Received: 13 December 2021 | Accepted: 9 May 2022

DOI: 10.1002/ajb2.16019

RESEARCH ARTICLE

Botany

Genetic diversity, gene flow, and differentiation among wild, semiwild, and landrace chile pepper (*Capsicum annuum*) populations in Oaxaca, Mexico

Ana L. Pérez-Martínez[1] | **Luis E. Eguiarte**[2] | **Kristin L. Mercer**[3] | **Natalia E. Martínez-Ainsworth**[1] | **Leah McHale**[3] | **Esther van der Knaap**[4] | **Lev Jardón-Barbolla**[1]

[1]Centro de Investigaciones Interdisciplinarias en Ciencias y Humanidades, Universidad Nacional Autónoma de México, Torre II de Humanidades 4°, 5° y 6° pisos, Circuito Interior, C.P. 04510, Ciudad Universitaria, Ciudad de México, México

[2]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México; Circuito exterior s/n anexo al Jardín Botánico. C.P. 04510. Ciudad Universitaria, Ciudad de México, México

[3]Department of Horticulture and Crop Science, Ohio State University, Columbus, OH 43210 USA

[4]Department of Horticulture, Institute of Plant Breeding, Genetics & Genomics, University of Georgia, Athens, GA 30602 USA

Correspondence

Lev Jardón-Barbolla, Centro de Investigaciones Interdisciplinarias en Ciencias y Humanidades, Universidad Nacional Autónoma de México, Torre II de Humanidades 4°, 5° y 6° pisos, Circuito Interior, C.P. 04510, Ciudad Universitaria, Ciudad de México. México.
Email: levjardon@ciencias.unam.mx

Abstract

Premise: *Capsicum annuum* (Solanaceae) was originally domesticated in Mexico, where wild (*C. annuum* var. *glabriusculum*) and cultivated (*C. annuum* var. *annuum*) chile pepper populations (>60 landraces) are common, and wild-resembling individuals (hereafter semiwild) grow spontaneously in anthropogenic environments. Here we analyze the role of elevation and domestication gradients in shaping the genetic diversity in *C. annuum* from the state of Oaxaca, Mexico.

Methods: We collected samples of 341 individuals from 28 populations, corresponding to wild, semiwild (*C. annuum* var. *glabriusculum*) and cultivated *C. annuum*, and closely related species *Capsicum frutescens* and *C. chinense*. From the genetic variation of 10 simple sequence repeat (SSR) loci, we assessed the population genetic structure, inbreeding, and gene flow through variance distribution analyses, genetic clustering, and connectivity estimations.

Results: Genetic diversity (H_E) did not differ across domestication levels. However, inbreeding coefficients were higher in semiwild and cultivated chiles than in wild populations. We found evidence for gene flow between wild populations and cultivated landraces along the coast. Genetic structure analysis revealed strong differentiation between most highland and lowland landraces.

Conclusions: Gene flow between wild and domesticated populations may be mediated by backyards and smallholder farms, while mating systems may facilitate gene flow between landraces and semiwild populations. Domestication and elevation may overlap in their influence on genetic differentiation. Lowland Gui'ña dani clustered with highland landraces perhaps due to the social history of the Zapotec peoples. In situ conservation may play an important role in preserving semiwild populations and private alleles found in landraces.

KEYWORDS

agrobiodiversity, *Capsicum*, crop evolution, domestication gradient, genetic structure, SSR markers

In regions where a crop has been cultivated for a long time, especially in those close to its original domestication area, the local genetic diversity contained in both wild and landrace populations acts as a genetic reservoir for the evolution of the crop (Brush, 2000). Environmental and/or altitudinal gradients can play an important role in shaping genetic variation across diversity centers (Vavilov, 1931; Brush, 2000; Long, 2009; Aguirre-Liguori et al., 2016; Mercer and Perales, 2019). Landraces show a great capacity of adaptation to diverse climate conditions, day length, elevation, and soil type (Long-Solís, 2012; Mercer and Perales, 2019). While crop plant domestication occurred first in pre-Columbian agricultural communities, in Mexico, crop evolution remains an ongoing process in campesino communities, where landrace seeds are saved and wild and semiwild plants remain part of the agroecosystem (Pressoir

and Berthaud, 2004). Often, a suite of different crops and landraces within a crop is maintained in traditional agricultural systems because of the significance of each one in local traditions or preferences (Idohou et al., 2014; Rivera et al., 2016). By studying this dynamic system, we can better understand the interactions of multiple natural and human-mediated evolutionary forces acting on centers of crop diversity.

Chile peppers (*Capsicum* spp.) belong to the Solanaceae and are among the plants domesticated in Mexico and Mesoamerica (Nuez et al., 2003; Long-Solís, 2012; Pickersgill, 2016). There is evidence of use and cultivation of *Capsicum annuum* since pre-Hispanic times and of its long cultural history in Mexico (Long-Solís, 2012; Kraft et al., 2014). The genus includes five domesticated species: *C. annuum* L., *C. chinense* Jacq. (habanero pepper), *C. frutescens* L. (tabasco pepper), *C. baccatum* L. and *C. pubescens* R. P. (apple pepper). All but *C. baccatum* are cultivated in Mexico, although *C. chinense* and *C. pubescens* originated and were domesticated in South America (Pickersgill, 1971, 1997, 2016; McLeod et al., 1983; Loaiza Figueroa et al., 1989; Hernández-Verdugo et al., 1999, 2001; Oyama et al., 2006). A great diversity of Mexican dishes and traditions in each region of the country make specific uses of particular types of pepper (Long-Solís, 2012; Muñoz-Zurita, 2015). These uses often have an extended cultural history associated with customs, food preferences rituals, and vocabulary. Chile peppers are also used as medicinal and ornamental plants (Long, 2009; Long-Solís, 2012; Hill et al., 2013; González-Pérez et al., 2014). Mexico is the second largest producer of fresh chile peppers (FAO, 2019) and is currently ranked as a worldwide leader exporting chile peppers to 43 countries (SAGARPA, 2017).

Capsicum annuum is one of the most important domesticated species in Mexico (Perry and Flannery, 2007; Long-Solís, 2012), with more than 60 landraces coexisting with their wild relative (*C. annuum* var. *glabriusculum*) throughout the country (Aguilar-Rincón et al., 2010; Montes-Hernández et al., 2010; SEMARNAT, 2018). Human communities play an important role in the evolution and conservation of wild, semiwild, and landrace populations of *C. annuum* in situ (Rai et al., 2013). Chile pepper landraces tend to be adapted to specific environments and have been selected for particular uses, resulting in a vast array of phenotypic diversity, including pungency, secondary flavors, fruit color and odor, and plant architecture (Aguilar-Rincón et al., 2010; Mapes and Basurto, 2016). Some of these landraces are adapted to grow in distinct environments. Moreover, there are also sites where peppers that are morphologically indistinguishable from wild peppers (hereafter, semiwild) grow spontaneously in human-managed spaces such as backyards or small milpa plots where maize and beans are the main crops (Caballero et al., 2004; Casas et al., 2007, 2016).

Those semiwild chiles are understood as a domestication category defined by Aguilar-Meléndez et al. (2009), who described plants found in human-managed spaces with traits that correspond to wild *C. annuum* var. *glabrisuculum*, i.e., perennial shrubs with small, deciduous, red mature fruits (Aguilar-Meléndez et al., 2009). Our use of this category emphasizes the management component of such a grouping. We refrain from using "weedy" or "feral" because they suggest either a nuisance plant or a domesticated–gone-wild sequence of events. Semiwild individuals correspond to chile perennial shrubs growing spontaneously and left to grow (let-standing) by farmers within anthropogenic environments and are also known as *arvenses* (from the Latin *arvum*, meaning ploughed, i.e., plants associated with anthropogenic disturbance). Farmers are aware that these semiwild peppers are dispersed by birds (Carlo and Tewksbury, 2014), and although not actively sown by humans, they are highly appreciated as part of the associated agrobiodiversity (as described by Perfecto et al., 2009). This cultural relevance and its associated traditional ecological knowledge are expressed in the fact that in Mexico wild and semiwild peppers receive different common names including Chilegole, Chiltepin, Chile de monte, Maax ik, Timpinchile, Amashito, and Chilgol (Aguilar-Rincón et al., 2010; names reported by people during our field collections). Here, we capitalize these common names for easy recognition. Inclusion of both semiwild and wild chile pepper types, along with diverse landraces will help us best capture existing genetic diversity in this system.

Domestication and genetic diversity of *C. annuum* populations in Mexico have been studied with different molecular markers including isoenzymes, RAPDs, SSRs, and more recently with SNPs (Hernández-Verdugo et al., 2001; Oyama et al., 2006; Contreras et al., 2011; González-Jara et al., 2011; Pacheco-Olvera et al., 2012; Toledo-Aguilar et al., 2016; Taitano et al., 2019). Many previous studies have explored the structure and genetic diversity of wild pepper populations while analyzing only a small number of domesticated populations (Hernández-Verdugo et al., 2001; Oyama et al., 2006; González-Jara et al., 2011; Pacheco-Olvera et al., 2012). Few, however, have focused on the diversity and genetic structure in chile peppers as a result of human management (González-Jara et al., 2011). Furthermore, in southern Mexico, a region of great cultural diversity, chile peppers are grown in many environments and cultivation systems—an optimal area for preserving the genetic resources of Mexican chile peppers (Latournerie et al., 2001; Aguilar-Meléndez et al., 2009; González-Jara et al., 2011; Long-Solís, 2012). Yet, few genetic and domestication studies have employed samples from southern areas of Mexico (but see, González-Jara et al., 2011; Kraft et al., 2014; Toledo-Aguilar et al., 2016; Taitano et al., 2019) or accounted for environment and management factors related to their origin. Thus, this region's landrace genetic diversity remains relatively unexplored and an ongoing study of this important species from a number of perspectives is warranted.

Within southern Mexico, the state of Oaxaca is characterized by extensive cultural and biological diversity (García-Mendoza et al., 2004). In this state, the long and

continuous history of highly complex interactions between human societies and nature (Caballero et al., 2004) has given rise to diverse management systems of wild and semiwild plants, and ongoing incipient plant domestication (Casas et al., 1997, 2001, 2016). The resulting diversity is especially notable for the case of *C. annuum* cultivars: Although Oaxaca contributes only about 0.3% of Mexico's chile production (SIAP, 2018), it harbors approximately 28 of the more than 60 major landraces of *C. annuum* found in Mexico (Aguilar-Rincón et al., 2010). Some of these landraces are endemic to this state and associated with specific cultural practices and different production systems that include backyards, campesino polyculture systems such as milpas, and monocultures. Interestingly, *C. annuum* var. *glabriusculum* semiwild and wild populations (found within the understory of semideciduous rainforests) also display a great diversity of local uses as well as biological and economic relevance in the towns and villages. Because the plant is cultivated, it is important to consider both the social and environmental landscapes that *C. annuum* inhabit to better understand and preserve the evolutionary processes and genetic diversity found in wild and domesticated populations (Samberg et al., 2013). Moreover, diversity has been identified as key to preserving the ability of populations to adapt to climate change (Bellon et al., 2015; Mercer and Perales, 2019).

The Mexican state of Oaxaca is a striking example of such diversity ranging from the cold highlands where the Pasilla Mixe landrace is found to the hotter and drier areas of the Tehuantepec Isthmus where Gui'ña dani is cultivated. The diversity of management regimes in Oaxaca occurs in the context of a topographically, climatically, and ecologically diverse region. This diversity includes the seasonal tropical rainforest of the Coastal lowlands, the temperate Central Valleys (around 1600 m a.s.l.), and the Cañada region in the northernmost part of the state where intertropical semidesert prevails at medium to low elevations around 800 m a.s.l., among others (Figure 1).

Previous genetic studies have used samples from the state of Oaxaca, but included only one wild and one let-standing populations (González-Jara et al., 2011), few landraces (Toledo-Aguilar et al., 2016), or both wild samples and four main landraces across the state (Taitano et al., 2019). A more extensive sampling across the domestication gradient and an inclusion of many more Oaxacan landraces has been needed. In sum, Oaxaca offers an extraordinary opportunity to study the joint effect of domestication, management, and environmental conditions on the structuring of genetic diversity.

Thus, here we comprehensively sampled chile peppers in the state of Oaxaca to analyze how *C. annuum* genetic diversity is structured in the state. One strength of this study is that we assessed diversity in a structured sampling of chile peppers across a domestication gradient—wild, semiwild, and domesticated populations—and across management regimes and environmental conditions in three contrasting regions. We specifically asked how these factors participate in structuring genetic diversity of chile pepper from this region of Mexico to better understand the roles of social, biotic, and abiotic factors in the evolutionary history of this important crop. Our study could then take into account direct and indirect human intervention on chile pepper genetic diversity, including the establishment of different kinds of anthropogenic environments with their corresponding effects on the ecological dynamics of pollinators and dispersers that can affect gene flow patterns. We hypothesized that differences in elevation between regions would be a factor of genetic differentiation coupled with the domestication process which, among other changes implied an expansion in the elevation range in which the cultivated chiles grow.

MATERIALS AND METHODS

Plant material collection

We collected plant material (leaves and fruits whenever they were available) from chile pepper populations across a wide environmental gradient including different management regimes and domestication levels. Between 2013 and 2016, we sampled 32 locations in the state of Oaxaca, Mexico. We grouped closely located samples (<9 km of distance) as single population, resulting in a total of 341 individuals from 28 populations (Table 1, Figure 1). The 9-km threshold to group populations was established considering the areas where farmers had reported to have commonly exchanged seeds; it is consistent with previous work suggesting domesticated chiles in Mexico appear to be exchanged by farmers across a narrow geographic range (Aguilar-Meléndez et al., 2009). Our collections included mainly landraces and wild individuals of *C. annuum*, as well as some individuals of the closely related *C. frutescens* and *C. chinense* species (Tables 1 and 2). Geographically, we sampled across two transects (Figure 1): 14 populations (175 individuals) from a mainly east–west transect covering the Oaxacan coast and 14 populations (166 individuals) along a mostly south to north transect from the middle of the Oaxacan coast at Pochutla to the Central Valleys region (12 populations, 128 individuals) and La Cañada region (2 populations, 38 individuals). The first transect along the coast presents a warm-subhumid climate and covers elevations from 0 to 600 m a.s.l., while the second transect that included Cañada (800 m a.s.l.) and Central Valleys (between 1300 and 1700 m a.s.l.) regions encompasses diverse climates from dry/semidry and warm-humid to temperate-humid and subhumid (García-Mendoza et al., 2004).

