacific unit; NP, North-Pacific unit. (Color figure online)

**Table 1** Genetic diversity estimators for six nuclear microsatellites in six evolutionary units and the domesticated group of papaya in Mexico

Populations	n	%P	A	$H_O$	$H_E$
North-Gulf	52	100	8.167 (2.639)	0.699 (0.092)	0.746 (0.103)
Isthmus-Gulf	95	100	11.333 (4.320)	0.693 (0.085)	0.783 (0.052)
South-Pacific	65	100	9.167 (2.229)	0.638 (0.063)	0.789 (0.044)
Yucatan Peninsula	103	100	10.000 (2.000)	0.672 (0.082)	0.729 (0.039)
Western Yucatan Peninsula	40	100	10.333 (3.502)	0.771 (0.090)	0.816 (0.045)
North-Pacific	92	100	8.500 (2.074)	0.485 (0.149)	0.624 (0.181)
Domesticated	109	100	2.800 (0.837)	0.018 (0.015)	0.106 (0.123)
Total, (without domesticated)	447	100	16.833 (4.119)	0.645 (0.037)	0.836 (0.065)

n=number of individuals per evolutionary units, %P=poly-morphic loci percentage, A=allele number,  $H_O$ =observed heterozygosity,  $H_E$ =expected heterozygosity. Standard deviation is indicated in parentheses

## Population structure and chloroplast DNA diversity

Using a total of 296 wild individuals, three genetic clusters were detected by posterior cluster membership (Geneland  $K=3$ ; Fig. 7). Group 1 included the following collection sites: Acayucan, Los Tuxtlas, Villa Guadalupe, Poza Rica, Matias Romero, Ventanilla, Santiago Astata, Marquelia, Caobas, Dzibilchaltún, Chichen Itzá, Río Lagartos, Cancún, Oxtankah, Copala, Tecpan and Tepic; Group 2 included: Cielo, Huasteca, Tamazunchale; and Group 3 included: Bucerías, Costa de Michoacán, Palenque and Mamantel. Overall wild groups showed a high genetic diversity of haplotype diversity ( $H_d=0.934$ ; ranges for  $H_d$  values=0.772–0.918). However, the nucleotide diversity was low ( $\pi=0.009$  to 0.016). In contrast, the domesticated group presented low values of haplotype and nucleotide diversity ( $H_d=0.094$ ;  $\pi=0.000$ ) (Table 2).

Genetic differentiation among the three wild genetic groups inferred by Geneland resulted in a  $N_{ST}=0.409$ . Overall (including wild and domesticated groups)  $N_{ST}$  value was 0.446. Between pairs of wild groups  $N_{ST}$  values ranged from 0.127 to 0.522 (Fig. 5b), while between pairs of the wild groups with the domesticated one,  $N_{ST}$  values ranged from 0.108 to 0.644. The number of migrants per generation

**Fig. 7** Evolutionary units inferred from Geneland using the *psbA-trnH* chloroplast region for wild papaya in its distribution in Mexico. Blue triangles: Group1; yellow triangles: Group2; purple triangles: Group3.



**Table 2** Genetic diversity estimators for the *psbA-trnH* marker in three wild genetic groups and the domesticated group in Mexico

Populations	n	h	S	Hd	$\pi$
Group 1	240	42	25	0.914 (0.010)	0.013 (0.007)
Group 2	47	12	17	0.772 (0.055)	0.016 (0.008)
Group 3	81	23	25	0.918 (0.014)	0.009 (0.005)
Domesticated	42	3	2	0.094 (0.061)	0.000 (0.000)
Total, (without domesticated)	368	68	37	0.934 (0.007)	0.015 (0.008)

n=number of individuals per genetic groups, h=number of haplotypes, S=polymorphic sites, Hd=haplotype diversity,  $\pi$ =nucleotide diversity. Standard deviation is indicated in parentheses

(Nm) among pairs of wild groups was low in some cases, but above 1 between Groups 1 and 3 (Nm values=0.23 to 1.71) (Fig. 5b). However, this estimator was moderate between the wild groups and the domesticated group (Nm < 1) except between Group 3 and the domesticated group (Nm = 2.06).

The haplotype network showed the distribution and frequency of 68 haplotypes for the 24 collection sites (Fig. 8). The most frequent haplotypes (H1, H2, H3) were present in all three wild units (Fig. 8a), except for group 3 that does not include H3 (Fig. 8b). In addition, the evolutionary units detected showed exclusive haplotypes (Fig. 8). The domesticated group was not considered in the haplotype network due to the almost null genetic variation found.

#### Evaluation of transgene monitoring

Of the 187 DNA pools examined, none amplified for the transgene regions (Table 3). To corroborate the absence of transgenes in papaya, individual samples were randomly chosen from each DNA pool to analyze the P35S region, with negative results for all samples.

## Discussion

Our study includes the analysis of the Mexican distribution of *Carica papaya*, area proposed as the center

of origin of the species, the center of distribution of genetic diversity and the site where its domestication began and continues to the present day. Our main objective was to identify the true evolutionary units of wild papaya and to provide necessary information on the areas where the processes that originate and maintain its variation occur for conservation purposes, under the premise that "you cannot conserve what you do not know". Our search for collection sites was exhaustive in order to identify the limits of the evolutionary units, using two different molecular markers (nuclear microsatellites and chloroplast sequences) which, due to their differences, provide complementary histories. The discovery of the evolutionary units will allow changes in the current management and conservation perspectives for the species, and to understand and manage the potential threats faced by the evolutionary units of wild papaya.

