Received: 4 May 2023 | Accepted: 20 June 2023

DOI: 10.1002/ajb2.16216

RESEARCH ARTICLE

Botany

Genomic analysis unveils reduced genetic variability but increased proportion of heterozygotic genotypes of the intensively managed mezcal agave, *Agave angustifolia*

Anastasia Klimova[1] | **Karen Y. Ruiz Mondragón**[1] | **Erika Aguirre-Planter**[1] | **Alfonso Valiente**[1] | **Rafael Lira**[2] | **Luis E. Eguiarte**[1]

[1]Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

[2]Laboratorio de Recursos Naturales, Unidad de Biotecnología y Prototipos (UBIPRO), Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

Correspondence

Luis E. Eguiarte and Anastasia Klimova, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico.
Email: fruns@unam.mx and anastasia_aleksandrovna@hotmail.com

Abstract

Premise: The central Oaxaca Basin has a century-long history of agave cultivation and is hypothesized to be the region of origin of other cultivated crops. Widely cultivated for mezcal production, the perennial crop known as "espadín" is putatively derived from wild *Agave angustifolia*. Nevertheless, little is known about its genetic relationship to the wild *A. angustifolia* or how the decades-long clonal propagation has affected its genetics.

Methods: Using restriction-site-associated DNA sequencing and over 8000 single-nucleotide polymorphisms, we studied aspects of the population genomics of wild and cultivated *A. angustifolia* in Puebla and Oaxaca, Mexico. We assessed patterns of genetic diversity, inbreeding, distribution of genetic variation, and differentiation among and within wild populations and plantations.

Results: Genetic differentiation between wild and cultivated plants was strong, and both gene pools harbored multiple unique alleles. Nevertheless, we found several cultivated individuals with high genetic affinity with wild samples. Higher heterozygosity was observed in the cultivated individuals, while in total, they harbored considerably fewer alleles and presented higher linkage disequilibrium compared to the wild plants. Independently of geographic distance among sampled plantations, the genetic relatedness of the cultivated plants was high, suggesting a common origin and prevalent role of clonal propagation.

Conclusions: The considerable heterozygosity found in espadín is contained within a network of highly related individuals, displaying high linkage disequilibrium generated by decades of clonal propagation and possibly by the accumulation of somatic mutations. Wild *A. angustifolia*, on the other hand, represents a significant genetic diversity reservoir that should be carefully studied and conserved.

KEYWORDS

Agavoideae, clonal propagation, crop domestication, crop wild relatives, genetic resources, RADseq

Mesoamerica has been recognized as a major world center of plant domestication (Vavilov, 1931; Smith, 2001; Clement et al., 2021). Pre-Columbian people domesticated over 200 native species (Pickersgill, 2016), some of which are now of global importance, including maize, beans, and squashes (Shiferaw et al., 2011; Bitocchi et al., 2017). Domestication is a complex evolutionary process in which human selection and cultivation of a plant leads to genetic, morphological, and physiological changes that distinguish domesticated plant populations from their wild relatives (Ross-Ibarra et al., 2007; Zeller and Göttert, 2019). In general, only a tiny proportion of wild progenitors are used for propagation;

this process may result in genetic bottlenecks and founder effects that vary in intensity and duration, which ultimately would lead to the erosion of genetic diversity (Khoury et al., 2021). A reduction of genetic diversity attributed to domestication has been reported in several species, such as soybean, tomato, maize, and rice (Hyten et al., 2006; Bai and Lindhout, 2007; Tian et al., 2009; Khoury et al., 2021). Besides the genetic and demographic effects of genetic drift, human selection further decreases genetic diversity in plants under domestication through the elimination of allelic variants of genes with undesirable traits (Kantar et al., 2017). Moreover, strong positive selection on loci controlling a trait of interest can result in a selective sweep at linked genetic regions. Many selective sweeps have been found at loci associated with domesticated traits (Shi and Lai, 2015). However, the degree of the selective sweep would depend on the present linkage among loci, selection intensity, and species reproductive system (Racimo et al., 2016).

Therefore, domestication and artificial selection may have eroded genetic diversity at specific loci in many modern crops, thus limiting to some extent their potential for developing novel varieties. Furthermore, in many crops species, during domestication, and due to clonal propagation and insufficient genomic recombination events, mildly deleterious mutations can reach high frequency (Ramu et al., 2017; Zhu et al., 2022), resulting in a fitness reduction of the domesticated plant and yield losses. Crop wild relatives, on the other hand, usually present higher levels of genetic variability in comparison to their domesticated conspecifics (Hübner and Kantar, 2021; Bohra et al., 2022). Usually, wild relatives possess more genetic diversity at the population and individual levels (e.g., heterozygosity), with the notorious exception of clonally propagated crops (McKey et al., 2010; Potato Genome Sequencing Consortium, 2011; Zhang et al., 2016). High heterozygosity in clonal species has been theoretically and empirically demonstrated (Balloux et al., 2003). In clonal crops, it has been attributed to the preservation of ancestral diversity and further accumulation of somatic mutations (Fischer and Van Kleunen, 2001; Schoen and Schultz, 2019; Yu et al., 2020).

Clonal propagation and its advantages for cultivation have significantly impacted the domestication of many crop species (e.g., cassava, hops, strawberry, agave, and potato) and were readily exploited by farmers (McKey et al., 2010). Clonal propagation allows the preservation of desirable genotypes that, in some cases, are highly heterozygous genotypes that show hybrid vigor, as well as rapid fixation of the agronomically valuable traits (McKey et al., 2010). Favorable mutations can easily be identified in the field and quickly propagated. Moreover, clonal propagation prevents genetic exchange with crop wild relatives, and farmers can control the introduction of genes from wild populations (McKey et al., 2010; Denham et al., 2020). Finally, clonal propagation is also the easiest way to multiply crop species compared to propagation by seed (Arizaga and Ezcurra, 2002; Gaut et al., 2015). Clonal propagation has a number of drawbacks, such as the accumulation of deleterious mutation, erosion of genetic diversity, and an increase in pathogen burden (McKey et al., 2010; Lian et al., 2019).

Agave (Asparagaceae, Agavoideae) species have had immense cultural and economic value in Mexico, where their use by humans dates back at least 9000 years. Mesoamericans historically used agave for food, fiber, and its sweet sap, consumed directly or used to prepare fermented beverages (Callen, 1965, 1967; MacNeish, 1967). Moreover, agaves have a wide array of ecological, reproductive, and morphological adaptations to arid environments, and they are an essential part of arid landscapes (Gentry, 2004; Eguiarte et al., 2013, 2021). Many agave species combine vegetative and sexual propagation, which allows them to extend generation time and secure survival in harsh desert environments (Gentry, 1972; Arizaga and Ezcurra, 2002). At least 40 agave species are used for mezcal and similar beverage production (Tetreault et al., 2021; COMERCAM, 2022). Some species are cultivated from wild seeds and subjected to almost no artificial selection and management; consequently, in some cases, cultivated *Agave* individuals are genetically indistinguishable from their wild relatives (Félix-Valdez et al., 2015; Cabrera-Toledo et al., 2020; Klimova et al., 2022). In contrast, others, such as *A. tequilana* var. Azul, have been intensively managed and clonally propagated for decades (Vargas-Ponce et al., 2009; Trejo et al., 2018; Ruiz-Mondragon et al., 2022).