The specific collection locations were chosen following the advice of local farmers and families, key informants who directed us where we might find chile pepper samples. In total, we collected 18 named landraces (indicated by capitalized names) of *C. annuum* in polyculture (*milpas*) and monoculture fields, while semiwild and wild individuals

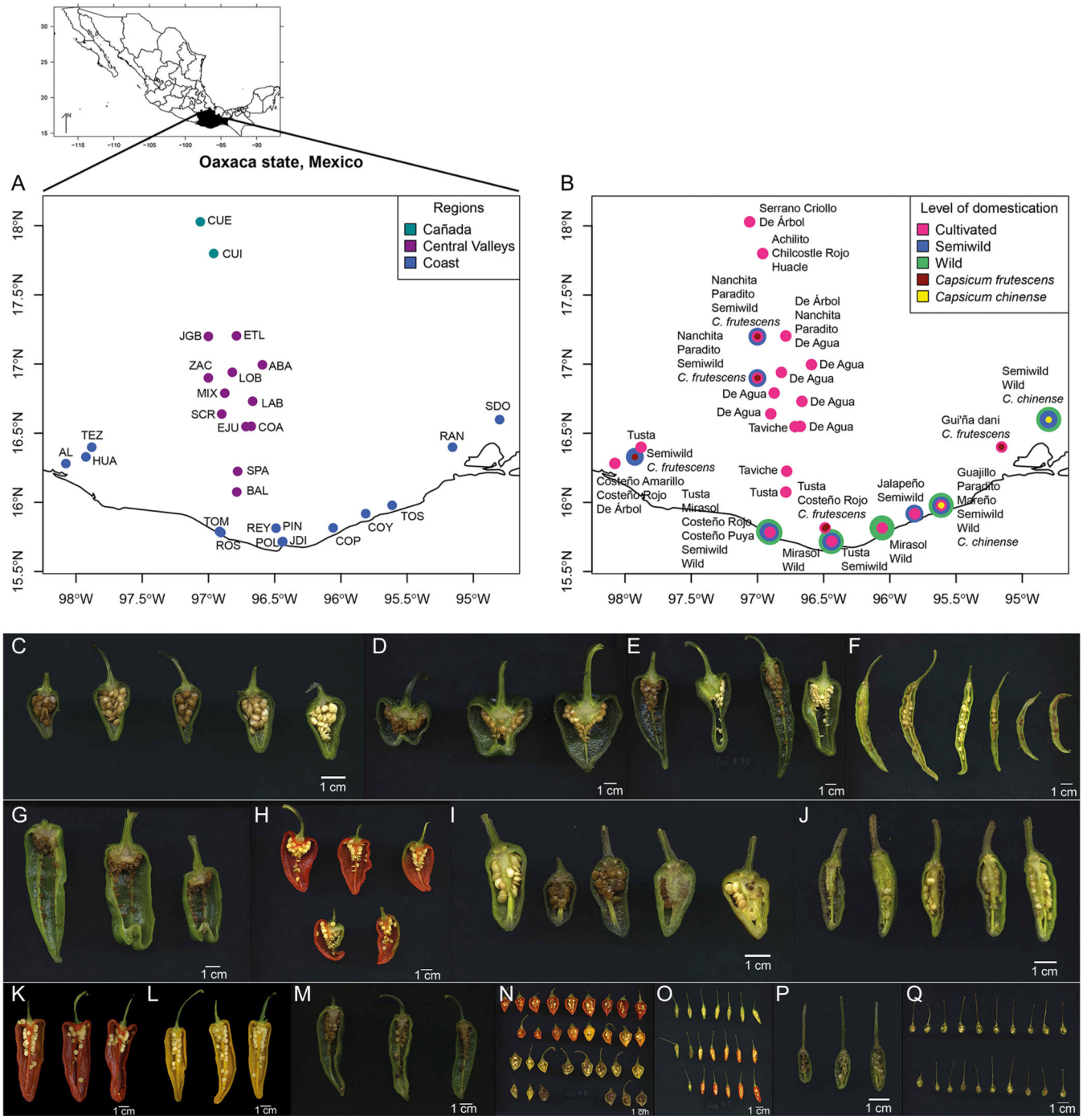

FIGURE 1 *Capsicum* sampling sites in Oaxaca, Mexico. (A, B) Wild, semiwild and landraces (cultivated) plants of *Capsicum annuum* were collected, along with some individuals of *Capsicum frutescens* and *Capsicum chinense* species. For a given locality sometimes *C. annuum* plants of all three domestication levels and/or the other species were sampled. Transverse section of representative fruits of each *C. annuum* landrace sampled in Oaxaca: (C) Achilito, (D) Huacle, (E) Chilcostle Rojo, (F) De Árbol (Cañada region); (G) de Agua, (H) Taviche, (I) Nanchita (J), Paradito (Central Valleys); (K) Costeño Rojo, (L) Costeño Amarillo, (M) Gui'ña dani, (N) Tusta, and (O) Mirasol landraces (coastal region). Also, (P) Piquín growing semiwild in some Oaxaca localities and (Q) Wild fruits (*C. annuum* var. *glabriusculum*). Chile pepper photos by Cristina Alonso.

were collected from backyards and remnants of tropical forest, respectively (Table 2). Additionally, we collected some accessions from the local markets for which no plants were available during our field trips. We also collected some samples of *C. frutescens* and *C. chinense* because they were growing very close to or with *C. annuum* individuals, mainly in backyard systems (Tables 1 and 2); these samples were included in the study to have a point of comparison for the shared genetic diversity with *C. annuum* populations and its overall genetic similarity/dissimilarity. For each

TABLE 1 Sampled *Capsicum* locations in the Mexican state of Oaxaca, including wild, semiwild and landrace (cultivated) individuals of *C. annuum*, *C. frutescens*, and *C. chinense* individuals (see map, Figure 1).

Location	ID	Region	Latitude	Longitude	Altitude	Named landrace or type[a]
Abasolo	ABA	Central Valleys	17.00	−96.59	1578	de Agua
Baltazar	BAL	Central Valleys	16.08	−96.79	1014	Tusta
Coatecas Altas and Paraje Coatequillas	COA	Central Valleys	16.55	−96.68	1531	de Agua
Ejutla	EJU	Central Valleys	16.55	−96.72	1470	Taviche
La Labor	LAB	Central Valleys	16.73	−96.67	1540	de Agua
La Lobera	LOB	Central Valleys	16.94	−96.82	1675	de Agua
San Juan and Gabriel Bautista Guelache	JGB	Central Valleys	17.20	−97.00	1579	Nanchita, Paradito, semiwild, *C. frutescens*
San Pablo Coatlán	SPA	Central Valleys	16.23	−96.78	1451	Taviche
Santa Cruz Mixtepec	MIX	Central Valleys	16.79	−96.88	1543	de Agua
Santa Cruz Nexila	SCR	Central Valleys	16.64	−96.90	1461	de Agua
Villa de Etla	ETL	Central Valleys	17.21	−96.79	1671	de Árbol, Nanchita, Paradito, de Agua
Zaachila	ZAC	Central Valleys	16.90	−97.00	1514	Nanchita, Paradito, semiwild, *C. frutescens*
Los Cues	CUE	Cañada	18.03	−97.06	776	de Árbol, Serrano Criollo
Cuicatlán	CUI	Cañada	17.80	−96.96	648	Achilito, Chilcostle Rojo, Huacle
Añil and Lagartero	AL	Coast	16.28	−98.08	153	de Árbol, Costeño Rojo, Costeño Amarillo
Copalita	COP	Coast	15.82	−96.06	58	Mirasol, wild
Coyul	COY	Coast	15.92	−95.81	36	Jalapeño, semiwild
Huaxpaltepec	HUA	Coast	16.33	−97.93	226	Semiwild, *C. frutescens*
Los Reyes	REY	Coast	15.82	−96.49	231	Costeño Rojo, Tusta, *C. frutescens*
Paso de los Indios and Lagunilla	PIN	Coast	15.83	−96.48	204	*C. frutescens*
El Polvorín	POL	Coast	15.72	−96.45	144	Mirasol, semiwild, wild
Juan Diegal	JDI	Coast	15.72	−96.43	115	Tusta, semiwild
Rancho Llano	RAN	Coast	16.40	−95.16	25	Gui'ña dani, *C. frutescens*
Rosedal	ROS	Coast	15.78	−96.91	23	Costeño Rojo, Costeño Puya, semiwild, wild
El Tomatal	TOM	Coast	15.79	−96.92	29	Tusta, Mirasol, semiwild
Santo Domingo del Ingenio	SDO	Coast	16.60	−94.80	48	Semiwild, wild, *C. chinense*
San Lorenzo Tezoluca	TEZ	Coast	16.40	−97.88	254	Tusta
La Tortolita and Saachilac	TOS	Coast	15.98	−95.61	50	Guajillo, Paradito, Mareño, semiwild, wild, *C. chinense*

[a]Landraces of *C. annuum* were found in each population. Wild and semiwild samples are indicated, and other species of *Capsicum* are specified.

population, our collections ranged from three to 33 individuals with a mean of 12.2 individuals per population. When arranged by domestication level (landraces, semiwild, and wild), our collection comprised 261, 33, and 23 individuals per group, respectively.

We evaluated the genetic diversity of chile peppers in Oaxaca, both in the context of a diverse topography and across a domestication gradient. We performed our analyses along those two axes. First, we analyzed all 341 individuals collected—322 individuals of *C. annuum*, 16 individuals of *C. frutescens* and 3 individuals of *C. chinense*—to evaluate the geographical distribution of genetic variation across the two transects with varying altitude and humidity conditions. In this case, we were able to find different semiwild and wild *C. annuum* landraces, as well as *C. frutescens* and *C. chinense* individuals within the same population. In the second approach, we evaluated genetic diversity within *C. annuum* individuals grouped by domestication level. To do so, we defined landraces as cultivated plants, distinguished by the farmers as such because of their characteristic

TABLE 2 Landraces, semiwild, and wild *Capsicum annuum* populations collected in Oaxaca state. Data include 16 individuals of *C. frutescens* and three individuals of *C.chinense* collected. For genetic analyses by landraces, Mareño, Jalapeño, Costeño Puya, and Guajillo were omitted because the number of comparable loci and the number of individuals were too low when analyzed as landrace. Location ID follows that described in Table 1.

***Capsicum annuum* local names, domestication categories, and congeneric species**	**Location ID**
Landraces	
de Agua	ETL, ABA, LOB, LAB, MIX, SCR, COA
Costeño Rojo	ROS, AL, REY
Tusta	TEZ, BAL, TOM, REY, JDI
Taviche	EJU, SPA
Gui'ña dani	RAN
Huacle (Red, Black, and Yellow)	CUI
de Árbol	CUE, ETL, AL
Paradito	ETL, ZAC JGB, TOS
Serrano Criollo	CUE
Nanchita	ETL, ZAC, JGB
Chilcostle Rojo	CUI
Mirasol	TOM, POL, COP
Achilito	CUI
Costeño Amarillo	AL
Mareño	TOS
Costeño Puya	ROS
Guajillo	TOS
Jalapeño	COY
Semiwild	JGB, ZAC, HUA, ROS, TOM, JDI, POL, COY, TOS, SDO
Wild	ROS, TOM, POL, COP, TOS, SDO
C. frutescens	RAN, REY, ZAC, HUA, JGB, PIN
C. chinense	TOS, SDO

morphological traits. We defined semiwild plants as morphologically similar to *C. annuum* var. *glabriusculum* wild chiles (perennial shrubs with small flowers with exerted stigma and small, erect, dehiscent fruits) but growing spontaneously in human-managed spaces, such as backyard and home gardens or as volunteers in nearby plots cultivated with other crops such as maize. Using the terminology of Aguilar-Meléndez et al. (2009), we use the term semiwild over weedy to recognize the putative origin of plants from the *C. annuum* var. *glabriusculum* wild populations. Last, we considered wild plants as *C. annuum* var. *glabriusculum* growing below the canopy of deciduous rainforests. Of the 322 individuals of *C. annuum* used for geographical analysis, we used 317 individuals to analyze the domestication gradient. We omitted four landraces for which we only obtained genetic information from a single individual (Costeño Puya, Guajillo, Jalapeño, and Mareño), keeping 14 landraces (261 cultivated individuals) and 33 semiwild and 23 wild individuals for this analysis.

In populations lacking suitable leaves, we collected fruits from senescent plants the farmers still had in the plots, totaling 37 individual seed samples of landraces Abasolo (9, Chile de Agua), San Baltazar Loxicha (5, Tusta), San Pablo Coatlán (4, Taviche), and Los Cues (19, Serrano and Jalapeño). We germinated those seeds in a greenhouse at the Institute of Ecology, at the National Autonomous University of Mexico (UNAM) in Mexico City, watering them every 72 h. From each maternal plant, just one seedling was kept for genotyping. The remaining samples came from leaves directly collected in the field and stored at −70°C. Thus, we worked with 317 individual samples among which we had no full siblings. We extracted DNA from frozen (if leaves were collected in the field) or fresh (if plants were grown in the greenhouse) leaf tissue, using a CTAB miniprep protocol (Doyle and Doyle, 1987) modified by Vázquez-Lobo et al. (1996) and stored DNA at −20°C.

Microsatellite genotyping

We used a set of 10 short sequence repeat (SSR) primers selected from those previously reported by Shirasawa et al. (2013; eight loci) and Nagy et al. (2007; two loci). These microsatellite loci primers are specific for *C. annuum* and included CaES3862, CaES1003, CaES0425, CaES4787, CaES2332, CaES4192, CaES5392, CaES4584 (Shirasawa et al., 2013), EPMS310, and EPMS386 (Nagy et al., 2007). These markers are found on 8 of the 12 *C. annuum* chromosomes. Each of the 10 reverse primers were labeled with one of the following fluorescent dyes PET, FAM, VIC, or NED.