#### Wild papaya distribution

In order to obtain the natural distribution of wild papaya, unlike other studies (Fuentes and Santamaría 2014; Espinosa et al. 2018; Hernández-Salinas et al. 2021), we ensured to use data exclusively from wild individuals, by distinguishing wild herbaria specimens from cultivated ones from seven national and international herbariums, and by performing an extensive field sampling. Thus, we are certain that only data from wild specimens was used to construct the potential distribution map. Wild papaya shows a wide distribution in Mexico, covering lowland tropical and subtropical vegetation areas ranging from southern Sinaloa in the west coast and from El Cielo, Tamaulipas, on the east coast, to southern Mexico, where it is widely distributed in the Yucatan Peninsula (Figs. 2, 3). No wild specimens from the states of Sinaloa, Nayarit and Michoacán were found in herbariums, which highlights the importance of the discovery and field collections of wild populations in these areas, since its distribution in these regions was unknown. Alarmingly, the current rate of deforestation and habitat fragmentation in the natural distribution of papaya is high enough as to endanger the existence of plant species (Novick et al. 2003). This, coupled with the lack of information on the state of the CWR, warns about the importance of studying and conserving wild populations of important crop species.



## Genetic diversity, structure and gene flow in wild populations of *Carica papaya* in Mexico

Genetic structure analyses using microsatellite markers and chloroplast DNA recognized six and three genetic groups, respectively. This indicates that wild papaya populations in Mexico behave as very large evolutionary units. The identification of evolutionary units is very useful because it allows to establish the set of individuals or populations with similar genetic characteristics that have diverged for a long time, which in turn helps to generate better conservation strategies that include the ecological and evolutionary processes of the species (Vogler and Desalle 1994; Moritz 1999). The greater area and dispersion that characterize large evolutionary units allows them to resist certain changes, such as environmental impacts and anthropogenic effects, and therefore their risk of local extinction decreases (Moritz 1999; Funk et al. 2012; Casacci et al. 2014). However, vulnerability to evolutionary processes deserves attention, since they can occur at any site of the periphery, or at any individual within the area, silently spreading and affecting a larger area than expected, which is why in-depth studies such as this are needed.

Genetic diversity was, in general, high for wild populations of papaya in Mexico for both markers. Overall, we found similar values of genetic diversity compared to previous studies in Mexico (Chávez-Pesqueira et al. 2014; Chávez-Pesqueira and Núñez-Farfán 2016), but higher than studies of wild papaya in Central America and the Caribbean (D'eeckenbrugge et al. 2007; Ocampo et al. 2007; Brown et al. 2012; Mardonovich et al. 2019). High haplotype diversity values were also obtained with the chloroplast marker, suggesting Mexico as an important genetic reservoir for the species, as genetic diversity has been maintained at high values over time. However, the difference in the number of groups identified with Geneland with the two types of markers could be related to recent anthropogenic effects. At the local level, Chávez-Pesqueira et al. (2014) found that populations located in a fragmented forest in Los Tuxtlas, Mexico, showed lower genetic diversity and higher genetic differentiation compared to populations from continuous forests. The increasing fragmentation of the natural habitat of wild papaya represents a worrisome scenario for the maintenance of genetic partitioning of the wild form of the

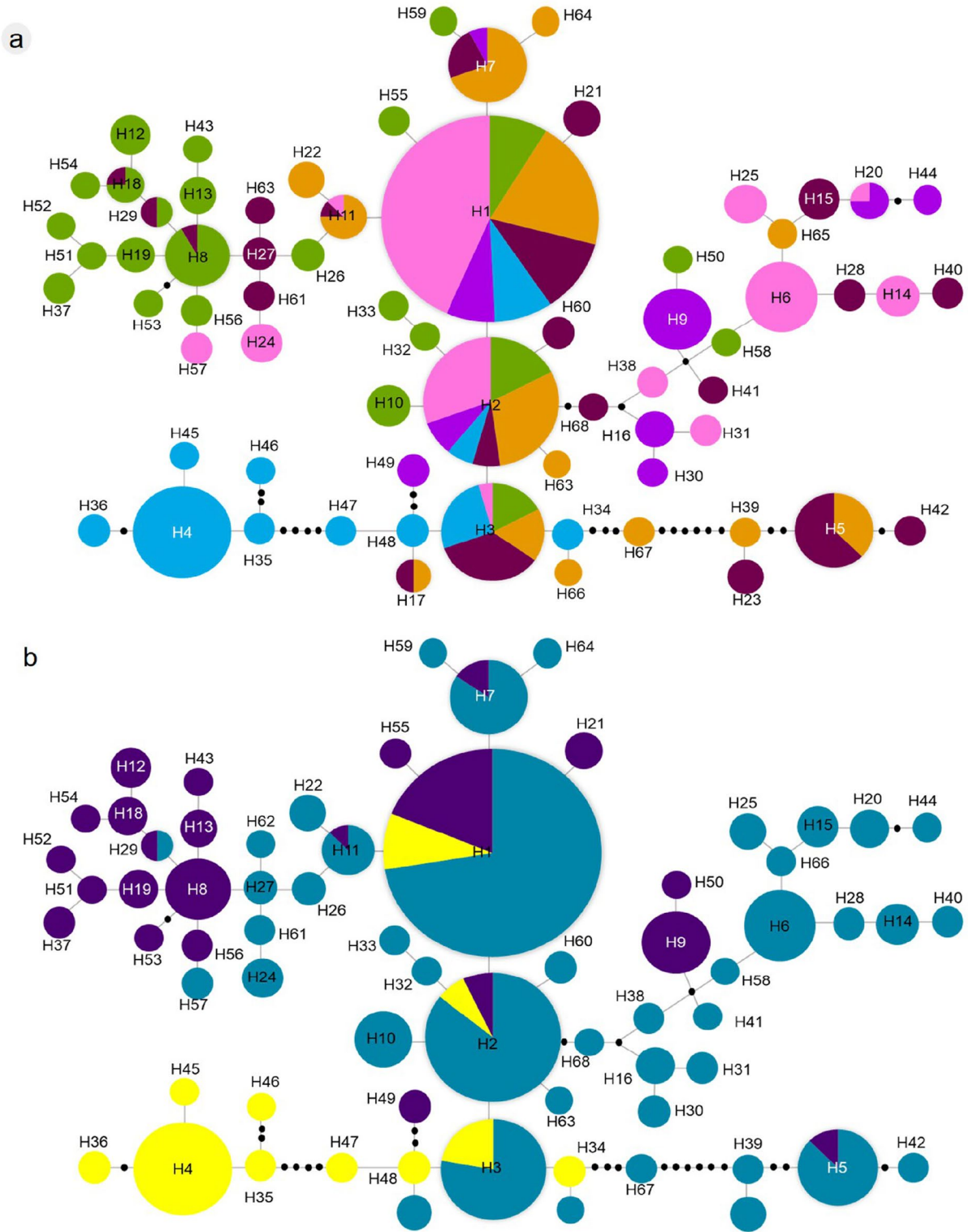
species. On the other hand, cpDNA data suggest that gene flow through seeds was probably important for the maintenance of genetic connectivity in the past, and that recent fragmentation has probably generated barriers for seed dispersers (Chávez-Pesqueira and Núñez-Farfán 2016).