Time under cultivation and the intensity of management practices have been reflected in the degree of genetic and morphological differentiation between wild and cultivated agave conspecifics (Félix-Valdez et al., 2015; Figueredo et al., 2015; Klimova et al., 2022). Several intensively managed agave species are reported to adjust to a domestication syndrome in different traits, including larger leaves and smaller leaf dentition (Valenzuela, 2010; Figueredo-Urbina et al., 2021). Furthermore, the genetic diversity of species such as *A. tequilana* var. Azul seems to be compromised (Gil-Vega et al., 2001; Vargas-Ponce et al., 2009; Trejo et al., 2018). Nevertheless, more recent studies based on single-nucleotide polymorphisms (SNPs) found heterozygosity excess and negative inbreeding index at plantations with clonally propagated agave plants (Ruiz-Mondragon et al., 2022; Cabrera-Toledo et al., 2022). These contrasting findings are probably the product of the low-resolution markers used in the older studies, which may be inappropriate for the research of the clonally propagated species (Arnaud-Haond et al., 2005).

In recent decades, the production of alcoholic beverages derived from agave distillation in Mexico has been growing exponentially (Arellano-Plaza et al., 2022), threatening wild agave populations, destroying natural vegetation, and introducing intensive management to cultivation (Tetreault et al., 2021; Lira et al., 2022). Alongside tequila, mezcal has become an iconic Mexican alcoholic beverage (Arellano-Plaza et al., 2022). Mezcal production reached 8 million liters in 2021 compared to only 1 million in 2011

(COMERCAM, 2022). According to the official Mexican standard (NOM-070-SCFI-2016), 13 Mexican states currently have a federal Mezcal Denomination of Origin. Each state specializes in different agave species; for example, the main species for mezcal production in the state of Michoacán is *A. cupreata*, *A. salmiana* in the state of San Luis Potosi, and *A. potatorum* and A. *marmorata* in Puebla. However, around 90% of mezcal production is concentrated in Oaxaca state in central Mexico (Arellano-Plaza et al., 2022). The main variety (88.11%) used to produce mezcal is known among producers as "espadín" (COMERCAM 2022). Cultivated espadín is believed to have originated from the wild *A. angustifolia* (Gentry, 2004), but the exact geographic and genetic origin of the cultivated espadín is still under debate (Rivera-Lugo et al., 2018).

Wild *A. angustifolia* has the broadest geographic distribution among the *Agave* species. It can be found from Sonora in northern Mexico to Costa Rica in Central America (Gentry, 2004). *Agave angustifolia* is a diploid, self-incompatible species (Molina-Freaner and Eguiarte, 2003; Moreno Salazar et al., 2007; Rivera-Lugo et al., 2018) that can be propagated sexually and vegetatively (Gentry, 2004). It has a broad group of pollinators, including *Leptonycteris* bats, birds, and bees (Molina-Freaner and Eguiarte, 2003).

Although central Mexico and particularly the state of Oaxaca potentially harbor an enormous reservoir of genetic diversity of *A. angustifolia* and has a century-long history of agave cultivation, virtually nothing is known about the genetic composition of either wild or cultivated agaves in this region (Rivera-Lugo et al., 2018). Moreover, very little is known about how intensive management may have affected the genomic composition of the cultivated espadín and whether or not it has differentiated from its wild relatives. Therefore, information on genomewide patterns of genetic variation and knowledge of population structure of wild and cultivated *A. angustifolia* in Oaxaca state is essential to define priorities for managing and conserving gene pools, to develop new sustainable cropping systems, and to study the impact of ongoing domestication on agave genome.

Genetic variation in crops represents a combination of diversity found in ancestral wild populations, possible secondary sexual contact with the wild conspecifics found close to the cultivation area, and intensity and direction of artificial selection (Kantar et al., 2017). However, in clonally propagated crops, once sexual recombination becomes infrequent, mitotic processes such as the accumulation of spontaneous somatic mutations (Balloux et al., 2003) may become increasingly important (Myles et al., 2011; Zhou et al., 2017; Favre et al., 2022). Therefore, we expect novel variations from these sources to accumulate over time in cultivated espadín and be detectable in the form of increased heterozygosity and reduced inbreeding index. Moreover, due to decades of intensive management, clonal propagation, and restricted gene flow, we expect detectable genetic differentiation between wild and cultivated samples.

To better understand the trade-offs associated with clonal propagation, we measured the effect of the ongoing domestication on the genomic diversity in one of the most economically important agave species, *A. angustifolia*, and its cultivated variety, espadín. We used RADseq and focused on the genomic diversity of wild and cultivated *A. angustifolia* in the Oaxaca and adjacent Puebla state. We addressed the following questions: (1) Are cultivated espadín plants genetically different from wild *A. angustifolia* individuals in the region? (2) How was the genetic diversity of cultivated plants affected by long-term intensive management and clonal propagation? (3) How much (local) genetic diversity is preserved in cultivated plants? (4) How ubiquitous is clonal propagation as the principal reproduction strategy across wild and domesticated plants? (5) What are the implications for conserving cultivated espadín and wild *A. angustifolia* genetic resources in Mexico?

MATERIALS AND METHODS

Sample collection

In 2021 and 2022, we sampled 16 plantations and 13 wild sites in Puebla and Oaxaca, Mexico (Figure 1; Appendix S1). At each plantation, we collected cultivated plants locally known as espadín. Wild sampling sites of *A. angustifolia* were located far enough from managed land, and the sampled plants were apparently not propagated or cultivated by humans. Although we did not measure the distance explicitly while sampling, the two closest (OW7 and OM14) wild and cultivated sites were 5 km from each other. From each sampling site, we collected one to six plants depending on availability, because wild plants of this species in the region seldom have large populations. Leaves from individual plants were collected fresh and kept in paper bags at room temperature, away from heat and direct sunlight. Once in the laboratory, samples were stored at −20°C until DNA extraction. The final number of the collected samples was 49 for wild and 47 for cultivated agaves. Details on geographic coordinates, the number of specimens collected at each site, and management type are in Appendix S1.