We performed PCRs in two reaction profiles as follows. First, the eight loci reported by Shirasawa et al., (2013) were amplified from 1 ng genomic DNA. The 25-μL reaction mixture contained 5 μL PCR buffer 1×, 3 μL $MgCl_2$ (3 mM), 0.08 μL Taq polymerase (0.4 U/25 μl), 2 μL dNTPs (0.8 mM), and 0.75 μL of each primer (0.3 μM) in H_2O. The thermal cycling conditions were modified with a touchdown in annealing temperature (Sato et al., 2005) as follows: 3 min at 94°C for initial denaturation; three cycles of 30 s denaturation at 94°C, 30 s annealing at 66°C and 30 s extension at 72°C; a touchdown phase of 20 cycles with 30 s denaturation at 94°C, 30 s annealing with initial temperature of 66°C; and then the annealing temperature was decreased by 1°C every 2 cycles until an annealing temperature of 56°C was reached and used for the final 10 cycles; every cycle had a 30 s extension at 72°C. Second, PCR amplification of the two loci from Nagy et al. (2007) were carried out in a 25-μL final volume with 1 ng genomic

DNA. The reaction mixture contained 5 µL PCR buffer 1×, 2.5 µl $MgCl_2$ (2.5 mM), 0.08 µL Taq polymerase (0.4 U/25 µL), 2 µL dNTPs (0.8 mM), 0.5 µL of each primer (0.3 µM) and 9.42 µL H_2O. PCR conditions were initial denaturation for 3 min at 94°C; 35 cycles of 30 s denaturation at 94°C, 30 s annealing at 60°C, and 30 s extension at 72°C; then a 2 min final extension at 72°C. PCR products were analyzed with a capillary sequencer at the UIUC Core Sequencing Facility at the University of Illinois Urbana-Champaign (Urbana, IL, USA; https://unicorn.biotech.illinois.edu/). The size of alleles was determined using Peak Scanner v1.0 Software (Applied Biosystems, Waltham, MA, USA).

Genetic diversity analyses

To characterize each marker's variability, for each locus, we estimated genetic diversity indexes using GENEPOP version 4.6 (Rousset, 2008). We ran Hardy–Weinberg tests using 10,000 Markov chain dememorizations, 20 batches and 5000 iterations per batch. The linkage disequilibrium test used 10,000 dememorizations, 1000 batches and 5000 iterations per batch (Rousset, 2008). Additionally, we calculated the frequency of null alleles and F_{ST} values with confidence intervals using FreeNa software (Chapuis and Estoup, 2007); we used 1000 replicates to evaluate whether it was necessary to exclude some loci.

To evaluate the genetic variation along elevation and domestication gradients, we used Arlequin version 3.5.2.2 (Excoffier et al., 2005) to calculate the mean number of sampled alleles (N), unbiased expected heterozygosity (H_E), observed heterozygosity (H_O), genetic differentiation estimator R_{ST} (Slatkin, 1995), and the reduction in effective population size (M index, Excoffier and Lischer, 2010; Peery et al., 2012), while inbreeding coefficient (R_{IS}) was calculated with GENEPOP version 4.6 (Rousset, 2008). After checking for normality and homogeneity of variances through Shapiro–Wilks and Bartlett's tests (Appendix S1), we performed an analysis of variance (ANOVA) using the R (version 3.4.2) package RcmdrMisc (Fox et al., 2020) to assess whether the mean of H_E changed significantly among geographical regions (Cañada, coastal, and Central Valleys). Additionally, allelic richness for each region was obtained through a rarefaction method implemented using ADZE version 1.0 (Szpiech et al., 2008) to incorporate sample size differences. Subsequently, we carried out an analysis of molecular variance (AMOVA) with Arlequin (Excoffier et al., 2005) using coastal, Cañada and Central Valleys regions (groups) as categories of the elevation gradient to analyze the hierarchical structure of genetic diversity among and within geographical regions and locations, including *C. frutescens* and *C. chinense* individuals. In a separate analysis, we carried out an AMOVA grouping only *C. annuum* individuals by domestication level, clustering the 14 landraces of chile peppers as "cultivated", all the let-stand populations as "semiwild", and all the wild populations as "wild".

To check for an isolation by distance (IBD) pattern of genetic diversity, we assessed the relationship between geographic and genetic distance matrices using Mantel tests in XLSTAT PLUS 2017 in Excel (Microsoft) with 10,000 permutations to compare locations. Paired geographic distances between populations were calculated using the R package ggmap (Kahle and Wickham, 2013). Geographical distance values were log-transformed to fit a two-dimension stepping-stone migration model. Correlation matrices of paired R_{ST} values were plotted with the R package corrplot (Wei et al., 2017) to evaluate genetic differentiation between pairs of populations or landraces.

For our AMOVAs and IBD tests described above, we used R_{ST} (Slatkin, 1995) because it is an adequate measure of genetic structure under the microsatellite stepwise mutation model. R_{ST} is defined as $R_{ST} = S - SW/S$, where S is the average squared difference in allele size between all pairs of alleles, and SW, the average sum of squares of the differences in allele size within each subpopulation (Slatkin, 1995; Balloux and Lugon-Moulin, 2002).

Population graph analysis

To look for patterns of genetic similarity among chile pepper landraces, we evaluated the genetic connectivity among locations and subsequently among landraces through population graph analyses using R packages popgraph (Dyer, 2014), gstudio (Dyer, 2012) and igraph (Csardi, 2013). The package popgraph is used to evaluate the connectedness of populations based on their genetic covariances (Dyer and Nason, 2004; Gotelli and Stanton-Geddes, 2015). The resulting network graph represents the populations as nodes (locations or landraces for each one of our analyses), with node size reflecting within-population genetic variance. Nodes are connected by "edges" (lines) whose length is inversely proportional to the genetic covariance between a pair of nodes, thus reflecting the among-population component of genetic variation (Dyer and Nason, 2004; Oyama et al., 2016). Only 26 locations and 12 landraces of *C. annuum* and one *C. frutescens* populations were included in this analysis; locations and landraces with less than four individuals cannot be analyzed by the program because their within-population variance and genetic covariances cannot be confidently assessed (Dyer and Nason, 2004).

Population structure analyses

Population structure was further explored using two complementary approaches. We used the Bayesian clustering program STRUCTURE version 2.3.4 (Pritchard et al., 2000) to infer genetic population structure by probabilistically assigning individuals to clusters, each one characterized by distinct allele frequencies as well as identifying the number (K) of clusters with maximum likelihood (Pritchard et al., 2000). We used an admixed model assuming correlation among allele frequencies, enabling us to assign each individual to more than one cluster with a specific

probability. Ten runs were considered for each value of K ranging from $K = 1$ to $K = 10$. For each run, we used 250,000 burn-in chains and 1,000,000 Markov chain Monte Carlo iterations. We performed the Evanno test (Evanno et al., 2005) using STRUCTURE HARVESTER (Earl and von Holdt, 2012) to infer the optimal K-value. Additionally, we conducted a discriminant analysis of principal components (DAPC) to identify genetic differentiation between groups through a multivariate method that is robust to linkage issues. This analysis performs an analysis of principal components previous to a discriminant analysis that clusters the data to optimize the variance between groups while minimizing the variance within groups (Jombart et al., 2010) and ensures that variables submitted to DA are uncorrelated without losing genetic information. We defined groups by population, geographic region, landrace and level of domestication, using R package adegenet (Jombart, 2008). We performed two separate analyses. The first one included all 341 individuals coming from 28 geographical populations, while for the second one we used 14 landraces, as well as semiwild, wild, *C. frutescens* and *C. chinense* groups ($n = 336$ individuals) with five individuals omitted from the domestication analysis due to the lack of enough individuals ($n \leq 2$) to group them as a landrace.

RESULTS

Genetic diversity

The diversity patterns of the 10 SSR loci in 341 individuals revealed a variable number of alleles, ranging from 3 to 21 distinct alleles per locus; H_E was highest at the EPMS386 locus ($H_E = 0.66$) and lowest in the CaES4584 locus ($H_E = 0.05$). Locus-by-locus analysis showed that all 10 loci displayed significant deviations from Hardy–Weinberg equilibrium ($p < 0.05$), concomitant with high and positive R_{IS} values (Table 3). In total, we found 36 private alleles, with the highest value in EPMS386 (Pa = 11), and no private alleles in CaES2332 and CaES4192. Moreover, we detected at least one private allele in the AL, JDI, SDO, RAN, CUI, MIX, ETL populations, as in the Costeño Amarillo, Costeño Rojo, de Agua, Tusta and Huacle landraces and in the semiwild and wild groups (see Table 4; Appendix S2).

Null alleles were detected in EPMS386 and EPMS310 in 23 and 20% of the individuals. Loci with null allele frequencies greater than 0.20 are routinely removed from microsatellite analyses (Dakin and Avise, 2004; Chapuis and Estoup, 2007); nevertheless, the estimating null allele (ENA) correction test indicated that the presence of null alleles did not affect the significance of F_{ST} estimations, as we obtained F_{ST} without ENA = 0.23 (0.13–0.29 confidence interval) and F_{ST} with ENA = 0.21 (0.15–0.26 confidence interval; for data by population and landrace, see Appendix S3). The linkage disequilibrium test detected significant values in five pairs of loci. These pairs were CaES4787–CaES1003 ($P = 0.0004$), CaES4787–CaES3868 ($P = 0.0198$), CaES1003–CaES3862 ($P = 0.0011$), CaES4584–CaES5392 ($P = 0.0040$) and EPMS386–EPMS310 ($P = 0.0104$). Among these, only CaES1003 and CaES3862 are located on the same chromosome (for data by population and landrace see Appendix S4).

We analyzed the effect of geography on the distribution of genetic diversity. We found slightly higher levels of genetic diversity in the coastal region ($H_E = 0.45$, SD = 0.11) in comparison to Central Valleys ($H_E = 0.41$, SD = 0.10) and La Cañada regions ($H_E = 0.41$, SD = 0.06; Appendix S2). However, genetic diversity H_E and H_O values were not significantly different between regions (ANOVA; H_E, $F_{2,\ 25} = 0.57$, $p = 0.57$; H_O, $F_{2,\ 25} = 0.01$ and $P = 0.99$; Appendix S1). Allelic richness calculated by rarefaction for the coastal, Central Valleys, and Cañada

TABLE 3 Genetic diversity analysis statistics and Hardy–Weinberg tests on 10 microsatellites (SSRs) in 28 populations of Oaxaca state. EPMS310 and EPMS386 microsatellites were described by Nagy et al. (2007) and the other eight SSRs correspond to those of Shirasawa et al. (2013).

SSR	Motif	Size	Dye	C	*A*	Pa	An	H_E	R_{IS}	χ^2	df	*P*
CaES3862	GGA	209	PET	1	3	2	0.05	0.18	0.349	52.37	26	0.0016*
CaES1003	AATC	176	VIC	1	4	2	0.03	0.07	0.603	31.23	10	0.0005*
CaES0425	AAC	255	VIC	2	4	4	0.14	0.25	0.989	182.38	30	0.0000*
EPMS310	$(CAT)_{13}$	140-172	PET	3	14	1	0.20^	0.59	0.666	I	46	0.0000*
CaES4787	ACT	132	NED	3	6	4	0.04	0.11	0.647	34.93	18	0.0097*
CaES2332	AAG	160	PET	4	5	0	0.11	0.44	0.332	103.66	46	0*
CaES4192	ATC	166	PET	5	3	0	0.17	0.33	0.620	211.84	44	0.0000*
CaES5392	GGA	201	VIC	6	8	8	0.02	0.07	0.001	37.20	14	0.0007*
CaES4584	AAC	108	NED	7	4	4	0.02	0.05	0.554	20.91	8	0.0074*
EPMS386	$(CA)_{15}$	122-170	FAM	8	21	11	0.23^	0.66	0.711	I	56	0.0000*

Notes: C, chromosome number; *A*, number of alleles; Pa, private alleles; An, mean frequencies of null alleles; H_E, expected heterozygosis; R_{IS}, inbreeding coefficient; χ^2, F_{IS} estimates by Fisher's method; I = high values (infinite). Significance: *$P < 0.05$. ^High frequencies ≥ 0.20.

TABLE 4 Genetic diversity values by level of domestication and landraces.