The haplotype network showed diversification processes and gene flow events (Fig. 8a, b). Most units show unique haplotypes and few mutational steps between them. Haplotype networks allow us to observe the spatial distribution of diversity and possible gene flow events, such as between the Yucatan Peninsula to the North Pacific (H24, H57) and vice versa (H50 and H58; Fig. 8). Therefore, gene flow between populations appears to be a frequent evolutionary process that occurs naturally.

## Gene flow between wild and domesticated papaya

The obtained values of  $Nm$  (for both markers) and  $m$  (for nuclear markers) revealed high gene flow among the wild evolutionary units (Figs. 5, 6), as well as the existence of gene flow between Maradol papaya plantations and the wild evolutionary units (Fig. 5). However, we found that the direction is mainly from domesticated to wild plants, and that the highest value of gene flow with the nuclear markers was between the domesticated group and the North-Pacific unit ( $Nm=2.47$ ; Fig. 5a), which presented the lowest values of genetic diversity (Table 1). This unit is embedded in one of the regions with the highest presence of papaya plantations in Mexico (SADER 2021), which could explain the higher gene flow values. Similarly, with the cpDNA markers, we found higher gene flow from plantations to group 3 ( $Nm=2.06$ ; Fig. 5b), which is also the group with the lowest nucleotide diversity. In both cases, we found the lowest levels of genetic differentiation with the domesticated group (Fig. 5), compared to the other wild units, which are more genetically differentiated from it.

The low levels of genetic diversity in domesticated papaya (Table 1, 2), together with the evidence of gene flow between domesticated and wild papaya (Fig. 5), highlight the importance of conserving wild genetic diversity, since gene flow from domesticated plants could affect the fitness of wild units; in addition to modifying the ecological and biological functionality of wild plants by homogenizing their diversity. Commercial plantations in Mexico represent



◀**Fig. 8.** 68 Haplotype network for the *psbA-trnH* chloroplast region of wild *Carica papaya* in its distribution in Mexico. The size of the circles is proportional to the frequency of each haplotype. Black dots represent unsampled haplotypes. Each section, size and color of haplotypes (pies) are proportional to the representation of the evolutionary units inferred from Geneland using (a) nuclear microsatellites (as in Fig. 4a), and (b) *psbA-trnH* chloroplast region (as in Fig. 7)

large extensions with thousands of plants per unit area, which promotes a higher probability of interbreeding with wild plants (Ellstrand et al. 1999). Furthermore, we recommend that future research include native varieties as it is highly likely that they maintain gene flow with commercial plantations and wild populations, as well.

The low diversity in the Maradol improved variety of papaya is determined from its origin, since only two parental lines were used to obtain it, and several rounds of artificial selection were carried out to gain genetic stability and fixation of desirable traits (Rodríguez and Corrales 1967; Rodríguez 2008). In this study, we confirmed that the genetic diversity obtained from plantations from two different regions of Mexico (Pacific and Yucatan Peninsula) is practically nil for both molecular markers (Tables 1, 2). It is known that the domestication process implies a genetic bottleneck that generally reduces the diversity found in wild plants (Yamasaki et al. 2005). In the case of improved varieties, such as Maradol, subsequent bottlenecks occur that further decrease genetic diversity, increasing their vulnerability to future environmental changes (Van De Wouw et al. 2010).

With the nuclear markers, the nine papaya plantations clustered into a single genetic group (Fig. 4b), indicating that plantations have a very similar genetic constitution. Likewise, with the cpDNA marker we found that most domesticated individuals contain the H1 haplotype, which is the most represented in wild plants and is found throughout the natural distribution, supporting that genetic diversity in plantations is limited. We suggest that the limited gene flow from wild to domesticated plants is explained mainly by farmers' management practices, since they generally use commercial seeds. As a result, the commercial cultivation of Maradol papaya creates a significant barrier to gene flow for wild plants. Wild pollen that lands on plantations represents a loss because it won't reach natural populations, reducing the likelihood of outcrossing. However, this study did not

consider papaya plants grown in backyards or home gardens, which could represent important sources of domesticated diversity. Furthermore, plants in home gardens could act as connectors for gene flow, and even facilitate the subsequent introgression of domesticated alleles into wild populations, as wild and domesticated plants share pollinators and floral visitors (Moo-Aldana et al. 2017; Badillo-Montaña et al. 2018; Pacheco-Huh et al. 2021).