Molecular analysis and SNP selection

For DNA extraction, one to four individuals were randomly chosen from each sampling site (Appendix S1). Total DNA was extracted from leaf tissue using a modified CTAB protocol (Doyle and Doyle, 1987; Klimova et al., 2022). DNA quality and quantity were checked in 1% agarose and a Qubit 3.0 fluorometer (Qubit dsDNA broad-range kit). DNA of adequate quantity and quality from 94 plants was sent to the University of Wisconsin Biotechnology Center for RADseq, library preparation, and 150-bp paired-end sequencing using the Illumina NovaSeq platform (Illumina, San Diego, CA, USA). A combination of methylation-sensitive restriction enzymes (PstI/MspI) was used for DNA digestion.

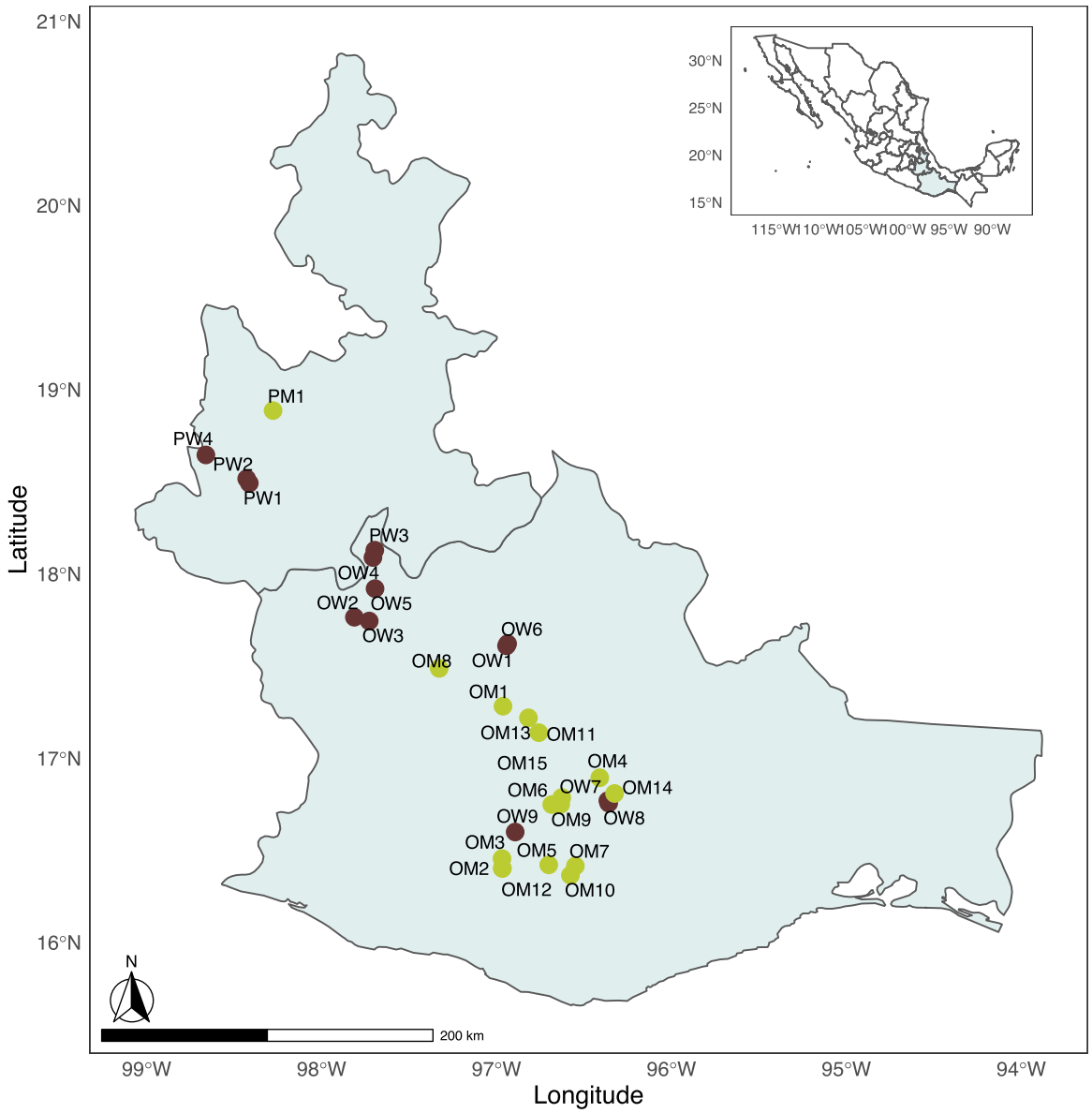

FIGURE 1 Map of sampling sites of wild *Agave angustifolia* (brown) and cultivated espadín (green) from the states of Puebla and Oaxaca, Mexico. The full names of each site, geographic coordinates, and sample sizes are given in Appendix S1.

Raw sequences were demultiplexed and filtered by removing adapters, short (<55 bp), and low-quality reads based on strict parameters as implemented in ipyrad v. 0.9.79 (Eaton and Overcast, 2020). Filtered reads were mapped to the reference transcriptome of *Agave tequilana* (GAHU00000000.1; Gross et al., 2013), and variants were called using the paired-end double enzyme methodology described in ipyrad. The posterior variant filtering was performed using VCFtools v.0.1.15 (Danecek et al., 2011). Only loci with mean depth (across individuals) of over five and less than 200, a maximum of two alleles with no insertions or deletions, were kept. Additionally, we set a minor allele frequency at 0.05 to reduce the possibility of including singletons or removing rare alleles necessary in elucidating fine-scale structure (Linck and Battey, 2019). We excluded DNA sites based on the proportion of missing data, keeping sites with no more than 20% missing data (–max-missing 0.8). We then filtered out the variants that significantly deviated from Hardy–Weinberg equilibrium ($P \leq 0.05$ after multiple test corrections). We also estimated the correlation between each pair of loci (r^2). Then, to ameliorate the confounding effects of linkage disequilibrium (LD), we eliminated markers with r^2 of over 0.2 as implemented in plink (--indep-pairwise 50 5 0.2) (Purcell et al., 2007). Finally, to determine the genotype's uniqueness, we used the mlg function implemented in the R package poppr (Kamvar et al., 2014).

Since some analyses were performed separately for the wild and cultivated samples, we prepared three data sets: wild samples only, cultivated samples only, and all samples.

All the filters mentioned above were performed separately for each of the three data sets. Whenever multiple tests were used, the resulting *P*-values were adjusted for false discovery rate (FDR) using the R package stats (R Core Team, 2021). The final VCF files were then produced for all downstream analyses. The same filtering parameters were applied to the complete data set of 94 samples and the two files for wild and cultivated samples.