Groups	*N*	*P*	*A*	*r*	Ar	Ae	Pa	H_O (SD)	H_E (SD)	*M*	R_{IS}
Level of domestication											
Cultivated	261	4	6.4	6.2	3.4	2.7	14	0.12 (0.12)	0.33 (0.31)	0.26	0.52
Semiwild	33	2	4.0	3.9	3.6	2.3	1	0.12 (0.11)	0.37 (0.23)	0.31	0.59
Wild	23	4	3.5	3.3	3.3	2.2	1	0.24 (0.15)	0.44 (0.26)	0.34	0.27
Mean		3	4.6	4.5	3.4	2.4	5.3	0.16 (0.13)	0.38 (0.27)	0.30	0.46
SD		1.2	1.6	1.6	0.2	0.3	7.6	0.07 (0.02)	0.06 (0.04)	0.04	0.17
Landraces											
Achilito	3	2	1.6	1.2	1.2	1.8	0	0.39 (0.35)	0.74 (0.18)	0.29	0.45
Costeño Amarillo	2	5	1.7	–	–	1.9	2	0.60 (0.22)	0.70 (0.18)	0.32	0.08
Mirasol	4	2	1.2	–	–	1.6	0	0.13 (0.18)	0.59 (0.23)	0.14	0.90
Chilcostle Rojo	4	4	1.6	1.5	1.3	1.9	0	0.15 (0.34)	0.55 (0.15)	0.43	0.99
Huacle	12	0	1.9	1.9	1.2	1.7	1	0.20 (0.14)	0.46 (0.27)	0.29	0.24
Paradito	10	3	2.4	2.4	1.3	1.9	0	0.22 (0.20)	0.39 (0.29)	0.27	0.26
Gui'ña dani	17	3	2.5	2.0	1.3	2.0	1	0.19 (0.19)	0.39 (0.22)	0.27	0.45
Serrano Criollo	9	2	1.8	–	–	1.6	0	0.26 (0.25)	0.39 (0.20)	0.37	0.08
Nanchita	8	3	1.8	1.8	1.2	1.6	0	0.21 (0.24)	0.37 (0.24)	0.30	0.10
Costeño Rojo	45	6	4.6	4.2	1.4	2.3	1	0.23 (0.24)	0.36 (0.28)	0.28	0.71
de Árbol	12	1	2.3	2.2	1.3	1.8	0	0.17 (0.17)	0.33 (0.16)	0.28	0.30
Taviche	21	3	2.4	2.3	1.2	1.9	0	0.05 (0.06)	0.28 (0.27)	0.34	0.43
de Agua	79	3	3.5	3.3	1.3	1.9	1	0.08 (0.11)	0.27 (0.23)	0.31	0.53
Tusta	35	1	3.3	3.0	1.2	1.6	1	0.09 (0.08)	0.21 (0.20)	0.24	0.16
Mean		3	2.3	–	–	1.8	0.6	0.21 (0.20)	0.43 (0.22)	0.30	0.41
SD		1.6	0.9	–	–	0.2	0.7	0.14 (0.08)	0.16 (0.04)	0.07	0.29

Notes: *N*, number of individuals; *P*, number of polymorphic loci with less than 5% missing data; *A*, mean number of alleles; *r*, allelic richness; Ar, mean allelic richness based on the minimum sample size (1 and 26 individuals for landraces and level of domestication respectively); Ae, mean number of effective alleles; Pa, number of private alleles; H_O, observed heterozygosis; H_E, expected heterozygosis; SD, standard deviation; *M*, index of Garza and Williamson (2001); R_{IS}, inbreeding coefficient.

regions was 6.6, 4.7, and 2.6, respectively (Appendix S5), while at the population level, the ROS (3.7), AL (3.6), JDI (2.9), and TOS (2.8) coastal populations had the highest values (Appendix S2).

Regarding analysis by level of domestication, we found slight differences among domestication categories (Table 4). H_E was estimated at 0.33 (SD = 0.31) for cultivated (including all landraces together); whereas semiwild and wild displayed $H_E = 0.37$ (SD = 0.23) and $H_E = 0.44$ (SD = 0.26) values respectively. When analyzing by landrace, we found the highest level of expected heterozygosity for Achilito ($H_E = 0.74$, SD = 0.18) and lowest for Tusta ($H_E = 0.21$, $SD = 0.20$). Highest allelic richness by rarefaction was found in the cultivated group (6.2) and for landraces in Costeño Rojo (4.2), semiwild (3.9), wild (3.3), de Agua (3.3), and Tusta (3.0) chile peppers.

Genetic structure

To further explore the geographic distribution of genetic diversity, we performed an AMOVA grouping the 22 populations in three different regions: Coast, Central Valleys, and Cañada. Results indicated that around 90% of genetic variation occurred within geographical populations (locations), irrespective of the inclusion of species *C. frutescens* and *C. chinense* in the samples (Appendix S6). When considering domestication level, the highest percentage of variation (92.97%) was found within populations, while the lowest percentage of variation (0.73%) occurred among populations within each domestication level (cultivated, semiwild, and wild groups; see Appendix S6). The total differentiation estimate by R_{ST} was 0.07 ($P = 0.2394$) by domestication level, and for geographic regions analysis we obtained $R_{ST} = 0.11$ ($P < 0.001$) including individuals of the

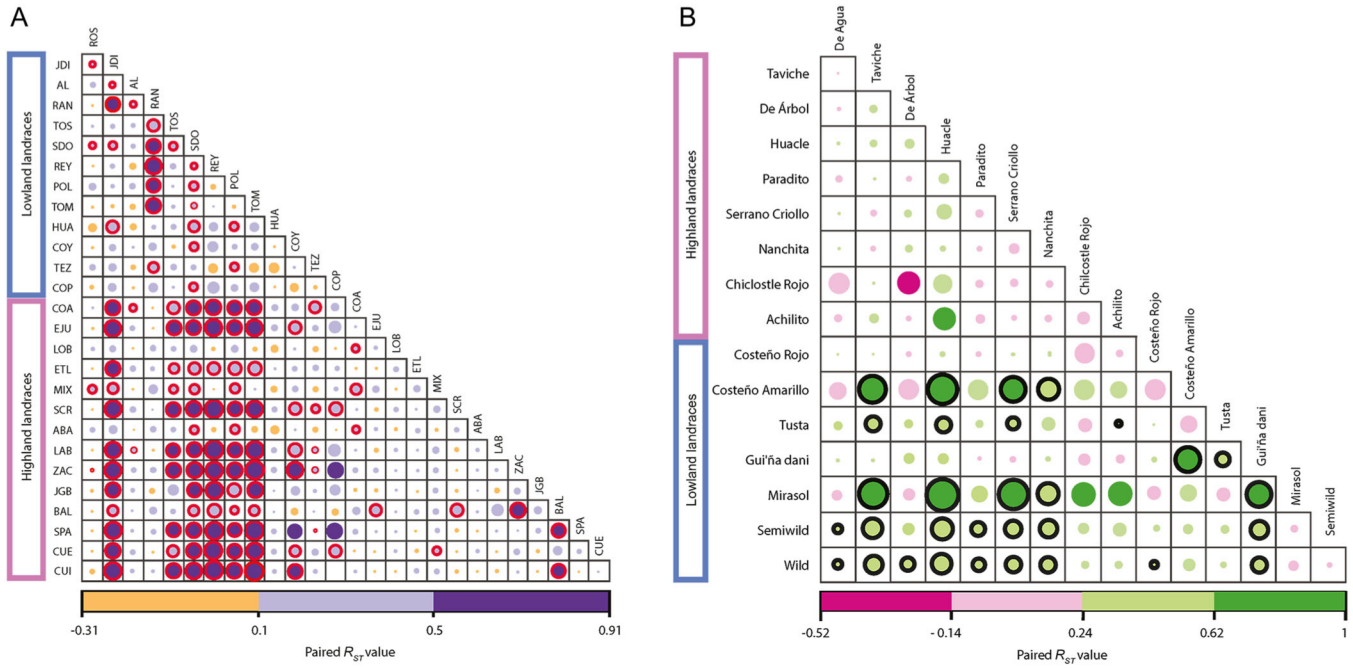

FIGURE 2 Heat maps showing paired R_{ST} values for (A) pairwise comparisons among Oaxacan localities where *Capsicum annuum* was sampled and (B) pairwise comparisons among pepper types; significant values of R_{ST} are encircled in red and black. The strongest genetic differentiation is found between coastal and highland populations. In turn, Mirasol landrace, semiwild, and wild types had high R_{ST} values when compared with highland landraces but little or no genetic structure with most of the lowland landraces (see Appendices S7, S8, S9, S10).

three collected *Capsicum* species (*C. annuum*–*C. frutescens*–*C. chinense*) and $R_{ST} = 0.09$ ($P < 0.001$) exclusively for *C. annuum* individuals by geographical level. In spite of the relatively homogeneous distribution of expected heterozygosity, inbreeding coefficients (R_{IS}) were lower in wild populations. Moreover, among all coastal populations, wild and semiwild populations had relatively low R_{ST} values with respect to cultivated populations, with the exception of the Gui'ña dani landrace (Figure 2; Appendices S7–S10). The strongest genetic differentiation was found between lowland (Coast) and highland (Cañada and Central Valleys) populations (Figure 2A). In turn, the Mirasol landrace and the semiwild and wild types had high R_{ST} values when compared with highland landraces but had little or no genetic structure with most of the lowland landraces (Figure 2B; Appendices S7–S10).

The role of geographic distance in structuring genetic variation was also evaluated through Mantel tests. These showed that in *C. annuum* populations genetic distance (R_{ST}) is positively but not significantly correlated with distance ($r = 0.03$, $P = 0.57$). Separate Mantel tests across each of the transects (coastal and through the highlands) also showed that the correlation between geographic and genetic distances was not significant (Appendix S11).

Population graph analyses

The population graphs showed high connectivity among sites and among landraces (Figure 3). The graph by location consisted of 26 nodes (each a location) and 47 edges. The interconnected architecture indicates tighter connections (shorter edges) among highland populations (Cañada and Central Valleys) (Figure 3A), which reflects a higher degree of genetic covariance among these populations.

The population graph for landraces, semiwild, wild, and *C. frutescens* populations was less complex, with a resulting topology of 15 nodes and 26 edges (*C. chinense* was excluded because the software requires more than three individuals per class). The wild, semiwild, and *C. frutescens* groups clustered to one side, while the domesticated landraces clustered together (Figure 3B). Clustering by geographical origin was also apparent in this network with landraces that formed clusters of covariation with other landraces cultivated in the same region (Figure 3B). For instance, with the exception of Chile de Agua, landraces from the Central Valleys were closely connected. Among lowland landraces, Gui'ña dani (30 m a.s.l.) shows high covariation both with Chile de Agua and Huacle pepper (1600 m a.s.l. and 900 m a.s.l., respectively). Interestingly, landraces Tusta, Mirasol, and de Agua were connected to semiwild and wild relatives as well as to *C. frutescens*; Tusta and Mirasol are sometimes found to be sympatric with wild and semiwild populations of *C. annuum* and semiwild populations of *C. frutescens*.

Structure analyses

The model-based genetic clustering algorithm obtained by STRUCTURE software provided further insight on the

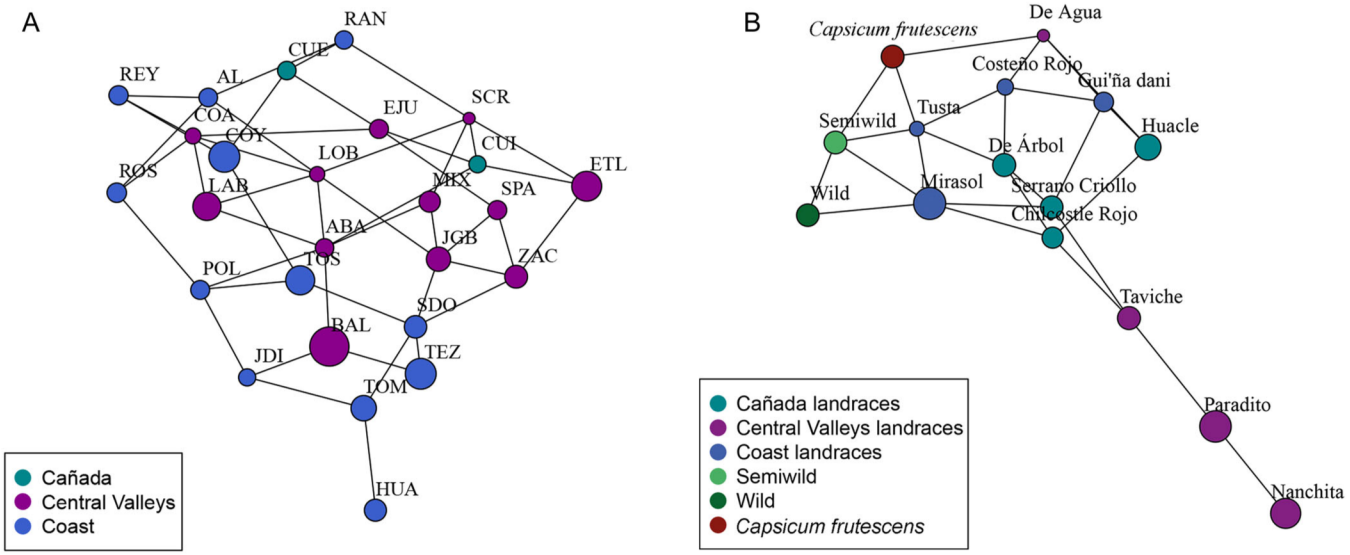

FIGURE 3 Population graph analyses (Dyer and Nason, 2004; Dyer, 2015) showing the genetic topography and connections (A) among 26 localities and (B) 12 landraces of *Capsicum annuum*, wild and semiwild individuals and *Capsicum frutescens* in Oaxaca state. Shorter edges indicate higher covariance between the nodes (i.e., localities or landraces), the size of node is proportional to the genetic variance within population or landrace.

differences among altitudinal and geographical areas. The Evanno test (Evanno et al., 2005; Appendix S12) resulted in a higher likelihood for two clusters ($K = 2$) in both runs (by location and by landrace). Nevertheless, here we present the assignment values per individual at $K = 2$, $K = 3$, and $K = 4$, because the pattern of cluster differentiation provides insight in the context of our additional analyses.

When the individuals were ordered by geographical location, at $K = 2$, the genetic assignment separated *C. annuum* into two clusters corresponding almost exactly to elevational ranges (Figure 4): one corresponding to highland populations (600–1700 m a.s.l.) and another to lowland populations (0–600 m). The exception was Rancho Llano (RAN), a Zapotec population that is mainly composed of Gui'ña dani pepper and whose individuals were assigned with a greater probability to the highland cluster. Similarly, most of the individuals of the Baltazar Loxicha (BAL) population, located in the transition zone between the coast and the Sierra Madre del Sur (1014 m a.s.l.) were assigned to the lowland cluster. At $K = 3$, we observed a more homogeneous cluster grouping mainly populations from the highlands, as well as RAN (see above) (Figure 4, middle row, in purple); coastal populations were subdivided into two clusters that had no clear geographical basis. At $K = 4$, the genetic assignation maintained a similar pattern, but a two-population substructure within highland populations became apparent (see Figure 4, for $K = 4$). Interestingly, the contrast between lowland and highland populations became less sharp as, like RAN, individuals from El Rosedal (ROS) were also assigned to one of the highlands clusters. These results are shown in an explicit geographic depiction in Figure 4B–D, a pie chart showing the probability of assignment to each cluster geographically. We detected a significant correlation ($P = 0.0001$, $r = -0.65$) between population elevation and the probability of genetic assignment to the lowland cluster (for $K = 2$), where elevation diminished significantly the probability to be assigned to the lowland cluster (Appendix S13).