In papaya, gene flow from domesticated to wild individuals can cause diverse consequences at different levels, e.g., (1) morphological consequences, such as increased size in fruits (Paz and Vázquez-Yanes 1998; Chávez-Pesqueira and Núñez-Farfán 2016; Fig. 1) and flowers (pers. obs.). Moreover, as individuals of Maradol papaya are mainly hermaphrodites, this characteristic can be transmitted to wild plants (pers. obs.) affecting pollination, since hermaphrodite flowers self-pollinate before anthesis (Manshardt et al. 2016). (2) Physiological consequences, such as decreasing the importance of specific environmental conditions for breaking dormancy and seed germination (Paz and Vázquez-Yanes 1998), affecting the presence of the species in natural seed banks. Moreover, in contrast to Maradol plants, wild plants are tolerant of heat deficit and water stress (Estrella-Maldonado et al. 2021), hence gene flow may lessen this feature in wild plants. Similarly, wild plants are more tolerant to antagonistic damage (Pacheco-Huh et al. 2021), therefore gene flow could increase the vulnerability of wild plants. (3) Ecological consequences, such as modifications to ecological interactions; wild plants interact with a wide range of organisms compared to Maradol plants (Badillo-Montaña et al. 2019; Pacheco-Huh et al. 2021). Therefore, gene flow from domesticated to wild plants could have an impact on the entire chain of ecological interactions, particularly specialized ones. (4) Evolutionary consequences, such as changes in allele frequencies and a tendency to homogenize genetic diversity, reducing diversity in wild units and increasing their vulnerability to environmental changes.

#### Transgene flow monitoring

Transgenic papaya was released in Hawaii in the late twentieth century (Tecson Mendoza et al. 2008; Silva-Rosales et al. 2010; Manshardt 2014) to combat the disease caused by PRSV. Gene flow events between

**Table 3** Pools of *Carica papaya* DNA analyzed and number of positive samples for the three transgenes analyzed

Sample type	n pools	n individuals	P35S (+)	Tnos (+)	PRSV-CP (+)
Wild	64	320	0	0	0
Plantations (domesticated)	48	240	0	0	0
Home garden (domesticated)	27	135	0	0	0
Commercials seeds (domesticated)	48	240	0	0	0

n pools = number of pools analyzed by type of sample; n individuals = number of individuals by pools; + means positive samples

transgenic varieties of papaya to other varieties of the species, have been reported (Gonsalves et al. 2012; Gonsalves 2014; Manshardt et al. 2016), showing the occurrence of gene flow from the hermaphrodite variety "Rainbow" (transgenic) to other non-hermaphrodite varieties of papaya. Currently, the countries where the importation of transgenic papaya from Hawaii is allowed are the United States, Canada and Japan (Chávez-Pesqueira and Núñez-Farfán 2017). In Mexico, there have been attempts to create and release transgenic papaya (Cabrera-Ponce et al. 1995; de la Fuente et al. 1997; Guzmán-González et al. 2006; Silva-Rosales et al. 2010). However, Mexican farmers have opposed to the introduction of transgenic papaya because they opt to improve their management practices to deal with PRSV (Silva-Rosales et al. 2010). Moreover, without transgenic papaya, Mexico represents the world's leading exporter of papaya (FAO 2022),

Transgene flow can be assessed with molecular markers to analyze introgression from domesticated to wild relatives, as well as other risks to genetic diversity within the wild-to-domesticated complex (Wegier et al. 2011; Hernández-Terán et al. 2017). Transgene flow is a cause for concern as the goal of CWR conservation is to maintain the processes that favor their continued diversification and potential in new scenarios (Tobón-Niedfeldt et al. 2022), and transgene introduction in wild populations have shown to modify native ecological interactions (Vázquez-Barrios et al. 2021). In the extensive sampling conducted in this research, from commercial plantations, local varieties and wild papaya individuals, transgenes were absent (Table 3). However, if the country's protection goals include maintaining the genetic diversity of the wild-to-domesticated papaya complex without transgenes, legal and accidental releases must be prevented.

### Conservation and management for *Carica papaya*

There is great concern about the conservation of many CWR. It has been estimated that up to 75% of wild species could be under some kind of threat (Dempewolf et al. 2014), mainly due to anthropogenic actions (Sala et al. 2000; Wright 2010; Haddad et al. 2015; Newbold et al. 2015), affecting the ecological processes that promote and maintain the genetic diversity of wild populations. In the case of wild papaya, although it has been recognized as a high priority species for conservation (Castañeda-Álvarez et al. 2016), assessments corresponding to the IUCN red list, NOM-059-SEMARNAT-2001, and other strategies to safeguard evolutionary resilience (Mastretta-Yanes et al. 2018; Tobón-Niedfeldt et al. 2022), are still needed. Our results are of great importance for global and local agrobiodiversity conservation plans, and show the importance of considering the wild-to-domesticated complex of species in conservation strategies. Although the crop's vulnerability is growing due to its low genetic variation, Mexico's cultivated areas and its exports have increased in recent years (FAO 2022). The likelihood that gene flow from domesticated to wild papaya plants will continue, should be addressed with comprehensive measures, governmental tools, awareness of markets and consumers, and considering the communities that still conserve the diversity of native varieties.

The integral conservation of agrobiodiversity depends on various governmental bodies in Mexico, which should be informed of these results in order to improve decision-making. In the same way, local farmers should be able to take better decisions by knowing and valuing the current state of their genetic resources. Crops that remain in their centers of origin, diversity and domestication face complexities that should be evaluated and recognized. Every



part of the wild-to-domesticated complex involves a rich and unique evolutionary history that silently offers us future opportunities but just as silently is lost unseen.

For papaya we recommend the constant evaluation of genetic diversity in the wild evolutionary units to detect possible reductions in genetic diversity, as well as a systematic transgene monitoring in the wild-to-domesticated complex, to report any case of transgene introduction. We suggest paying special attention to the wild evolutionary units that remain in areas of high papaya production, since gene flow from plantations is more likely to reduce their genetic diversity.