Genetic diversity and relatedness

Basic diversity statistics, such as expected (H_E) and observed (H_O) heterozygosity, were calculated overall and for wild and cultivated samples separately. For each type of sample, we also estimated the weighted average nucleotide diversity per site (Π), where the weights are determined by the number of genotyped samples at each site as implemented in pixy software (Korunes and Samuk, 2021). We also estimated multilocus heterozygosity (MLH) and standardized multilocus heterozygosity (sMLH) for each individual (Stoffel et al., 2016). Multilocus heterozygosity is defined as the total number of heterozygous loci in an individual divided by the number of loci typed in this individual. It is easily interpreted; for example, an MLH of 0.2 means that 20% of loci are heterozygous in an individual in question. Whereas sMLH is defined as the total number of heterozygous loci in an individual divided by the sum of average observed heterozygosities in the population over the subset of loci successfully typed in the focal individual. Thus, the standardization ensures that all individuals are measured on an identical scale, despite the differences in marker information (Slate et al., 2004). The diversity statistics mentioned above were calculated using R packages adegenet and inbreedR (Stoffel et al., 2016). In addition, we used plink 1.9 (Purcell et al., 2007) to calculate the individual-based inbreeding index Fhat3 (Yang et al., 2011) and Wright's F_{IS} statistic. Diversity estimates for the wild and cultivated samples were compared for significant differences using the Wilcox test in the R package ggpurb.

We also estimated the relatedness coefficient between each pair of individuals using the relatedness2 function and methodology described by Manichaikul et al. (2010) and implemented in VCFtools. The possible kinship coefficient ranges from over 0.354, corresponding to duplicate/MZ twin, while relationships from 0.177 to 0.354, from 0.0884 to 0.177, and from 0.0442 to 0.0884 correspond to 1st-, 2nd-, and 3rd-degree relationships, respectively; a negative relatedness estimate means that individuals are less related than the average, indicating a genetic structure within the analyzed group of samples (Manichaikul et al., 2010). The relatedness coefficient was estimated for each data set separately.

The relatedness pattern among and within wild and cultivated plants was further evaluated using a pairwise identity-by-state (IBS) allele-sharing analysis as implemented in plink. The IBS for each pair of individuals represents the average proportion of alleles shared at genotyped SNPs and is estimated using wild and cultivated individuals (Purcell et al., 2007; Chang et al., 2015).

To evaluate the impact of a population bottleneck during domestication, we identified alleles that were (1) private to the wild population (lost during domestication), (2) shared between the wild and domesticated population (retained during domestication, or introgressed after domestication), or (3) private to the domesticated population (novel in the domesticated population). For example, the proportion of wild variation retained during domestication was calculated as shared alleles divided by private wild alleles plus shared alleles. The number of private alleles within wild and cultivated samples was determined using the population module in STACKS (Catchen et al., 2013).

Genetic differentiation and population structure

Genetic differentiation between wild and cultivated samples and among individuals was explored using several complementary approaches, such as principal component analysis (PCA), individual genetic clustering, and distance-based relationships. First, we visualized genetic relationships among samples using PCA, an efficient nonmodel-based method for assessing population structure in high-dimensional data sets (Patterson et al., 2006), using the R package SNPrelate (Zheng et al., 2012).

We also used a clustering approach without a priori grouping, as implemented in ADMIXTURE v.1.23 (Alexander et al., 2009). Admixture analysis was run using 2000 bootstraps; clusters were set from 1 to 10 (*K*), with five replicates for each *K* value. The support for different values of *K* was assessed according to the likelihood distribution (i.e., lowest cross-validation error) among replicates and by visual inspection of the coancestry values for each individual. Clustering analysis was performed on the complete data set and exclusively for the wild samples using the same parameters.

To further assess and visualize the genetic relationships among samples, we also constructed a distance tree using the R package poppr (Kamvar et al., 2014). The distance tree was based on the UPGMA algorithm, with 1000 bootstrap replicates to assess branch support. In addition, we estimated pairwise F_{ST} between wild and cultivated samples and among sampling sites using the StAMPP package (Pembleton et al., 2013).

Moreover, we used fineRADstructure (Malinsky et al., 2018) to infer fine-scale genetic structure using a model-based Bayesian clustering approach that groups individuals with high levels of shared coancestry. A "coancestry matrix" of all wild and cultivated individuals, defined as a summary of nearest neighbor haplotype relationships, is required as input and was generated from a VCF file using the population module of STACKS. We subsequently used 1,000,000 Markov chain Monte Carlo

(MCMC) iterations with a burn-in of 1,000,000 and sampling occurring every 1000 iterations. Finally, a tree was constructed with 100,000 hill-climbing iterations, and the results were visualized using the scripts fineradstructureplot.R and finestructurelibrary.R, which are available from https://github.com/edgardomortiz/fineRADstructure-tools.

To better understand the geographic structuring of genetic variation within wild samples, we used the R package vegan (Oksanen et al., 2017) to perform the Mantel test based on 10,000 permutations of the data sets. Geographic distance and pairwise F_{ST} genetic distances among sampling sites were calculated using the Geographic Distance Matrix Generator 1.2.3 (Ersts, 2022) and the R package StAMPP (Pembleton et al., 2013), respectively.

Finally, we used the clustering method implemented in TESS3R (Caye et al., 2015); this methodology considers genetic and geographic data to determine the most probable number of clusters in a geographic space. We tested $K = 1–5$ possible genetic groups with 20 replicates of each K and kept the most supported model (i.e., "best" based on cross-entropy scores) within each of the 20 replicates. Locations on the map were colored according to the resulting dominant ancestry cluster. TESS3R analysis was performed exclusively for the wild samples.

RESULTS

The RADseq of 94 wild and cultivated *A. angustifolia* plants from Oaxaca and Puebla resulted in a total of 579,581,048 paired-end raw reads, with an average of 6,037,302 reads per sample (range: 4,670,374 to 7,376,597). After quality filtering and adapters removal, we obtained 4,297,472 to 6,388,091 high-quality reads per sample. From this data, 511,309 variants were called using the *A. tequilana* transcriptome (Gross et al., 2013) as a reference. Posteriorly, we removed three individuals sampled from one particular plantation that presented high differentiation levels (F_{ST}: 0.24, 95% CI: 0.23–0.25) compared to the rest of the data. These samples, although they were identified as espadín by the plantation owner, were morphologically similar to managed *A. rhodacantha*, so we decided to remove them.

After filtering, our final data set consisted of 91 individuals and 8616 SNPs, with 9.5% missing data and a mean individual depth of 56.2 (7.46) (Appendix S2). All the multilocus genotypes were unique, indicating that no individuals have been inadvertently sampled more than once and that our SNPs data set had good discriminatory power. The average sequencing depth for each individual and missingness per site and per individual are presented in Appendix S2. Interestingly, when wild and cultivated samples were filtered and analyzed separately, considerably fewer SNPs were recovered for the cultivated samples (8395 in cultivated vs. 33,396 in wild plants before filtering for LD; Table 1). Accordingly, the average r^2 was higher for cultivated plants than the wild agaves (0.69 vs. 0.19, respectively) before filtering for LD (Table 1).