When individuals were ordered by landrace (Figure 5), at $K = 2$ a pattern of differentiation by elevation was apparent again (highlands vs. lowlands); it is relevant to remember that most of the wild and semiwild samples came from low-elevation locations. In this case, only the Gui'ña dani landrace (cultivated in the lowlands of the Tehuantepec Isthmus) was assigned with high probability to the highland cluster. Individuals belonging to Costeño Rojo (lowlands landrace) were assigned to either the lowland or the highland cluster. At $K = 3$, the first genic pool (Figure 5, middle row, in purple) contained all landraces from the highlands, along with the Gui'ña dani landrace, some individuals of the Costeño Rojo landrace, and a few *C. frutescens*. Interestingly, some individuals of the Tusta landrace were assigned to a separate cluster, and not to the rest of the lowland landraces. At $K = 4$, the first 10 landraces had a mixed composition of individuals assigned to any of two highland clusters (still including the lowland Gui'ña dani landrace); the lowland Tusta landrace was assigned to a separate cluster (blue) shared with some semiwild and *C. frutescens* individuals, whereas Costeño Rojo individuals had assignment probabilities to all clusters. Wild and semiwild peppers maintained their affiliation to the lowland cluster (green, Figure 5, lower row).

DAPC analyses

The DAPC analyses showed additional footprints of the domestication gradient as a factor underlying the genetic

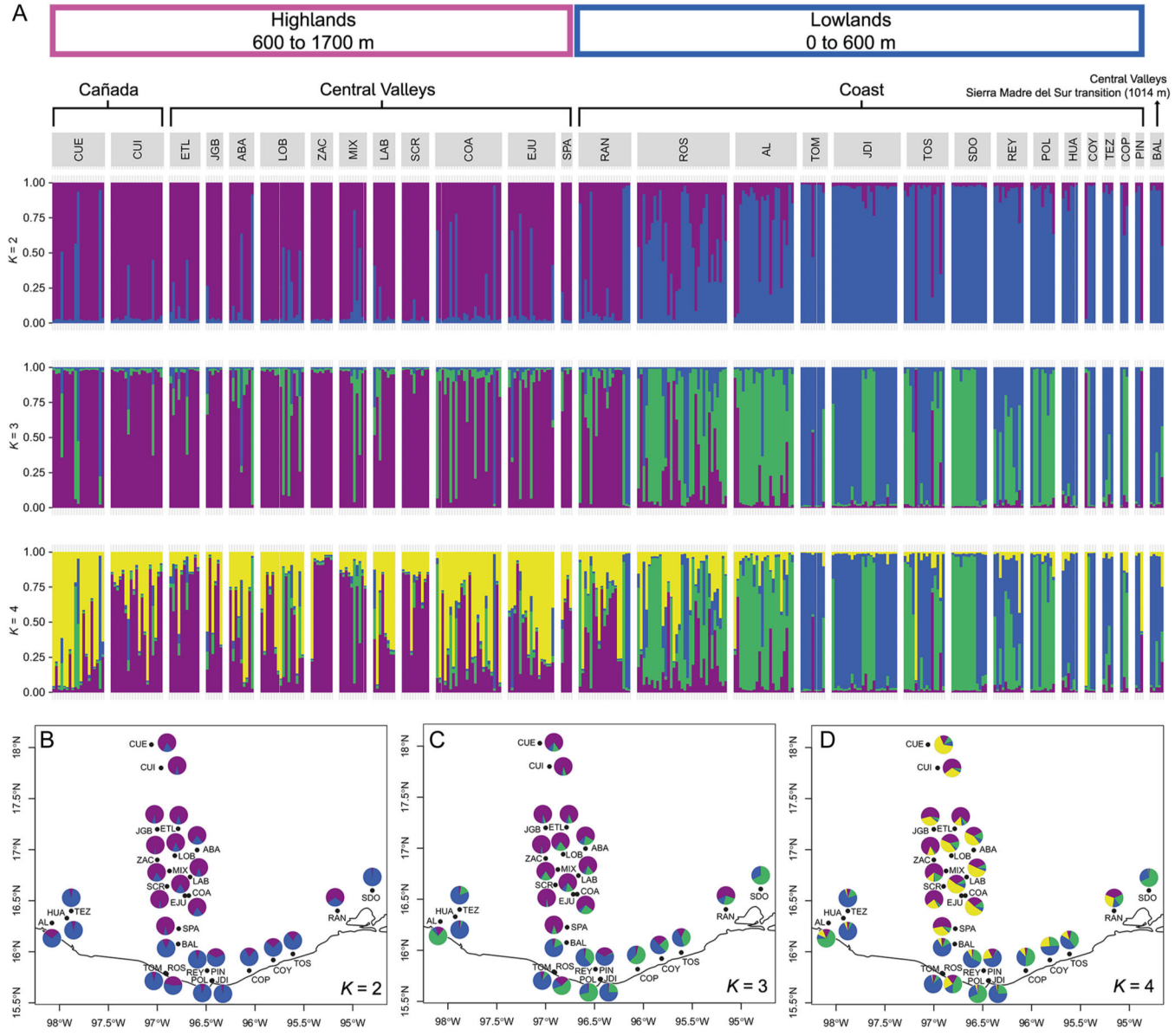

FIGURE 4 Probabilities of assignment of each individual obtained by STRUCTURE ordered by population. (A) We present $K = 2$, $K = 3$ and $K = 4$. In $K = 2$, group 1 clusters populations from the Cañada and Central Valleys regions and just one from lowlands (Rancho Llano, RAN), while group 2 encompasses coastal populations and one highland population (Baltazar, BAL). Maps of assignment probability to clusters obtained by STRUCTURE make visibly clear the genetic separation between highlands and lowlands. (A, B) At $K = 2$ localities of Central Valleys and Cañada were assigned to cluster 1, while localities in the coastal region were assigned to cluster 2, with the exception of Rancho Llano (RAN). (C) At $K = 3$ and (D) $K = 4$, an equivalent pattern was observed.

structure: wild, semiwild and cultivated (Figure 6). DAPC results showed an assignment probability (space of genetic variation represented by ellipses) of 96% to cultivated, 70% to semiwild, and 57% to wild groups (Appendix S14). Thus, the genetic spaces of the semiwild and wild peppers overlapped each other and also with that of cultivated peppers. For *C. frutescens* and *C. chinense*, the assignment probability was 0% because all its individuals were encompassed within the genetic variation space of *C. annuum* (Figure 6A).

When plotted points were grouped by landrace, the Tusta landrace segregated from the rest of the cultivated peppers, as 80% of Tusta individuals were exclusively within the confidence interval of the Tusta landrace (i.e., they had far less overlap with other landraces; see Figure 6B and Appendix S14). To a lesser extent, the same was observed for the Mirasol landrace, which is morphologically intermediate between domesticated and wild peppers (it has erect fruits, but they are 3–5 cm long, whereas the fruits of wild var. *glabriusculum* are 0.8–2 cm). Variation for the semiwild peppers strongly overlapped with the wild (*C. annuum* var. *glabriusculum*) and the cultivated types (for probability values and size of groups, see Appendix S14).

The DAPC was also performed for geographic regions (coastal, Central Valleys, and Cañada; Appendix S15). The

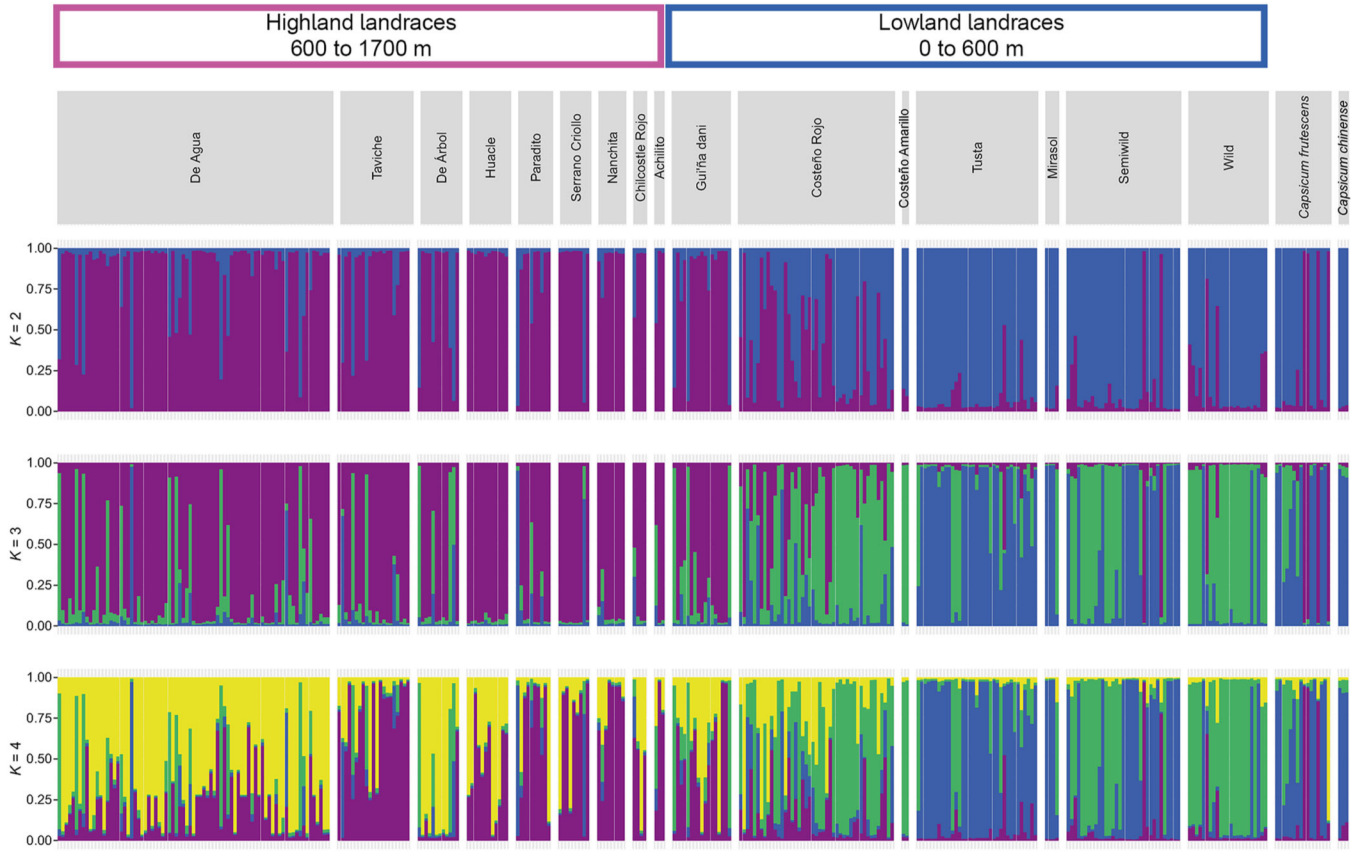

FIGURE 5 STRUCTURE results with individuals ordered by landrace and domestication level. We present $K = 2$, $K = 3$ and $K = 4$. The inferred genetic structure of landraces coincides with landrace-growing altitudes dividing them into lowland and highland landraces. However, Gui'ña dani landrace was grouped with highland landraces and at $K = 4$ Tusta was differentiated from the other landraces, with some individuals being assigned to a cluster with semiwild and *Capsicum frutescens* peppers (lower row, blue color).

results seem to agree with those of the STRUCTURE analysis since they showed evidence of some differentiation among highlands and lowlands populations. The space of genetic variation corresponding to Central Valleys and the coast covered 87% and 82%, respectively, of the individuals collected in those zones (Appendix S14). Meanwhile, variance space of the Cañada region retained 20 of 38 individuals with an assignment probability of 32% (assignment probabilities of the three DAPC graphs are in Appendix S16).

DISCUSSION

Our results highlight the effect of elevation and domestication gradients in shaping the genetic structure of *C. annuum* populations as genetic clustering correspond to elevation zones. Nevertheless, exceptions to these patterns illustrate how human processes—such as historical human migrations—may influence genetic structure. The different approaches we used consistently highlighted the gene flow between wild and domesticated populations, especially in the coastal region where semiwild populations also appear to be involved likely through backyard and small-holder cultivation contexts. By clarifying the effects of natural and human-mediated processes on this complex array of interacting species and domestication levels, we can better facilitate effective in situ conservation of these important genetic resources.