**Acknowledgments** We thank Pilar Suárez-Montes and Rosalinda Tapia-López for their assistance in obtaining part of the genetic data. To the Laboratory of Molecular Biology of the Institute of Biology of the UNAM and to its academic technicians in charge, for their support in the laboratory work. We thank Gabriel Dzib, Matilde Otriz García and Erika Dzib Quintal for their support with field and laboratory work. We are very grateful to the herbariums CICY, MEXU, UADY, NYBG, ENCB, XAL and Los Tuxtlas Biological Station, UNAM for the support to our study. This work was supported by the grants from Dirección General del Sector Primario y Recursos Naturales Renovables (DGSPRNR) belonging to the SEMARMAT and coordinated by CONABIO (WQ003) awarded to JNF and MCP, CONACyT Ciencia Básica 2017–2018 (A1-S-14559) awarded to MCP, and PAPIIT-UNAM (IN214719) awarded to AW. This study constitutes part of the PhD research of PJRG, who received a scholarship from the Consejo Nacional de Ciencia y Tecnología (CONACyT, CVU. 774517) and extends thanks to the Posgrado en Ciencias Biológicas, CICY. VA (CONACyT, CVU. 1033774) and SRP (CONACyT, CVU. 776199) also thank their CONACyT support respectively and the Posgrado en Ciencias, UNAM. Finally, we are grateful for the support of all people in the communities where our study was developed for their time and interest in sharing their knowledge.

**Author contributions** This project was conceptualized by MCP, JNF and AW. Data collection was led by MCP, JNF and AW. Material preparation and labwork were led by MCP and SRP. PJRG, VA, SRP, AW and MCP conducted the analyses. PJRG, AW, and MCP wrote and revised the manuscript. VA and JNF provided valuable inputs for the improvement of the final manuscript. The funding for this work was acquired by MCP, JNF and AW. All authors read and approved the final manuscript.

**Funding** The authors were supported by grants from Dirección General del Sector Primario y Recursos Naturales Renovables (DGSPRNR) belonging to the SEMARMAT and coordinated by CONABIO, project number WQ003, and from Ciencia Básica CONACyT 2017–2018 (A1-S-14559), and PAPIIT-UNAM (IN214719).

**Data availability** The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

#### Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

#### References

- Aradhya MK, Manshardt RM, Zee F, Morden CW (1999) A phylogenetic analysis of the genus *Carica* L. (Caricaceae) based on restriction fragment length variation in a cpDNA intergenic spacer region. *Genet Resour Crop Evol* 46:579–586. <https://doi.org/10.1023/A:1008786531609>
- Badillo-Montaña R, Aguirre A, Santamaría F et al (2018) Importancia de los visitantes florales y la expresión sexual en el éxito reproductivo de un cultivo de papaya (*Carica papaya* L.). In: Ramírez-Bautista A, Pineda-López R (eds) *Ecología y Conservación de Fauna en Ambientes Antropizados*. CONACyT-UAQ, Mexico, pp 5–17
- Badillo-Montaña R, Aguirre A, Munguía-Rosas MA (2019) Pollinator-mediated interactions between cultivated papaya and co-flowering plant species. *Ecol Evol* 9:587–597. <https://doi.org/10.1002/ece3.4781>
- Bartsch D, Lehnen M, Clegg J et al (1999) Impact of gene flow from cultivated beet on genetic diversity of wild sea beet populations. *Mol Ecol* 8:1733–1741. <https://doi.org/10.1046/j.1365-294X.1999.00769.x>
- Barve N, Barve V, Jiménez-Valverde A et al (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecol Model* 222:1810–1819. <https://doi.org/10.1016/J.ECOLMODEL.2011.02.011>
- Brown JE, Bauman JM, Lawrie JF et al (2012) The structure of morphological and genetic diversity in natural populations of *Carica papaya* (Caricaceae) in Costa Rica. *Biotropica* 44:179–188. <https://doi.org/10.1111/j.1744-7429.2011.00779.x>
- Brozynska M, Furtado A, Henry RJ (2016) Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnol J* 14:1070–1085. <https://doi.org/10.1111/pbi.12454>
- Cabrera-Ponce JL, Vegas-García A, Herrera-Estrella L (1995) Herbicide resistant transgenic papaya plants produced by an efficient particle bombardment transformation method. *Plant Cell Rep* 15(15):1–7. <https://doi.org/10.1007/BF01690242>
- Carvalho FA, Renner SS (2012) A dated phylogeny of the papaya family (Caricaceae) reveals the crop's closest relatives and the family's biogeographic history. *Mol Phylogenet Evol* 65:46–53. <https://doi.org/10.1016/J.YMPEV.2012.05.019>
- Casacci LP, Barbero F, Balletto E (2014) The “Evolutionarily Significant Unit” concept and its applicability in biological conservation. *Ital J Zool* 81:182–193. <https://doi.org/10.1080/11250003.2013.870240>