TABLE 1 The number of SNPs obtained after filtering for wild and cultivated samples of *A. angustifolia* analyzed separately. N = number of individuals; r^2 = correlation between pairs of loci; H_O = observed heterozygosity; H_E = expected heterozygosity; F_{IS} = Wright's inbreeding index; Fhat3 = inbreeding index; MLH = multilocus heterozygosity. All diversity estimates were presented with respective standard deviations in parenthesis.

Management	N	No. SNPs before quality filtering	No. SNPs after quality filtering and before LD filtering	Mean r^2	No. SNPs after quality filtering and LD filtering ($r^2 < 0.2$)	H_O	H_E	F_{IS}	Fhat3	MLH
Wild	42	511,309	33,396	0.19	20,342	0.19 (0.12)	0.24 (0.13)	0.23 (0.03)	0.23 (0.04)	0.19 (0.007)
Cultivated	49	511,309	8395	0.69	4,096	0.35 (0.33)	0.29 (0.16)	−0.16 (0.22)	0.16 (1.06)	0.34 (0.06)

Genetic diversity

Based on 8616 SNPs, the overall mean observed heterozygosity (H_o: 0.22, SD: 0.17) was lower than the expected heterozygosity (H_E: 0.25, SD: 0.12). The difference between expected and observed heterozygosity was significant even after multiple test corrections (Bartlett's test, all samples K-squared = 774.8, $P < 0.001$). For the wild plants, the mean H_o among 42 individuals was lower (0.17, SD: 0.12) than the expected heterozygosity ($H_E = 0.24$, SD: 0.14), indicating a deficit of heterozygous individuals. Contrasting results were obtained for the cultivated samples that presented significant excess of observed heterozygosity, H_o: 0.27, SD: 0.34, H_E: 0.19, SD: 0.20 (Table 2). When each data set was analyzed separately (wild samples 20,342 SNPs and cultivated samples 4096 SNPs), the main trend remained. We found heterozygosity excess for cultivated samples and heterozygosity deficiency for wild samples (Table 1). However, the heterozygosity of the cultivated samples increased (H_o: 0.35, SD: 0.33, and H_E: 0.29, SD: 0.16), whereas, for the wild samples, the estimates were almost identical to the estimates based on 8616 SNPs (H_o: 0.19, SD: 0.12, and H_E: 0.24, SD: 0.13), indicating that in the case of the wild samples the increase in number SNPs did not affect diversity estimates (Table 1; Appendix S3).

Cultivated samples presented significantly lower average nucleotide diversity (0.04, SD: 0.05) than wild samples (0.07, SD: 0.05). Individual-based multilocus heterozygosity (MLH) was similar to the H_o for wild and cultivated plants, independent of the sample partition strategy (Table 2; Appendix S3). Similar to our other heterozygosity estimates, sMLH was significantly higher in cultivated samples than in wild plants, ranging from 0.71 in one wild sample from Oaxaca to as high as 1.41 in a cultivated plant from Oaxaca. Furthermore, cultivated samples had negative values for F_{IS} and Fhat3. In contrast, wild samples had high and positive inbreeding estimates (Table 2; Appendix S4). Seven cultivated plants had high positive F_{IS}, similar to the values found in wild plants (Appendix S4). Nevertheless, wild samples overall had significantly lower levels of heterozygosity and higher inbreeding than in cultivated samples (Table 2; Appendix S3).

In general, the wild samples were unrelated to each other (Figure 2; Appendix S5), and their mean relatedness was –0.27 (SD 0.13). Nevertheless, we found close genetic relationships between several individuals collected at two sites (Figure 2; Appendix S5). Contrasting results were found for the cultivated individuals (Figure 2; Appendix S5). To avoid biases in relatedness estimation, we removed four additional individuals; these samples had the genetic profile of the wild samples, which resulted in highly negative relatedness with the rest of the cultivated individuals, and therefore skewed the mean estimate. Furthermore, those samples were represented by four individuals from a unique plantation in Puebla state (site PM1). Plants at that site were not under intensive management (i.e., they looked like they had been planted in a wild setting). After removing the samples mentioned above, the mean relatedness for the cultivated samples was 0.24 (SD: 0.27), which suggested that the remaining 45 cultivated individuals had, on average, a first-degree relationship among them (Figure 2; Appendix S5).

Relationships among wild and cultivated agave individuals were also explored by estimating identity-by-state (IBS) allele-sharing values for all pairwise comparisons. The frequency distribution of IBS estimates (Appendix S6) showed that wild samples presented low allele-sharing values within them and with cultivated samples. In contrast, most cultivated individuals fall in the bin from 0.75 to 0.8 (mean: 0.7, SD: 0.28), suggesting that cultivated plants are highly related.

Similarly, private allele analyses revealed that alleles that were private (exclusive) to the wild population (1285) exceeded the number of private alleles in the domesticated population (717). Overall, we estimated that approximately 84% of wild alleles were retained in the cultivated plants. However, we caution that our sampling of wild populations may be incomplete, and some populations may have become locally extinct due to extensive land-use change in Oaxaca state. We suggest that the private alleles found in cultivated espadín may be the product of accumulated somatic mutations (because of clonal propagation for several generations) or may indicate that we have not yet found the ancestral espadín population.

Population genetic structure

The genetic differentiation between wild and cultivated samples was strong, $F_{ST} = 0.24$, and significant (95% CI: 0.235–0.242). The lowest differentiation was found between two cultivated sites from Oaxaca ($F_{ST} = -0.063$) 55 km from each other. In general, all cultivated sites had low F_{ST} among them but were considerably differentiated from wild samples (Appendix S7). The highest differentiation was found between one cultivated site from Oaxaca and a wild site from Puebla (F_{ST} over 0.4). These findings were confirmed by PCA, distance tree, and clustering algorithm implemented in ADMIXTURE (Figure 3).

TABLE 2 Descriptive genomic diversity estimates and standard deviation in parenthesis as determined for wild *Agave angustifolia* and cultivated espadín plants in the states of Oaxaca and Puebla, Mexico, using 8618 SNPs. N = number of samples; MLG = multilocus genotype; MLH = multilocus heterozygosity; sMLH = standardized multilocus heterozygosity; F_{IS} = Wright's inbreeding index; Fhat3 = inbreeding index; Π = standardized nucleotide diversity. The standard deviation for each estimate is presented in parentheses.