Patterns of genetic diversity

Genetic diversity patterns in *C. annuum* populations are defined by elevation and the level of domestication (wild, semiwild, and cultivated). Although genetic diversity values were not significantly different between a priori defined regions (Central Valleys, Cañada and coastal); allelic richness, STRUCTURE, pairwise F_{ST}, and population graph analyses suggested an altitudinal pattern. While one could argue that such pattern could be biased by our lack of wild samples from the highland populations, we found that this pattern is congruent with the area where the wild relative of domesticated peppers (*C. annuum* var. *glabriusculum*) grow. Wild chile peppers are rarely found at high altitudes, whereas cultivated chiles seem to have followed an altitudinal expansion from their native range (Pickersgill, 2016). Our results are consistent with previous

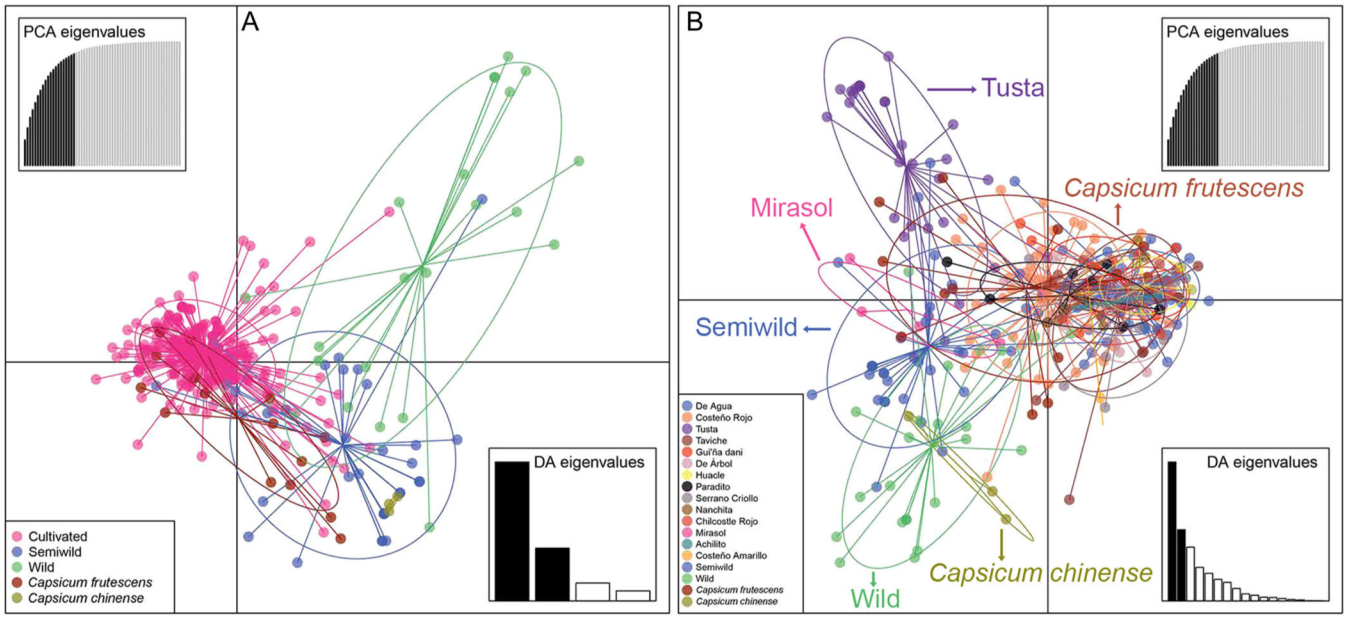

FIGURE 6 Discriminant analysis of principal components (DAPC). These graphs represent the individuals as dots and the groups as ellipses. (A) By level of domestication where each category is differentiated and *Capsicum frutescens* individuals are mixed with individuals of *Capsicum annuum* unlike *Capsicum chinense* individuals. (B) DAPC by landrace shows that Tusta and Mirasol landraces (located in the Coast of Oaxaca) were differentiated of other *C. annuum* landraces. Top boxes indicate number of retained principal components and bottom boxes show retained discriminant analysis eigenvalues used to run analyses. We performed DAPC clustering by geographical region and retrieved the same distinction between lowlands and highlands already evident from STRUCTURE (see Appendix S15).

findings, which have highlighted the relevance of the lowland populations in *C. annuum* domestication (Aguilar-Meléndez et al., 2009; González-Jara et al., 2011; Kraft et al., 2014). Moreover, this pattern is also similar to that found by Toledo-Aguilar and collaborators (2016), who analyzed 24 microsatellite loci in chile Ancho cultivars finding that higher levels of H_E are associated with low elevations. It is possible that at least part of the differentiation between lowland and highland clusters was also reinforced by domestication. Domestication involved an expansion in the elevation range of the species towards areas where wild populations of *C. annuum* var. *glabrisculum* are scarce or nonexistent. Thus, interaction of the domestication and altitudinal gradients may well be generating the patterns we found.

The domestication gradient as studied here, shows that wild peppers have slightly higher levels of genetic diversity than cultivated and semiwild peppers. However, some landraces presented higher genetic diversity values than the wild populations, which suggests that there has not been an important loss of genetic variation in these cultivated peppers. This pattern has been reported in other areas of Mexico, like the Yucatan Peninsula for chile pepper (Aguilar-Meléndez et al., 2009) and for ayocote bean (*Phaseolus coccineous*) in central and southern Mexico (Guerra-García et al., 2017). One possible explanation for this pattern is that the high-diversity landraces (Achilito, Costeño Amarillo, Mirasol, and Chilcostle) have had either more stable and larger population sizes and/or have been enriched by alleles coming from both wild and cultivated gene pools, coupled with the comparatively low number of individuals found in each wild population (six on average), which would lead to stronger effects of genetic drift. Another factor could be the proximity of wild peppers in the lowlands where Costeño Amarillo and Mirasol are grown (the latter even being collected side by side with the semiwild and wild *C. annuum* var. *glabriusculum*), which would facilitate the introgression from wild to cultivated germplasm.

Inbreeding coefficients (R_{IS}) in this study suggest a higher frequency of self-pollination among peppers growing in anthropogenic environments (both cultivated and semiwild peppers), whereas wild populations appear to outcross more frequently. The inferred difference in self-pollination may be partially explained by the prevalence of exerted stigma in *C. annuum* var. *glabriusculum* flowers, while the domesticated types have a wide range of variation in stigma anatomy from protruding to shorter than the anthers (Pickersgill, 1971; Bosland and Votava, 2012). Nevertheless, other factors such as differences in pollinator availability and selection pressures in the anthropogenic environments are expected to be at play.

Gene flow between wild and cultivated chiles

Many of the populations we sampled in the Isthmus of Tehuantepec and Oaxacan coast are separated from wild populations by just a few kilometers, making gene flow between cultivated and wild populations feasible. The feasibility of gene flow is strengthened by the reports of

farmers of the Oaxacan coast that birds disperse seeds of *C. annuum* var. *glabriusculum* in their fields, while local milpa polyculture practices allow for higher species diversity and may attract generalist pollinators that mediate pepper pollination (Landaverde-González et al., 2017; Taitano et al., 2019). Signals of gene flow were revealed by the genetic overlap of population areas between wild and semiwild groups in DAPC analysis (ellipses), as well as by the low paired R_{ST} values between coastal wild and cultivated populations (Figure 2B). This inferred gene flow is also supported by our STRUCTURE results, in which the lowland landraces clustered with the wild populations of *C. annuum* var. *glabriusculum*. Although our data are insufficient to precisely elucidate the direction of gene flow, we must point out that *C. annuum* var. *glabriusculum* can be dispersed by birds from wild to human-managed environments and vice versa. By contrast, most of the cultivated types cannot be dispersed by birds into the rainforest due to their domestication syndrome (Luna-Ruiz et al., 2018; Mares-Quiñones and Valiente-Banuet, 2019). Moreover, pollinators have restricted ranges of activity, with only local effects on cross-pollination (Raw, 2000). Thus, seed dispersal of *C. annuum* var. *glabriusculum* may be the most important agent for gene flow, which we expect to be facilitated in the context of small-plot milpa and backyard production systems where associated agrobiodiversity includes trees and shrubs, ideal for bird perching (Perfecto et al., 2009; Galluzzi et al., 2010).

Our finding that low-elevation populations display higher genetic diversity supports a relevant role of gene flow from wild populations and/or may reflect the origin of domesticated *C. annuum* from lowland populations at its wild range. Whether these patterns are related to adaptation or bottlenecks due to the expansion of the altitudinal range of this crop, cannot be elucidated from our data. Nevertheless, it is interesting that five of six private alleles in cultivated types were found in coastal landraces, suggesting a greater capacity of lowland populations to retain variation, either through gene flow with wild relatives or due to the absence of elevation-related, selective sweeps.

Additional signals of historic gene flow between wild populations and cultivated types can be deducted from the population graph analysis and the DAPC results. Wild and semiwild populations show a high level of covariance (Figure 2B), in concordance with both the geographic proximity and the strong phenotypical resemblance between backyard semiwild and forest understory wild individuals. Interestingly, both wild and semiwild nodes are directly connected to populations of Mirasol (Figure 3B), a landrace that is grown almost exclusively in backyards and small plots (less than 1/8 ha) close to farmers' houses. DAPC shows that the genetic variation area of the semiwild populations overlaps both with wild and with cultivated types, possibly acting as a link between them.

Our evidence suggests that Tusta and Mirasol landraces are related with semiwild and wild populations and may be interacting with nearby *C. frutescens* through gene flow, which would explain why Tusta and Mirasol landraces appear closely correlated in the population graph analyses (Figure 3B) and in the overlapping variance space in the DAPC (Figure 6B). Barrios et al. (2007) identified intra- and interspecific hybridizations within and between species of the *C. annuum–frutescens–chinense* complex in traditional agricultural systems. Additionally, reports of interspecific crossing within this complex have found higher fertility values when wild forms of *Capsicum* species are involved (Eshbaugh, 2012). Potentially, there could be hybridization between backyard-grown landraces of *C. annuum*, such as Tusta and Mirasol, and the let-standing wild-resembling *C. frutescens* and/or wild *C. annuum* var. *glabriusculum* that we found in the same geographical zone (central part of the Oaxacan coast: COP, TOM, JDI, REY, and TEZ locations). Nevertheless, further genome-wide studies are necessary to assess this issue. So far, Tusta and Mirasol landraces highlight the relevance of small-scale production units (see Appendix S17) as genetic diversity reservoirs and may function as a bridge between wild and cultivated gene pools.

Genetic resources and landrace management practices

Previous studies have suggested that backyard and farmer's family garden populations act as genetic reservoirs for *C. annuum* var. *glabriusculum*, since semiwild chile peppers thrive in those spaces (Aguilar-Meléndez et al., 2009; González-Jara et al., 2011). Along the same line, our evidence supports a role for let-standing semiwild peppers, most of them growing in backyards or in the border of small milpas, as a link between wild and cultivated populations. Backyard populations are especially frequent in the coastal region, an area where agricultural plots are interspersed as part of an agroecological matrix with backyards and small milpas either directly surrounded or very close to patches of semideciduous rainforest (Perevochtchikova et al., 2018). Lowland cultivated populations showed lower values of R_{ST} indexes in pairwise comparisons with wild and semiwild types, with the last found mostly in backyards.

Our evidence further reveals the existence of genetic variation specifically associated to landraces. The presence of private alleles in Costeño Rojo, Costeño Amarillo, Gui'ña dani, Tusta, de Agua, and Huacle landraces (see Table 4) as well as the clustering of the landraces in both population graph and DAPC analyses point in this direction. Thus, domesticated peppers have specific combinations of alleles that are found in the highly morphologically divergent landraces; a fact that should be considered in any in situ conservation effort.

Moreover, the distribution of genetic variation reflects to some extent the strong endemism and human management of chile pepper landraces characteristic to the state of Oaxaca. From the landraces that contained private alleles, Chile de Agua and Chile Huacle are endemic to the Central Valleys and the La Cañada regions, respectively (Montaño-Lugo et al., 2014; García-Gaytán et al., 2017), and they are

key ingredients of Oaxacan dishes such as mole (a sauce type including chocolate) or stuffed chiles (Muñoz-Zurita, 2015). In turn, Costeño Rojo, Costeño Amarillo, and Gui'ña dani peppers, for which farmers reported very little seed exchange, also had private alleles. These private alleles contribute to the genetic variation in certain landraces and ultimately add to the high level of variation within populations and within regions detected through AMOVA (Conner and Hartl, 2004). Such patterns offer additional insight, since some landraces have been clearly differentiated through domestication, for example, the big rounded almost black fruits found in the Huacle chile.

Among the coastal landraces, Costeño, Tusta, and Gui'ña dani contain private alleles. Costeño is grown in a relatively small area of the Oaxacan Coast, the Costa Chica, and from there it was brought to Rosedal (130 km to the east). Tusta and Gui'ña dani are endemic to the State of Oaxaca and are grown in relatively small-scale polyculture plots (milpas) and backyards (see Appendix S17). In fact, Gui'ña dani landrace is grown almost exclusively by the Binnizá indigenous people who usually sow the seeds within other main production systems, in two or three rows at the edges of 1–2-ha maize, beans, or tomato growing plots. Costeño, Tusta, and Gui'ña dani are of special importance for local indigenous peoples and are highly valued for local salsas, used either fresh or dry (Muñoz-Zurita, 2015).

The Gui'ña dani landrace showed an interesting pattern. This lowland landrace grouped with the highlands group in the STRUCTURE analysis, and the population graph showed a high covariance between Gui'ña dani and three Valleys-Cañada highland peppers (de Agua, Huacle and Serrano Criollo). There may be a social driver to this pattern. The Binnizá people of the Tehuantepec Isthmus speak Diidxazá, one of the many Zapotec languages spoken in Oaxaca (Smith-Stark, 2007). The literal translation of Gui'ña dani means *chile de monte* or pepper from the mountain. *Monte* is used in two senses across Mexico, to refer to a forested area or to a proper mountain. Linguistic studies point to a close relationship between Zapotecan languages of Central Valleys and those of the lowland Isthmus of Tehuantepec (Smith-Stark, 2007; Beam de Azcona, 2016). Moreover, archeological evidence indicates that after the 14th century there was a large migration from the Central Valleys, resulting in the establishment of a Zapotec *señorío* (kingdom) on the Isthmus of Tehuantepec (Oudijk, 2008). In fact, Binnizá translates literally as people from the clouds in the Diidxazá language. Thus, one possible explanation for the observed genetic clustering between Gui'ña dani and the highland landraces may have to do with a shared historical origin; this same history could also explain the high R_{ST} values between the Gui'ña dani landrace and the rest of the coastal landraces (Figure 2B).

CONCLUSIONS

This study focused on the patterns of genetic diversity and structure in cultivated, semiwild, and wild *C. annuum* populations in three contrasting regions of Oaxaca in southern Mexico. We found evidence that in this area of high landrace diversity, human management and traditional agricultural systems have played and continue to play an important role in the conservation of the genetic resources of *C. annuum* chile peppers.

Gene flow and past evolutionary history have modelled the genetic structure of chile pepper populations and the genetic relationship among wild and domesticated forms. Since pre-Hispanic through modern times, management practices have been acting on *C. annuum* populations in Oaxaca in combination with environmental and ecological aspects such as elevation, migration, mating systems, and pollination dynamics. We found genetic differentiation among highland and lowland populations, that could be indicative of local adaptation to specific environmental conditions probably due to selection, management practices, and local uses of chile peppers in Oaxaca.