- Castañeda-Álvarez NP, Khoury CK, Achicanoy HA et al (2016) Global conservation priorities for crop wild relatives. *Nat Plants* 2:1–6. <https://doi.org/10.1038/NPLANT.S.2016.22>
- Chávez-Pesqueira M, Núñez-Farfán J (2016) Genetic diversity and structure of wild populations of *Carica papaya* in Northern Mesoamerica inferred by nuclear microsatellites and chloroplast markers. *Ann Bot* 118:1293–1306. <https://doi.org/10.1093/aob/mcw183>
- Chávez-Pesqueira M, Núñez-Farfán J (2017) Domestication and genetics of papaya: a review. *Front Ecol Evol* 5:1–9. <https://doi.org/10.3389/fevo.2017.00155>
- Chávez-Pesqueira M, Suárez-Montes P, Castillo G, Núñez-Farfán J (2014) Habitat fragmentation threatens wild populations of *Carica papaya* (Caricaceae) in a lowland rainforest. *Am J Bot* 101:1092–1101. <https://doi.org/10.3732/ajb.1400051>
- Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Mol Ecol* 9:1657–1659. <https://doi.org/10.1046/j.1365-294x.2000.01020.x>
- European Commission (2014) JRC GMO-Matrix Join research Centre. <https://gmo-crl.jrc.ec.europa.eu/jrcgmomatrix/matrices/full>. Accessed 1 Dec 2022
- Cuervo-Robayo AP, Téllez-Valdés O, Gómez-Albores MA et al (2014) An update of high-resolution monthly climate surfaces for Mexico. *Int J Climatol* 34:2427–2437. <https://doi.org/10.1002/JOC.3848>
- D'eeckenbrugge GC, Restrepo MT, Jiménez D, Mora E (2007) Morphological and isozyme characterization of common papaya in Costa Rica. *Acta Hortic* 740:109–120. <https://doi.org/10.17660/ACTAHORTIC.2007.740.11>
- de Fuente JM, Ramirez-Rodríguez V, Cabrera-Ponce JL, Herrera-Estrella L (1997) Aluminum tolerance in transgenic plants by alteration of citrate synthesis. *Science* (80-) 276:1566–1568. <https://doi.org/10.1126/SCIENCE.276.5318.1566>
- Dempewolf H, Eastwood RJ, Guarino L et al (2014) Adapting agriculture to climate change: a global initiative to collect, conserve, and use crop wild relatives. *Agroecol Sustain Food Syst* 38:369–377. <https://doi.org/10.1080/21683565.2013.870629>
- Dempewolf H, Baute G, Anderson J et al (2017) Past and future use of wild relatives in crop breeding. *Crop Sci* 57:1070–1082. <https://doi.org/10.2135/cropsci2016.10.0885>
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15
- Ehrlich PR, Raven PH (1969) Differentiation of populations: gene flow seems to be less important in speciation than the neo-Darwinians thought. *Science* 165:1228–1232. <https://doi.org/10.1126/science.165.3899.1228>
- Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. *Annu Rev Ecol Syst* 30:539–563. <https://doi.org/10.1146/annurev.ecolsys.30.1.539>
- Escobar LE, Lira-Noriega A, Medina-Vogel G, Townsend Peterson A (2014) Potential for spread of the white-nose fungus (*Pseudogymnoascus destructans*) in the Americas: use of Maxent and NicheA to assure strict model transference. *Geospat Health* 9:221–229. <https://doi.org/10.4081/gh.2014.19>
- Espinosa E, Gámez AJ, Avila MA, et al (2018) Distribución geográfica potencial de papaya silvestre cultivada en México. *Rev Mex Ciencias Agrícolas* 9:1377–1388. <https://doi.org/10.29312/remexca.v9i7.550>
- ESRI (2015) Data and Maps for ArcGIS (2015). <https://www.esri.com/arcgis-blog/products/analytics/analytics/data-and-maps-for-arcgis-2015-is-now-available/>. Accessed 1 Dec 2022
- Estrella-Maldonado H, Ramírez AG, Ortiz GF et al (2021) Transcriptomic analysis reveals key transcription factors associated to drought tolerance in a wild papaya (*Carica papaya*) genotype. *PLoS ONE* 16:e0245855. <https://doi.org/10.1371/journal.pone.0245855>
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* 14:2611–2620. <https://doi.org/10.1111/j.1365-294x.2005.02553.x>
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491. <https://doi.org/10.1007/s00424-009-0730-7>
- FAO (2022) Major tropical fruits: market review 2021. Rome
- Fuentes G, Santamaría JM (2014) Papaya (*Carica papaya* L.): Origin, domestication, and production. In: Ming R, Moore PH (eds) Genetics and genomics of papaya. Springer, New York, pp 3–16
- Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends Ecol Evol* 27:489–496. <https://doi.org/10.1016/j.tree.2012.05.012>
- Gepts P, Papa R (2003) Evolution during domestication. In: *Encyclopedia of life sciences*. Wiley, Chichester
- Goettsch B, Urquiza-Haas T, Koleff P et al (2021) Extinction risk of Mesoamerican crop wild relatives. *Plants People Planet* 3:775–795. <https://doi.org/10.1002/PPP3.10225>
- Gonsalves D (2014) Hawaii's transgenic papaya story 1978–2012: a personal account. In: Ming R, Moore PH (eds) Genetics and genomics of papaya. Springer, New York, pp 115–142
- Gonsalves D, Gonsalves C, Carr J et al (2012) Assaying for pollen drift from transgenic 'Rainbow' to nontransgenic 'Kapoho' papaya under commercial and experimental field conditions in Hawaii. *Trop Plant Biol* 5:153–160. <https://doi.org/10.1007/S12042-011-9090-5>
- Guillot G, Santos F, Estoup A (2008) Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics* 24:1406–1407. <https://doi.org/10.1093/BIOINFORMATICS/BTN136>
- Guzmán-González S, Valadez-Ramírez P, Robles-Berber R-E et al (2006) Biolistic genetic transformation of *Carica papaya* L. using the Helios™ Gene Gun. *HortScience* 41:1053A – 1053. <https://doi.org/10.2127/HORTSCI.41.4.1053A>
- Haddad NM, Brudvig LA, Clobert J et al (2015) Habitat fragmentation and its lasting impact on Earth's ecosystems.