Management	N	MLG	MLH	sMLH	F_{IS}	Fhat3	Π
Wild	42	42	0.17 (0.01)	0.77 (0.03)	0.31 (0.03)	0.47 (0.03)	0.07 (0.05)
Cultivated	49	49	0.26 (0.03)	1.19 (0.15)	–0.07 (0.13)	–0.07 (0.19)	0.04 (0.05)

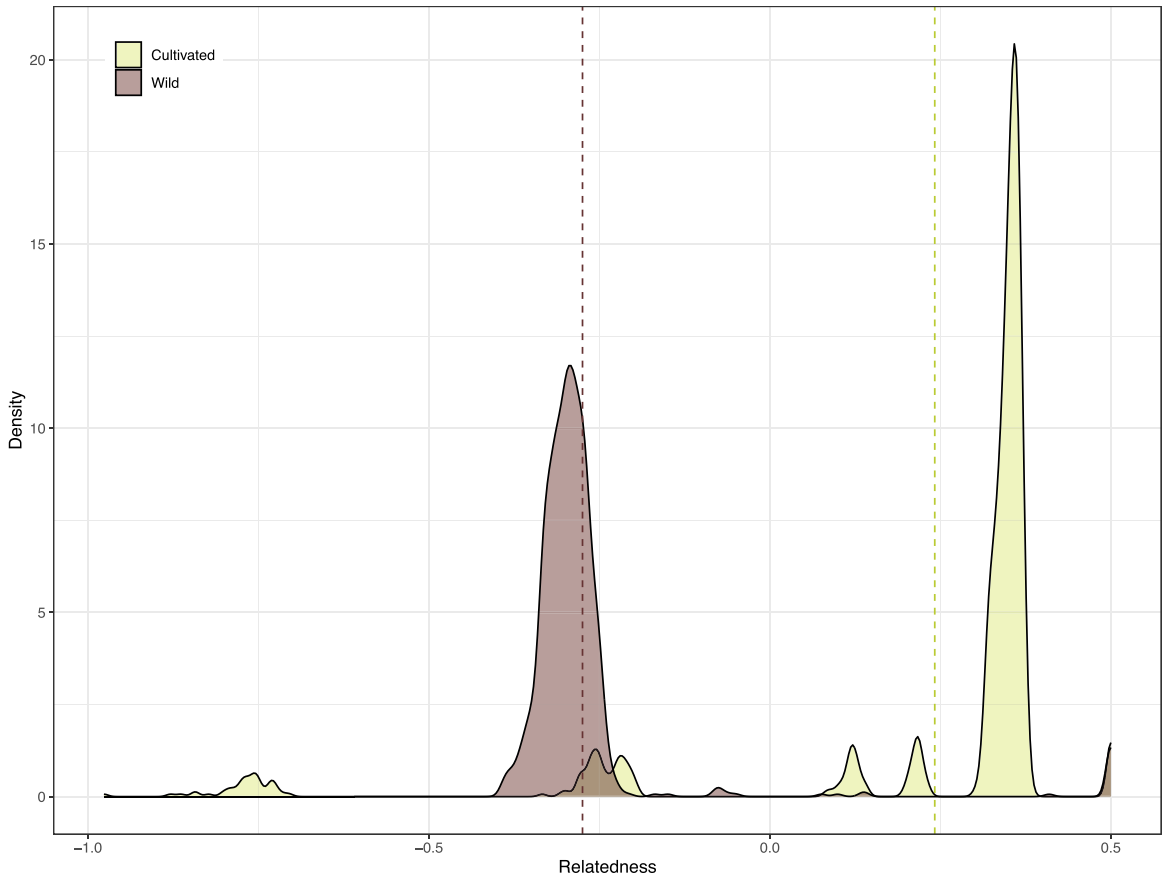

FIGURE 2 Density plots of individual relatedness estimates among wild *A. angustifolia* (brown) and cultivated espadín (green) plants from the states of Puebla and Oaxaca, Mexico. The methodology described by Manichaikul et al. (2010) was used. The dashed lines represent the mean of the respective data set. The relatedness was estimated separately for wild and cultivated plants.

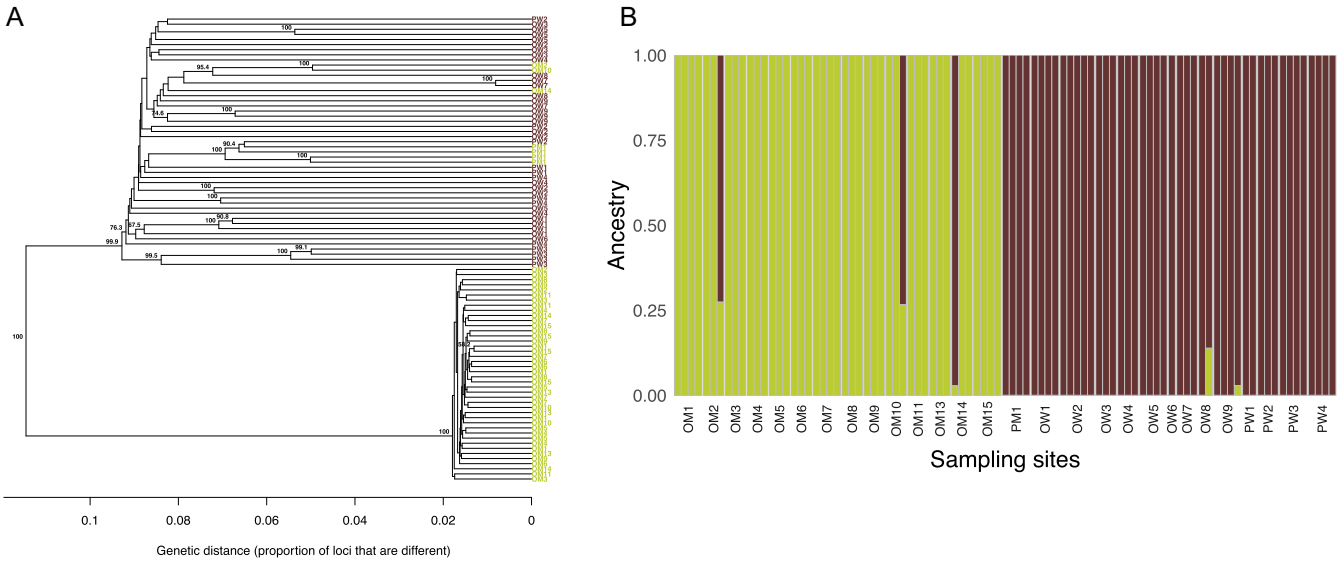

FIGURE 3 Population genetic structure of the wild *A. angustifolia* and cultivated espadín samples from Puebla and Oaxaca states, Mexico, based on 8618 SNPs. (A) UPGMA tree based on Nei's genetic distance and (B) bar plot of the individual assignment probabilities (vertical axis) for the most likely number of genetic clusters $K = 2$ inferred using the program ADMIXTURE. Samples were clustered according to the site as specified in Appendix S1. Green: cultivated samples; brown: wild samples.