Our work highlights the potential for interaction between human management and environmental conditions in shaping the evolution of this crop species within its center of origin. In the context of wild populations declining due to deforestation of their habitats for electricity generation from wind (see Avila-Calero, 2017; Dunlap, 2018) and because of the transition to monoculture (i.e., non-milpa) systems, places like backyards could play an important role for in situ conservation. We found reiterative evidence of an important role for two landraces (Tusta and Mirasol) as possible gene-flow bridges linking cultivated germplasm with semiwild and wild chile peppers. Mating systems and hybridization in chile peppers need to be further explored for a better understanding of how they contribute to shaping the genetic variation of wild, semiwild, and chile pepper landraces.

AUTHOR CONTRIBUTIONS

A.P.M. and L.J.-B. conceived and designed the study. A.P.M. performed the laboratory work and led the data analysis. L.J.-B., K.L.M., L.M., E.K., and L.E. secured funding. A.P.M., L.J.-B., L.E., and N.M.A. led the writing of the manuscript. A.P.M., L.J.-B., L.E., N.M.A., K.L.M., and L.M. reviewed and edited the final version of the manuscript. All authors contributed to ideas, discussed the results, edited the manuscript, and approved its final version.

ACKNOWLEDGMENTS

The authors thank Dr. Erika Aguirre Planter and Dr. Laura Espinosa Asuar for technical support. Thanks to Dr. Gabriela Castellanos Morales for help with the STRUCTURE analyses. We thank Dr. Andrew Michel, Dr. Salvador Montes, Dr. Alejandra Moreno Letelier, Dr. Enrique Scheinvar, Dr. Mariana Benítez, Dr. Catarino Perales, Dr. Brian Pace, Dr. José Carrillo, Dr. Rachel Capouya, Dr. Jessica Pérez, Dr. Nathan Taitano, M. Geog. Cristina Alonso, M. Sc. Fernanda Herce, M. Sc. Adriana Uscanga, Biol. Julia Moreno, and Tania Lara for help during the field collection. We thank Dr. Daniel Piñero for sharing financial support. Special

thanks to farmers, families, and people that shared their knowledge, work, and their chile peppers; their help made this research possible. We thank Cristina Alonso for generously providing photographs of the chile fruits. We thank Marcela San Giacomo Trinidad, Gabriela Pérez Báez, Sandra Smith Aguilar, Eduardo David Vicente, and Pedro Cardona for their insights and literature about linguistic relationships between the Zapotec languages. We thank three anonymous reviewers for their helpful comments to enhance our manuscript. L.J.-B. appreciates not being part of the Mexican Sistema Nacional de Investigadores. This research was financed through the PAPIIT IA-202515 (UNAM-DGAPA) project, CONACyT-Problemas Nacionales Project No. 247730, and by UNAM's annual budget assignment to L.J.-B. and L.E.E. Additional funding came from the Center for Applied Plant Sciences at Ohio State University.

CONFLICTS OF INTEREST/COMPETING INTEREST

Authors declare that they have no conflict of interest nor competing interest, including commercial or intellectual property interest in local germplasm.

DATA AVAILABILITY STATEMENT

Microsatellite data matrix and R scripts used throughout this paper are available at Figshare: https://doi.org/10.6084/m9.figshare.19660929.

ETHICS APPROVAL

This research has been approved by the ethics committee of CEIICH, UNAM as part of Lev Jardón Barbolla research project.

CONSENT TO PARTICIPATE

All plant tissue was collected upon approval of local farmers following the Nagoya Protocol, farmers were informed about the nature and objectives of the research project and promised that no patent nor commercial benefit would be obtained from their plants. All collections were made under the UNAM's Institute of Ecology collection permit for non-endangered cultivated species.

ORCID

Luis E. Eguiarte https://orcid.org/0000-0002-5906-9737
Kristin L. Mercer https://orcid.org/0000-0003-4990-2227
Natalia E. Martínez-Ainsworth https://orcid.org/0000-0001-8043-9483
Leah McHale https://orcid.org/0000-0003-1028-2315
Esther van der Knaap https://orcid.org/0000-0003-4963-7427
Lev Jardón-Barbolla http://orcid.org/0000-0002-3555-1211

REFERENCES

Aguilar-Meléndez, A., P. L. Morrell, M. L. Roose, and S. C. Kim. 2009. Genetic diversity and structure in semiwild and domesticated chiles (*Capsicum annuum*; Solanaceae) from Mexico. *American Journal of Botany* 96: 1190–1202.

Aguilar-Rincón, V. H., T. Corona-Torres, P. López-López, L. Latournerie-Moreno, M. Ramírez-Meraz, H. Villalón-Mendoza, and J. A. Aguilar-Castillo. 2010. Los chiles de México y su distribución. SINAREFI, Colegio de Postgraduados, INIFAP, ITConkal, UANL, UAN. Montecillo, Texcoco, Estado de México.

Aguirre-Liguori, J. A., E. Aguirre-Planter, and L. E. Eguiarte. 2016. Genetics and ecology of wild and cultivated maize: domestication and introgression. *In* R. Lira, A. Casas, and J. Blancas [eds.], Ethnobotany of Mexico: interactions of people and plants in Mesoamerica, 403–416. Springer, NY, NY, USA.

Avila-Calero, S. 2017. Contesting energy transitions: wind power and conflicts in the Isthmus of Tehuantepec. *Journal of Political Ecology* 24: 992–1012.

Balloux, F., and N. Lugon-Moulin. 2002. The estimation of population differentiation with microsatellite markers. *Molecular Ecology* 11: 155–165.

Barrios, O., V. Fuentes, T. Shagarosdky, R. Cristóbal, L. Castiñeiras, Z. Fundora, M. García, et al. 2007. Nuevas combinaciones híbridas de *Capsicum* spp. en sistemas de agricultura tradicional de occidente y oriente de Cuba. *Agrotecnia de Cuba* 31: 327–335.

Beam de Azcona, R. 2016. Zapotecan languages. Oxford Research Encyclopedia of Linguistics. Oxford University Press, Oxford, UK. Website: https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-73 [accessed 12 May 2020].

Bellon, M. R., E. Gotor, and F. Caracciolo. 2015. Conserving landraces and improving livelihoods: how to assess the success of on-farm conservation projects? *International Journal of Agricultural Sustainability* 13: 167–182.

Boslan, P. W., and E. J. Votava. 2012. Peppers: vegetable and spice capsicums. CABI, Wallingford, UK.

Brush, S. B. [ed.]. 2000. Genes in the field: on-farm conservation of crop diversity. Lewis Publishers, Boca Raton, FL, USA.

Caballero, J., L. Cortés, M. A. Martínez-Alfaro, and R. Lira-Saade. 2004. Uso y manejo tradicional de la diversidad vegetal. *In* G. Mendoza, M. J. Ordoñez, and M. Briones-Salas [eds.], Biodiversidad de Oaxaca, 541–564. Instituto de Biología, Universidad Nacional Autónoma de México-Fondo Oaxaqueño para la Conservación de la Naturaleza, Mexico D.F., Mexico; México-World Wildlife Fund, Oaxaca, Mexico.

Carlo, T., and J. Tewksbury. 2014. Directness and tempo of avian seed dispersal increases emergence of wild chiltepins in desert grasslands. *Journal of Ecology* 102: 248–255.

Casas, A., J. Caballero, C. Mapes, and S. Zárate. 1997. Manejo de la vegetación, domesticación de plantas y origen de la agricultura en Mesoamérica. *Botanical Sciences* 61: 31–34.

Casas, A., R. Lira, I. Torres, A. Delgado, A. I. Moreno-Calles, S. Rangel-Landa, J. Blancas, et al. 2016. Ethnobotany for sustainable ecosystem management: a regional perspective in the Tehuacán Valley. *In* R. Lira, A. Casas, and J. Blancas [eds.], Ethnobotany of Mexico: interactions of people and plants in Mesoamerica, 179–206. Springer, NY, NY, USA.

Casas, A., A. Otero-Arnaiz, E. Pérez-Negrón, and A. Valiente-Banuet. 2007. In situ management and domestication of plants in Mesoamerica. *Annals of Botany* 100: 1101–1115.

Casas, A., A. Valiente-Banuet, J. L. Viveros, J. Caballero, L. Cortés, P. Dávila, R. Lira, and I. Rodríguez. 2001. Plant resources of the Tehuacán-Cuicatlán valley, México. *Economic Botany* 55: 129–166.

Chapuis, M. P., and A. Estoup. 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24: 621–631.

Conner, J. K., and D. L. Hartl. 2004. A primer of ecological genetics. Sinauer, Sunderland, MA, USA.

Contreras-Toledo, A. R., H. López-Sánchez, A. Santacruz-Varela, E. Valadez-Moctezuma, V. H. Aguilar-Rincón, T. Corona-Torres, and P. A. López. 2011. Diversidad genética en México de variedades

nativas de chile poblano mediante microsatélites. *Revista Fitotecnia Mexicana* 34: 225–232.

Csardi, M. G. 2013. Package 'igraph'. Website: https://cran.microsoft.com/snapshot/2017-05-27/web/packages/igraph/igraph.pdf [accessed 10 October 2020].

Dakin, E. E., and J. C. Avise. 2004. Microsatellite null alleles in parentage analysis. *Heredity* 93: 504–509.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf material. *Phytochemical Bulletin* 19: 11–15.

Dunlap, A. 2018. The 'solution' is now the 'problem': wind energy, colonisation and the 'genocide-ecocide nexus' in the Isthmus of Tehuantepec, Oaxaca. *International Journal of Human Rights* 22(4): 550–573.

Dyer, R. J. 2012. The gstudio package. Virginia Commonwealth University, Richmond, VA, USA. Website: https://www.researchgate.net/profile/Rodney_Dyer/publication/267710510_The_gstudio_Package/links/569d596008aed27a702fa791/The-gstudio-Package.pdf [accessed 10 October 2020].

Dyer, R. J. 2014. popgraph: This is an R package that constructs and manipulates population graphs. R package version 1.4. Website: http://cran.r-project.org/package=popgraph [accessed 8 October 2020].

Dyer, R. J. 2015. Population graphs and landscape genetics. *Annual Review of Ecology, Evolution, and Systematics* 46: 327–342.

Dyer, R. J., and N. D. Nason. 2004. Population graphs: the graph theoretic shape of genetic structure. *Molecular Ecology* 13: 1713–1727.

Earl, D. A., and B. M. von Holdt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.

Eshbaugh, W. H. 2012. The taxonomy of the genus *Capsicum*. *In* V. M. Russo [ed.], Peppers: botany, production and uses, 14–28. CABI, Cambridge, MA, USA.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin (version 3.5): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* 1: 47–50.

Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.

FAO [Food and Agricultural Organizations of the United Nations]. 2019. FAOSTAT statistics database. Website http://www.fao.org/faostat/en/ [accessed August 2019].

Fox, J., R. Muenchen, D. Putler, and M. J. Fox. 2020. Package 'RcmdrMisc'. Website http://mirrors.nic.cz/R/web/packages/RcmdrMisc/RcmdrMisc.pdf [accessed 31 September 2020].

Galluzzi, G., P. Eyzaguirre, and V. Negri. 2010. Home gardens: neglected hotspots of agro-biodiversity and cultural diversity. *Biodiversity and Conservation* 19: 3635–3654.

García-Gaytán, V., F. C. Gómez-Merino, L. I. Trejo-Téllez, G. A. Baca-Castillo, and S. García-Morales. 2017. The Chilhuacle chili (*Capsicum annuum* L.) in Mexico: description of the variety, its cultivation, and uses. *International Journal of Agronomy* 2017: 1–14.

García-Mendoza, A. J., M. J. Ordoñez, and M. Briones-Salas. 2004. Biodiversidad de Oaxaca. Instituto de Biología, Universidad Nacional Autónoma de México-Fondo Oaxaqueño para la Conservación de la Naturaleza, Mexico City, Mexico; México-World Wildlife Fund, Oaxaca, México.

Garza, J. C., and E. G. Williamson. 2001. Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* 10: 305–318.

González-Jara, P., A. Moreno-Letelier, A. Fraile, D. Piñero, and F. García-Arenal. 2011. Impact of human management on the genetic variation of wild pepper, *Capsicum annuum* var. *glabriusculum*. *PLoS One* 6: 1–11.

González-Pérez, S., A. Garcés-Claver, C. Mallor, L. E. S. de Miera, O. Fayos, F. Pomar, F. Merino, and C. Silvar. 2014. New insights into *Capsicum* spp. relatedness and the diversification process of *Capsicum annuum* in Spain. *PLoS One* 9: e116276.

Gotelli, N. J., and J. Stanton-Geddes. 2015. Climate change, genetic markers and species distribution modelling. *Journal of Biogeography* 42: 1577–1585.

Guerra-García, A., M. Suárez-Atilano, A. Mastretta-Yanes, A. Delgado-Salinas, and D. Piñero. 2017. Domestication genomics of the open-pollinated scarlet runner bean (*Phaseolus coccineus* L.). *Frontiers in Plant Science* 8: 1891.

Hernández-Verdugo, S., A. P. Dávila, and K. Oyama. 1999. Síntesis del conocimiento taxonómico, origen y domesticación del género *Capsicum*. *Boletín de la Sociedad Botánica de México* 64: 65–84.

Hernández-Verdugo, S., R. Luna-Reyes, and K. Oyama. 2001. Genetic structure and differentiation of wild and domesticated populations of *Capsicum annuum* (Solanaceae) from Mexico. *Plant Systematics and Evolution* 226: 129–142.

Hill, T. A., H. Ashrafi, S. Reyes-Chin-Wo, J. Yao, K. Stoffel, M. J. Truco, A. Kozik, et al. 2013. Characterization of *Capsicum annuum* genetic diversity and population structure based on parallel polymorphism discovery with a 30 K unigene Pepper GeneChip. *PLoS One* 8: e56200.

Idohou, R., B. Fandohan, V. K. Salako, B. Kassa, R. C. Gbèdomon, H. Yédomonhan, R. L. G. Kakaï, and A. E. Assogbadjo. 2014. Biodiversity conservation in home gardens: traditional knowledge, use patterns and implications for management. *International Journal of Biodiversity Science, Ecosystem Services & Management* 10: 89–100.

Jombart, T. 2008. adegenet: an R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11: 94–108.

Kahle, D., and H. Wickham. 2013. ggmap: Spatial visualization with ggplot2. R Journal 5: 144–161.