- Sci Adv 1:e1500052. <https://doi.org/10.1126/sciadv.1500052>
- Hails RS (2000) Genetically modified plants: the debate continues. *Trends Ecol Evol* 15:14–18. [https://doi.org/10.1016/S0169-5347\(99\)01751-6](https://doi.org/10.1016/S0169-5347(99)01751-6)
- Harlan JR, de Wet MJM (1971) Toward a rational classification of cultivated plants. *Taxon* 20:509. <https://doi.org/10.2307/1218252>
- Hernández-Salinas G, Luna-Cavazos M, Soto-Estrada A et al (2021) Distribution and eco-geographic characterization of *Carica papaya* L. native to Mexico. *Genet Resour Crop Evol* 69(69):99–116. <https://doi.org/10.1007/S10722-021-01207-3>
- Hernández-Terán A, Wegier A, Benítez M et al (2017) Domesticated, genetically engineered, and wild plant relatives exhibit unintended phenotypic differences: a comparative meta-analysis profiling rice, canola, maize, sunflower, and pumpkin. *Front Plant Sci* 8:2030. <https://doi.org/10.3389/FPLS.2017.02030/BIBTEX>
- Jiménez VM, Mora-Newcomer E, Gutiérrez-Soto MV (2014) Biology of the papaya plant. In: Ming R, Moore PH (eds) *Genetics and genomics of papaya*. Springer, New York, pp 17–33
- Lynch M, Crease TJ (1990) The analysis of population survey data on DNA sequence variation. *Mol Biol Evol* 7:377–394. <https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A040607>
- Manshardt R (2014) History and future of the solo papaya. In: Paul H, Moore PH (eds) *Genetics and genomics of papaya*. Springer, New York, pp 95–111
- Manshardt R, Bishaw D, Pitz K, Stewart CN (2016) Gene flow from commercial transgenic papaya fields into feral populations in Hawaii. *Acta Hort* 1124:33–40. <https://doi.org/10.17660/ACTAHORTIC.2016.1124.5>
- Mardonovich S, Tepe EJ, Moore RC (2019) The natural diversity of *Carica papaya* in Panama. *Acta Hort* 1250:99–109. <https://doi.org/10.17660/ACTAHORTIC.2019.1250.15>
- Mastretta-Yanes A, Acevedo Gasman F, Burgeff C et al (2018) An initiative for the study and use of genetic diversity of domesticated plants and their wild relatives. *Front Plant Sci* 9:209. <https://doi.org/10.3389/FPLS.2018.00209/BIBTEX>
- Moo-Aldana RD, Munguía-Rosas MA, Serralta LP et al (2017) Can the introduction of modern crop varieties in their centre of origin affect local ecological knowledge? A case study of papaya in the Yucatan Peninsula. *Hum Ecol* 45:367–375. <https://doi.org/10.1007/s10745-017-9903-8>
- Moritz C (1999) Conservation units and translocations: strategies for conserving evolutionary processes. *Hereditas* 130:217–228. <https://doi.org/10.1111/J.1601-5223.1999.00217.X>
- Muscarella R, Galante PJ, Soley-Guardia M et al (2018) Package ‘ENMeval’. Automated Runs and Evaluations of Ecological Niche Models. Version 0.3.0 2
- Newbold T, Hudson LN, Hill SLL et al (2015) Global effects of land use on local terrestrial biodiversity. *Nature* 520:45–50. <https://doi.org/10.1038/nature14324>
- Novick RR, Dick CW, Lemes MR et al (2003) Genetic structure of Mesoamerican populations of Big-leaf mahogany (*Swietenia macrophylla*) inferred from microsatellite analysis. *Mol Ecol* 12:2885–2893. <https://doi.org/10.1046/J.1365-294X.2003.01951.X>
- Núñez-Farfán J, Dirzo R (1988) Within-gap spatial heterogeneity and seedling performance in a Mexican tropical forest. *Oikos* 51:274. <https://doi.org/10.2307/3565308>
- Ocampo J, Dambier D, Ollitrault P et al (2006) Microsatellite markers in *Carica papaya* L.: isolation, characterization and transferability to *Vasconcellea* species. *Mol Ecol Notes* 6:212–217. <https://doi.org/10.1111/J.1471-8286.2006.01197.X>
- Ocampo PJ, Coppens d’Eeckenbrugge G, Risterucci AM et al (2007) Papaya genetic diversity assessed with microsatellite markers in germplasm from the Caribbean region. *Acta Hort*. <https://doi.org/10.17660/ActaHortic.2007.740.9>
- Osorio-Olvera L, Lira-Noriega A, Soberón J et al (2020) ntbox: an r package with graphical user interface for modelling and evaluating multidimensional ecological niches. *Methods Ecol Evol* 11:1199–1206. <https://doi.org/10.1111/2041-210X.13452>
- Pacheco-Huh J, Carmona D, Dzib G, Chávez-Pesqueira M (2021) Mutualistic and antagonistic interactions differ in wild and domesticated papaya (*Carica papaya*) in its center of origin. *Plant Biol*. <https://doi.org/10.1111/plb.13214>
- Papa R, Gepts P (2003) Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theor Appl Genet* 106:239–250. <https://doi.org/10.1007/s00122-002-1085-z>
- Papa R (2005) Gene flow and introgression between domesticated crops and their wild relatives. Turin, Italy
- Paz L, Vázquez-Yanes C (1998) Comparative seed ecophysiology of wild and cultivated *Carica papaya* trees from a tropical rain forest region in Mexico. *Tree Physiol* 18:277–280. <https://doi.org/10.1093/TREEPHYS/18.4.277>
- Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Peterson AT, Soberón J (2012) Species distribution modeling and ecological niche modeling: getting the concepts right. *Nat Conserv* 10:102–107. <https://doi.org/10.4322/natcon.2012.019>
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Model* 190:231–259. <https://doi.org/10.1016/J.ECOLMODEL.2005.03.026>
- Phillips SJ, Dudík M, Schapire RE (2004) A maximum entropy approach to species distribution modeling. In: *Proceedings, Twenty-First Int Conf Mach Learn ICML 2004*, pp 655–662. <https://doi.org/10.1145/1015330.1015412>
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype Data. *Genetics* 155:945–959
- R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.Rproject.org/>