Estimates of the cross-validation error (CVE) in the ADMIXTURE analysis showed that the model fit was optimized at $K = 2$ (Appendix S8). Variation in CVE among the replicates was relatively low across all K-values (Appendix S8). Wild populations were clearly separated from the cultivated plants (at $K = 2$), except for one cultivated site in Puebla (PM1) that genetically belonged to the wild samples, as commented above. ADMIXTURE was also able to distinguish three cultivated individuals with mixed ancestry collected at three different plantations in Oaxaca. Nevertheless, the cultivated plants were generally consistently distinct from the wild samples (Figure 3). A UPGMA tree analysis (Figure 3) performed with all the individuals and 8618 SNPs showed strong agreement with the clustering analysis results. However, the tree allowed better insight into the within-group relationships. For example, genetic differentiation within cultivated samples was extremely low, with relatively short branches of similar length among samples (Figure 3). Wild samples, on the other hand, showed a clustering based on the sampling site, suggesting fine-scale structuring within them.

Similar results were obtained with the PCA (Appendix S9); the cultivated samples formed a small cluster strongly differentiated from the wild plants (EV1, 22.8%), whereas the cloud of wild samples was dispersed, indicating possible population genetic structure (Appendix S9). Six cultivated samples were also separated from the main cloud and genetically similar to wild ones. Four of those plants were represented by one cultivated site in Puebla (PM1), and two individuals were from two intensively managed plantations. Interestingly, the owner recognized one of these samples (site OM14) as a "lumbre" variety. This plant was morphologically similar to wild *A. angustifolia* found in the area but had fewer spines and a reddish color.

Next, to infer fine-scale population structure via shared coancestry among individuals, we used a model-based Bayesian clustering approach implemented in fineRADstructure. The resulting coancestry matrix and cladogram shown in Figure 4 largely confirmed the results of the previous analyses. First, wild and cultivated plants were resolved as distinct, well-supported groups, except for seven samples. Four of those belonged to one managed population PM1, which clustered with wild samples from Puebla and northern Oaxaca; another belonged to the managed site (OM14) and clustered with northern Oaxaca wild samples. Yet, the other two samples collected at two different plantations in Oaxaca (OM2 and AM10) formed a separate group but were more related to the wild samples. Finally, substantially higher levels of shared coancestry were found among cultivated individuals, reflecting lower levels of genetic diversity (Figure 4).

An interesting pattern emerged when wild samples were analyzed separately (Appendices S10 and S11). Although ADMIXTURE analysis indicated that the best K for wild samples was one (Appendix S10), when the individual ancestry was plotted (Appendix S11), we found several sampling sites, mainly from the southern part of Oaxaca, presenting distinct ancestry. Nevertheless, each site had one or more individuals of mixed ancestry (Appendix S11).

A Mantel test indicated that there was not a significant relationship between genetic distance (pairwise F_{ST}) and geographic distance among wild *A. angustifolia* (Appendix S12). A spatially explicit TESS analysis, in accordance with ADMIXTURE results, indicated that southern wild samples (south to Oaxaca city) were genetically separated from the northern group that comprised all samples from Puebla state and samples from north to Oaxaca city. Nevertheless, the majority of individuals presented mixed ancestry (Appendix S13).

DISCUSSION

The genetic resources maintained in the crop wild relatives provide the most promising option for improving domesticated species and conserving and managing the remnant wild populations (Bohra et al., 2022). Furthermore, this genetic variation may be crucial for the gene pools depleted by intensive breeding and artificial selection (Dempewolf et al., 2017). Therefore, if the latest genotyping tools are used to describe the standing genetic variation in the wild populations and the remaining diversity in cultivated varieties (Andrews et al., 2016), using this variation, the eroded diversity in crop species might be improved and genomically informed management strategies can be implemented.

In this study, we used over 8000 SNPs to determine the population structure and genomewide diversity of wild *A. angustifolia* and cultivated espadín individuals from Oaxaca and Puebla, Mexico. The resulting data set provides unique insight into the genomic consequences of intensive management and clonal propagation in an economically important agave species. We found that cultivated espadín plants are genetically different from wild *A. angustifolia* individuals of the Oaxaca and Puebla states. Moreover, we found that clonal propagation seems to be the principal reproduction strategy for espadín, with its diversity affected accordingly (i.e., an increase in heterozygosity at the individual level but a reduction in genotypic diversity at the population level). This information will be important for developing conservation and management strategies to preserve genetic resources in this agave and preventing rapid, unsustainable destruction of wild populations.

Genetic diversity in wild *A. angustifolia* and clonally propagated espadín

The significance of genetic diversity for wild plant and animal species is indisputable (Allendorf et al., 2022). However, diversity is also crucial for cultivated plant species because it can be used for future crop improvement and adaptation to changing climate and evolving pests (Khoury et al., 2021; Swarup et al., 2021). Moreover, this genetic