Kraft, K. H., C. H. Brown, G. P. Nabhan, E. Luedeling, J. D. J. L. Ruiz, G. C. d'Eeckenbrugge, R. J. Hijmans, and P. Gepts. 2014. Multiple lines of evidence for the origin of domesticated chili pepper, *Capsicum annuum*, in Mexico. *Proceedings of the National Academy of Sciences, USA* 111: 6165–6170.

Landaverde-González, P., J. J. G. Quezada-Euán, P. Theodorou, T. E. Murray, M. Husemann, R. Ayala, H. Moo-Valle, et al. 2017. Sweat bees on hot chillies: provision of pollination services by native bees in traditional slash-and-burn agriculture in the Yucatán Peninsula of tropical Mexico. *Journal of Applied Ecology* 54: 1814–1824.

Latournerie, L., J. L. Chávez, M. Pérez, C. F. Hernández, R. Martínez, L. M. Arias, and G. Castañon. 2001. Exploración de la diversidad morfológica de chiles regionales en Yaxcaba, Yucatán, México. *Agronomía Mesoamericana* 12: 41–47.

Loaiza-Figueroa, F., K. Ritland, J. A. L. Cancino, and S. D. Tanksley. 1989. Patterns of genetic variation of the genus *Capsicum* (Solanaceae) in Mexico. *Plant Systematics and Evolution* 165: 159–188.

Long Towell, J. 2009. Los senderos prehispánicos del *Capsicum*. J. Long Towell and A. León [coordinators], Caminos y mercados de México. *Serie Histórica General* 23: 79–105.

Long-Solís, J. 2012. Capsicum y cultura: la historia del chilli, 203. Fondo de Cultura Económica, Mexico City, Mexico.

de Jesus Luna-Ruiz, J., G. P. Nabhan, and A. Aguilar-Meléndez. 2018. Shifts in plant chemical defenses of chile pepper (*Capsicum annuum* L.) due to domestication in Mesoamerica. *Frontiers in Ecology and Evolution* 6: 48.

Mapes, C., and F. Basurto. 2016. Biodiversity and edible plants of Mexico. *In* R. Lira, A. Casas, and J. Blancas [eds.], Ethnobotany of Mexico: interactions of people and plants in Mesoamerica, 83–131. Springer, NY, NY, USA.

Mares-Quiñones, M. D., and J. I. Valiente-Banuet. 2019. Horticultural aspects for the cultivated production of piquin peppers (*Capsicum annuum* L. var. *glabriusculum*)—A review. *HortScience* 54: 70–75.

McLeod, M. J., S. I. Guttman, W. H. Eshbaugh, and R. E. Rayle. 1983. An electrophoretic study of evolution in *Capsicum* (Solanaceae). *Evolution* 37: 562–574.

Mercer, K. L., and H. R. Perales. 2019. Structure of local adaptation across the landscape: flowering time and fitness in Mexican maize (*Zea mays* L. subsp. *mays*) landraces. *Genetic Resources and Crop Evolution* 66: 27–45.

Montaño-Lugo, M. L., V. A. Velasco-Velasco, J. Ruíz-Luna, C. Ángeles, G. Virginia, G. Rodríguez-Ortiz, and M. L. Martínez. 2014. Contribución al conocimiento etnobotánico del chile de agua (*Capsicum annuum* L.) en los Valles Centrales de Oaxaca, México. *Revista Mexicana de Ciencias Agrícolas* 5: 503–511.

Montes-Hernández, S. P., S. López-López, and M. Ramírez-Meraz. 2010. Recopilación y análisis de la información existente de las especies del género *Capsicum* que crecen y se cultivan en México (informe final). Comisión Nacional para el Conocimiento y Uso de la Biodiversidad [CONABIO], technical report. National Institute of Forestry, Agricultural and Livestock [INIFAP], México D.F., Mexico.

Muñoz-Zurita, R. 2015. Los chiles nativos de México. DGE Equilibrista, Mexico City, Mexico.

Nagy, I., A. Stagel, Z. Sasvari, M. Röder, and M. Ganal. 2007. Development, characterization, and transferability to other Solanaceae of microsatellite markers in pepper (*Capsicum annuum* L.). *Genome* 50: 668–688.

Nuez, F. V., R. Gil-Ortega, and J. G. Costa. 2003. El cultivo de pimientos, chiles y ajíes, 586. Ediciones Mundi-Prensa, Madrid, Spain.

Oudijk, M. 2008. The Postclassic period in the valley of Oaxaca: The archaeological and ethnohistorical records. *In* J. Blomster [ed.], After Monte Albán: transformation and negotiation in Oaxaca, Mexico. University Press of Colorado, Boulder, CO, USA.

Oyama, K., S. Hernández-Verdugo, C. Sánchez, A. González-Rodríguez, P. Sánchez-Peña, J. A. Garzón-Tiznado, and A. Casas. 2006. Genetic structure of wild and domesticated populations of *Capsicum annuum* (Solanaceae) from northwestern Mexico analyzed by RAPDs. *Genetic Resources and Crop Evolution* 53: 553–562.

Oyama, K., M. Martínez-Ramos, J. M. Peñaloza-Ramírez, V. Rocha-Ramírez, E. G. Armenta-Medina, and P. Hernández-Soto. 2016. Population genetic structure of an extremely logged tree species *Guaiacum sanctum* L. in the Yucatan Peninsula, Mexico. *Botanical Sciences* 94: 345–356.

Pacheco-Olvera, A., S. Hernández-Verdugo, V. Rocha-Ramírez, A. González-Rodríguez, and K. Oyama. 2012. Genetic diversity and structure of pepper (L.) from northwestern Mexico analyzed by microsatellite markers. *Crop Science* 52: 231–241.

Peery, M. Z., R. Kirby, B. N. Reid, R. Stoelting, E. Doucet-Bëer, S. Robinson, C. Vásquez-Carrillo, et al. 2012. Reliability of genetic bottleneck tests for detecting recent population declines. *Molecular Ecology* 21: 3403–3418.

Perevochtchikova, M., J. A. Hernández, and V. S. Avila-Foucat. 2018. Recursos naturales y diversificación productiva en cuatro localidades rurales del Estado de Oaxaca, México. *Cuadernos de Desarrollo Rural* 15: 1–25.

Perfecto, I., J. Vandermeer, and A. Wright. 2009. Nature's matrix: linking agriculture, conservation and food sovereignty, 242. Earthscan, London, UK.

Perry, L., and K. V. Flannery. 2007. Precolumbian use of chili peppers in the Valley of Oaxaca, Mexico. *Proceedings of the National Academy of Sciences, USA* 104: 11905–11909.

Pickersgill, B. 1971. Relationships between weedy and cultivated forms in some species of chili peppers (genus *Capsicum*). *Evolution* 25: 683–691.

Pickersgill, B. 1997. Genetic resources and breeding of *Capsicum* spp. *Euphytica* 96: 129–133.

Pickersgill, B. 2016. Chile peppers (*Capsicum* spp.). *In* R. Lira, A. Casas, and J. Blancas [eds.], Ethnobotany of Mexico: interactions of people and plants in Mesoamerica, 207–231. Springer, NY, NY, USA.

Pressoir, G., and J. Berthaud. 2004. Population structure and strong divergent selection shape phenotypic diversification in maize landraces. *Heredity* 92: 95–101.

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

Rai, V. P., R. Kumar, S. Kumar, A. Rai, S. Kumar, M. Singh, S. P. Singh, et al. 2013. Genetic diversity in *Capsicum* germplasm based on microsatellite and random amplified microsatellite polymorphism markers. *Physiology and Molecular Biology of Plants* 19: 575–586.

Raw, A. 2000. Foraging behaviour of wild bees at hot pepper flowers (*Capsicum annuum*) and its possible influence on cross pollination. *Annals of Botany* 85: 487–492.

Rivera, A., A. B. Monteagudo, E. Igartua, A. Taboada, A. García-Ulloa, F. Pomar, M. Riveiro-Leira, and C. Silvar. 2016. Assessing genetic and phenotypic diversity in pepper (*Capsicum annuum* L.) landraces from North-West Spain. *Scientia Horticulturae* 203: 1–11.

Rousset, F. 2008. GENEPOP'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106.

SAGARPA [Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación]. 2017. Planeación Agrícola Nacional 2017–2030: Chiles y pimientos Mexicanos. SAGARPA, Mexico DF, Mexico. Website: https://www.gob.mx/cms/uploads/attachment/file/257072/Potencial-Chiles_y_Pimientos-parte_uno.pdf [accessed 26 January 2020].

Samberg, L. H., L. Fishman, and F. W. Allendorf. 2013. Population genetic structure in a social landscape: barley in a traditional Ethiopian agricultural system. *Evolutionary Applications* 6: 1133–1145.

Sato, S., S. Isobe, E. Asamizu, N. Ohmido, R. Kataoka, Y. Nakamura, T. Kaneko, et al. 2005. Comprehensive structural analysis of the genome of red clover (*Trifolium pratense* L.). *DNA Research* 12: 301–364.

SEMARNAT [Secretaría de Medio Ambiente y Recursos Naturales]. 2018. Las 64 variedades de chile en México. Website: https://www.gob.mx/semarnat/articulos/las-64-variedades-de-chile-en-mexico?idiom=es [accessed 26 January 2020].

Shirasawa, K., K. Ishii, C. Kim, T. Ban, M. Suzuki, T. Ito, T. Muranaka, et al. 2013. Development of *Capsicum* EST-SSR markers for species identification and in silico mapping onto the tomato genome sequence. *Molecular Breeding* 31: 101–111.

SIAP [Servicio de Información Agroalimentaria y Pesquera, reporte producción]. 2018. México. Website: http://infosiap.siap.gob.mx:8080/agricola_siap_gobmx/ResumenProducto.do [accessed 5 August 2019].

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457–462.

Smith-Stark, T. 2007. Algunas isoglosas zapotecas. *In* C. Buenrostro, S. Herrera Castro, Y. Lastra, F. Nava, J. J. Rendón, O. Schumann, L. Valiñas, and M. A. Vargas Monroy [eds.], Clasificación de las lenguas indígenas de México. UNAM-INALI, Mexico DF, Mexico.

Szpiech, Z. A., M. Jakobsson, and N. A. Rosenberg. 2008. ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* 24: 2498–2504.

Taitano, N., V. Bernau, L. Jardón-Barbolla, B. Leckie, M. Mazourek, K. Mercer, L. McHale, et al. 2019. Genome-wide genotyping of a novel Mexican chile pepper collection illuminates the history of landrace differentiation after *Capsicum annuum* L. domestication. *Evolutionary Applications* 12: 78–92.

Toledo-Aguilar, R., H. López-Sánchez, A. Santacruz-Varela, E. Valadez-Moctezuma, P. A. López, V. H. Aguilar-Rincón, V. A. González-Hernández, and H. Vaquera-Huerta. 2016. Characterization of genetic diversity of native Ancho chili populations of Mexico using microsatellite markers. *Chilean Journal of Agricultural Research* 76: 18–26.

Vavilov, N. I. 1931. Mexico and Central America as the principal centre of origin of cultivated plants of New World. *In* N. I. Vavilov. 2009. Origin and geography of cultivated plants [translated from Russian by D. Love]. Cambridge University Press, Cambridge, UK.

Vázquez-Lobo, A., S. Ortiz, and E. Alvarez-Buylla. 1996. Mini-prep for easy DNA extraction of conifer species. Appendix 6 in A. Vázquez-Lobo Yuren, Filogenia de hongos endófitos del género *Pinus* L.: implementación de técnicas moleculares y resultados preliminares. Tesis de Licenciatura [Bachelor's thesis]. Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico D.F., Mexico. Website: https://repositorio.unam.mx/contenidos/filogenia-de-hongos-endofitos-del-genero-pinus-l-implementacion-de-tecnicas-moleculares-y-resultados-preliminares-214755?c=r37gKJ%26d=false%26q=%2A:%2A%26i=9%26v=1%26t=search_0%26as=0

Wei, T., V. Simko, M. Levy, Y. Xie, Y. Jin, and J. Zemla. 2017. Package 'corrplot'. *Statistician* 56: e24.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Results of statistical analyses using the RCMDRMISC package (by region and by level of domestication).

Appendix S2. Genetic diversity values at the population and geographical level.

Appendix S3. Estimated frequencies of null alleles by population and landrace obtained with FreeNA software.

Appendix S4. Genotypic linkage disequilibrium for *Capsicum annuum* populations obtained by GENEPOP software.

Appendix S5. Allelic richness by sample size of Oaxaca cultivated landraces, semiwild and wild chile peppers obtained by rarefaction using ADZE.

Appendix S6. Analysis of molecular variance (AMOVA) of microsatellite loci by geographical region and by level of domestication.

Appendix S7. Population pairwise R_{ST} values.

Appendix S8. Matrix of significant R_{ST} *P*-values across localities.

Appendix S9. Landraces, semiwild, and wild populations pairwise R_{ST}.

Appendix S10. Matrix of significant R_{ST} *P*-values (corrected) by landraces, semiwild, and wild populations.

Appendix S11. Mantel test using genetic differentiation (R_{ST}) and Euclidean distances (geographical).

Appendix S12. Graphic results of Evanno test.

Appendix S13. Correlation between cluster assignment and population elevation.

Appendix S14. Discriminant analysis of principal components (DAPC) assignment probabilities by geographical region, by level of domestication and landrace name.

Appendix S15. Discriminant analysis of principal components (DAPC) graph by geographical region.

Appendix S16. Assignment probabilities to a discriminant analysis of principal components (DAPC) cluster plotted for each individual.

Appendix S17. Information about management regimes and endemism of each landrace.

How to cite this article: Pérez-Martínez, A. L., L. E. Eguiarte, K. L. Mercer, N. E. Martínez-Ainsworth, L. McHale, E. van der Knaap, and L. Jardón-Barbolla. 2022. Genetic diversity, gene flow, and differentiation among wild, semiwild, and landrace chile pepper (*Capsicum annuum*) populations in Oaxaca, Mexico. *American Journal of Botany* 109(7): 1157–1176. https://doi.org/10.1002/ajb2.16019