- Rodríguez NA, Corrales SR (1967) El papayo ‘Maradol.’ La Habana, Cuba
- Rodríguez MA (2008) De la ciencia popular a la industria: la variedad cubana de papaya Maradol. La Habana, Cuba
- Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC et al (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 34:3299–3302. <https://doi.org/10.1093/MOLBEV/MSX248>
- SADER (2021) México, principal exportador de papaya en el mundo; crece producción 3.2 por ciento en 2020. *Secr. Agric. y Desarro. Rural*
- Sala OE, Chapin FS, Armesto JJ et al (2000) Global biodiversity scenarios for the year 2100. *Science* (80-) 287:1770–1774
- Salamon M, Lévêque L, Ballenghien M, Viard F (2020) Spill-back events followed by self-sustainment explain the fast colonization of a newly built marina by a notorious invasive seaweed. *Biol Invasions* 22:1411–1429. <https://doi.org/10.1007/s10530-019-02193-5>
- SIAP (2018) Servicio de Información Agroalimentaria y Pesquera. <https://nube.siap.gob.mx/cierreagricola/>. Accessed 12 May 2020
- Silva-Rosales L, Roberto Gonzalez De Leon D, Chauvet M (2010) Why there is no transgenic papaya in Mexico. *Transgenic Plant J* 4:45–51
- Simoes M, Romero-Alvarez D, Nuñez-Penichet C et al (2020) General theory and good practices in ecological niche modeling: a basic guide. *Biodivers Inform* 15:67–68. <https://doi.org/10.17161/BI.V15I2.13376>
- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* (80-) 236:787–792. <https://doi.org/10.1126/science.3576198>
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462. <https://doi.org/10.1093/GENETICS/139.1.457>
- Snow AA (2002) Transgenic crops—why gene flow matters. *Nat Biotechnol* 20:542–542. <https://doi.org/10.1038/nbt0602-542>
- Soriano-Melgar LL, Alcaraz-Meléndez L, Rodríguez-Álvarez M, Real-Cosío S (2016) Colecta y conservación in vitro y ex situ de recursos fitogenéticos de *carica papaya* L. *Agroproductividad* 9:28–32
- Tecson Mendoza EM, Laurena AC, Botella JR (2008) Recent advances in the development of transgenic papaya technology. *Biotechnol Annu Rev* 14:423–462. [https://doi.org/10.1016/S1387-2656\(08\)00019-7](https://doi.org/10.1016/S1387-2656(08)00019-7)
- Teixeira Da Silva JA, Rashid Z, Nhut DT et al (2007) Papaya (*Carica papaya* L.) biology and biotechnology. In: *Tree and forestry science and biotechnology*, 1 edn. pp 47–73
- Thomas E, Ramirez M, van Zonneveld M et al (2016) Assessment of the conservation status of Mesoamerican crop species and their wild relatives in light of climate change. In: Maxted N, Dullo ME, Ford-Lloyd BV (eds) *Enhancing crop genepool use: capturing wild relative and landrace diversity for crop improvement*. CABI, pp 248–270
- Tobón-Niedfeldt W, Mastretta-Yanes A, Urquiza-Haas T et al (2022) Incorporating evolutionary and threat processes into crop wild relatives conservation. *Nat Commun* 13:1–18. <https://doi.org/10.1038/s41467-022-33703-0>
- Van De Wouw M, Kik C, Van Hintum T et al (2010) Genetic erosion in crops: concept, research results and challenges. *Plant Genet Resour* 8:1–15. <https://doi.org/10.1017/S1479262109990062>
- Vázquez E, Mata H, Ariza R, Santamaría F (2010) Producción y manejo postcosecha de papaya Maradol en la planicie Huaesteca. INIFAP-SAGARPA, Mexico
- Vázquez-Barrios V, Boege K, Sosa-Fuentes TG et al (2021) Ongoing ecological and evolutionary consequences by the presence of transgenes in a wild cotton population. *Sci Rep* 11:1–10. <https://doi.org/10.1038/s41598-021-81567-z>
- Vogler AP, Desalle R (1994) Diagnosing units of conservation management. *Conserv Biol* 8:354–363. <https://doi.org/10.1046/J.1523-1739.1994.08020354.X>
- Wegier A, Piñeyro-Nelson A, Alarcón J, et al (2011) Recent long-distance transgene flow into wild populations conforms to historical patterns of gene flow in cotton (*Gossypium hirsutum*) at its centre of origin. *Mol Ecol* 20:4182–4194. <https://doi.org/10.1111/J.1365-294X.2011.05258.X>
- Wegier A (2013) Diversidad genética y conservación de *Gossypium hirsutum* silvestre y cultivado en México. Universidad Nacional Autónoma de México
- Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163:1177. <https://doi.org/10.1093/genetics/163.3.1177>
- Wright SJ (2010) The future of tropical forests. *Ann N Y Acad Sci* 1195:1–27. <https://doi.org/10.1111/j.1749-6632.2010.05455.x>
- Wu M, Lewis J, Moore RC (2017) A wild origin of the loss-of-function lycopene beta cyclase (CYC-b) allele in cultivated, red-fleshed papaya (*Carica papaya*). *Am J Bot* 104:116–126. <https://doi.org/10.3732/AJB.1600301>
- Yamasaki M, Tenaillon MI, Bi IV, et al (2005) A Large-Scale Screen for Artificial Selection in Maize Identifies Candidate Agronomic Loci for Domestication and Crop Improvement. *Plant Cell* 17:2859–2872. <https://doi.org/10.1105/TPC.105.037242>

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.