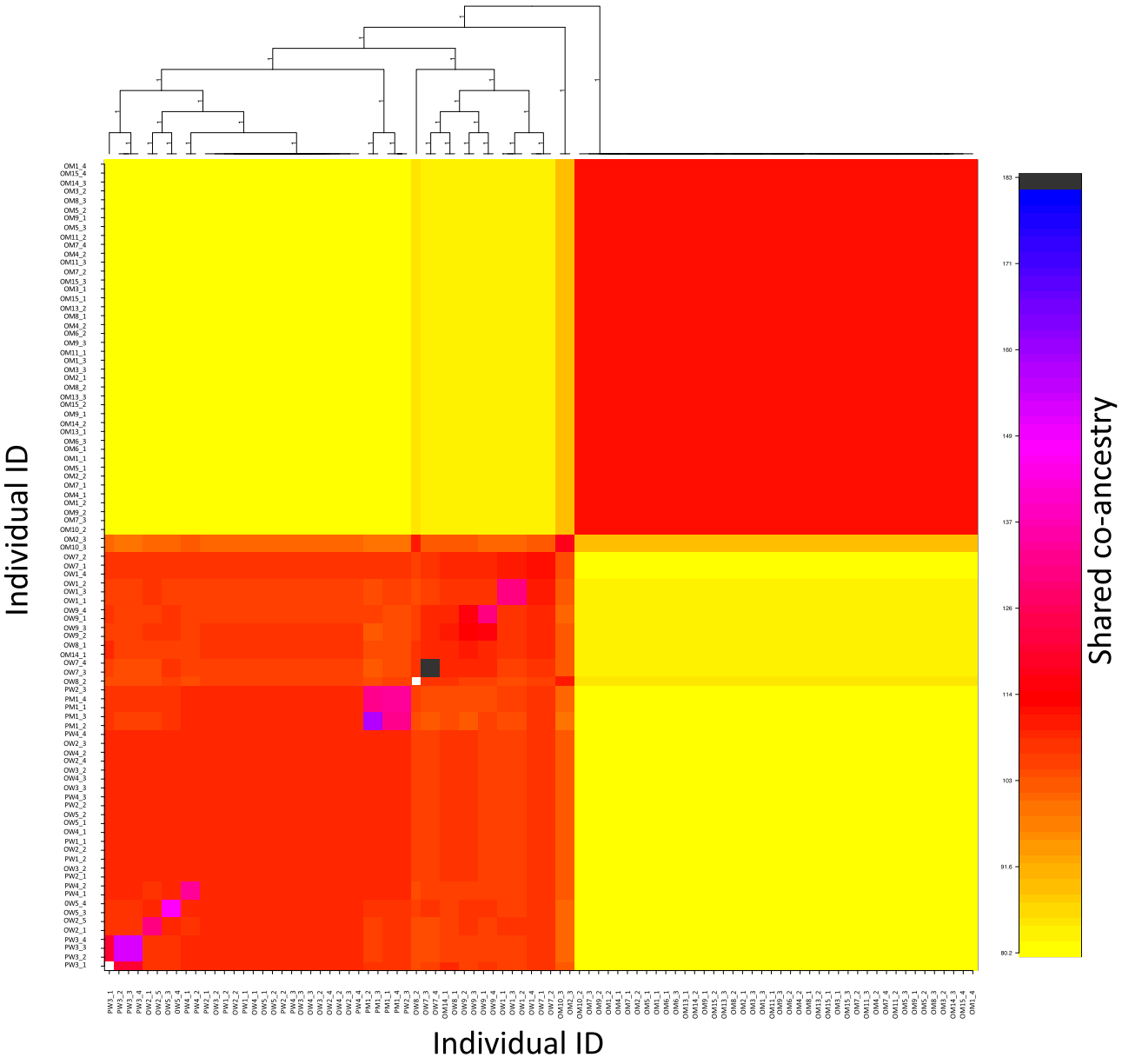

FIGURE 4 Results of the fineRADstructure analysis of the genomic data for wild and cultivated samples of *A. angustifolia*, based on 8618 SNPs. The heat map depicts variation in pairwise coancestry among individuals according to the color scale on the right, which indicates relatedness between individuals: yellow, low relatedness; blue/black, high relatedness. The values next to the dendrogram branches are posterior probabilities of population assignments: they reflect the proportion of MCMC samples in which the individuals in question formed a specific group.

diversity can be preserved in germplasm collections (e.g., seeds or plant nurseries) that eventually can be used to develop new crop varieties resistant to pests and diseases and that are better adapted to different climates, soils, and agricultural systems (Gutaker et al., 2019; Azhar and Wani, 2021). However, the genetic diversity of the cultivated crop species is usually reduced because of the evolutionary events related to domestication and development of the local landraces (Gaut et al., 2018; Khoury et al., 2021), such as genetic drift and artificial selection (Doebley et al., 2006). Accordingly, an overall reduction of genetic diversity has been reported in various cultivated plant species (Doebley et al., 2006; Swarup et al., 2021) and managed agaves (Vargas-Ponce et al., 2009; Trejo et al., 2018).

In clonally propagated crops, once sexual reproduction becomes rare, mitotic processes such as the accumulation of somatic mutations (Balloux et al., 2003) may become increasingly important (Foster and Aranzana, 2018; Zheng et al., 2022). In clonally propagated agave varieties, this novel variation will accumulate over time and may be detectable as increased observed heterozygosity (Cabrera-Toledo et al., 2022; Ruiz-Mondragon et al., 2022). Accordingly, observed heterozygosity (H_O) levels in the genome of wild *A. angustifolia* were significantly lower ($H_O = 0.17$) compared to cultivated samples

($H_O = 0.27$). Moreover, wild samples were characterized by homozygote excess, suggesting frequent inbreeding. Interestingly, the observed heterozygosity of the wild individuals from Oaxaca and Puebla was also lower than those reported for *A. angustifolia* var. *pacifica* from Sonora ($H_O = 0.22$) (Klimova et al., 2022) and in wild *A. angustifolia* from Jalisco ($H_O = 0.28$) (Cabrera-Toledo et al., 2022). For all wild *A. angustifolia* populations analyzed to date and in other agave species (e.g., Cabrera-Toledo et al., 2020; Figueredo-Urbina et al., 2021; Cabrera-Toledo et al., 2022), inbreeding has been reported, suggesting that it may be a hallmark of agave as a group.

High levels of inbreeding increase homozygote frequencies; thus, individuals will express recessive or partially recessive deleterious mutations and, consequently, suffer from inbreeding depression, expressed as lower survival and reduced fertility (Charlesworth et al., 2009). Therefore, the considerable inbreeding levels in wild agave populations demand an explanation. An interesting hypothesis that may explain high levels of inbreeding in natural plant populations was proposed by Robertson (1964). He suggested that when a population splits into sublines with occasional mixing of these subpopulations, the overall rate of inbreeding would be minimized. Nevertheless, within the sublines, inbreeding and genetic drift would be increased. In support of this hypothesis, we found that the relatedness among wild *A. angustifolia* was very low. However, within some wild sampling sites, we found levels of relatedness ranging from 0.08 to 0.41. Moreover, the possibility of clonal reproduction, clonal longevity, limited seed dispersal, a high number of inbred progeny (Trame et al., 1995), and marked discontinuity in the distribution of natural populations may also contribute to the observed pattern of heterozygosity. Further population genetic studies are needed to precisely determine the causes and consequences of inbreeding on these species in the wild.

Although agaves have been extensively studied, historically, the main focus has been concentrated on *A. tequilana*, the species used for tequila production (Gil-Vega et al., 2006; Rodríguez-Garay et al., 2009; Cabrera-Toledo et al., 2022; Ruiz-Mondragon et al., 2022). In contrast, other species and varieties have received less attention (Félix-Valdez et al., 2015; Klimova et al., 2022). One previous study, based on AFLPs, found that the genetic diversity—estimated as Nei's heterozygosity—of the cultivated espadín in Oaxaca was similar to that observed in wild *A. angustifolia* populations (Rivera-Lugo et al., 2018). Moreover, several recent SNP-based studies reported comparatively high levels of heterozygosity and a negative inbreeding index in cultivated *A. angustifolia* in Jalisco (Cabrera-Toledo et al., 2022) and in intensively managed *A. tequilana* (Ruiz-Mondragón et al., 2022).

A high level of clonal propagation increases heterozygosity (Balloux et al., 2003; Yu et al., 2020). Long-established clones of several crop species have been reported to be highly heterozygous, for example, cassava landraces (Pujol et al., 2005), vanilla (Favre et al., 2022), grapes (Aradhya et al., 2003; Laucou et al., 2018), and potatoes (Manrique-Carpintero et al., 2018). Moreover, LD in clonally propagated samples is also expected to be higher than in their wild relatives (Balloux et al., 2003; De Meeûs and Balloux, 2004). Li et al. (2018), Hyten et al. (2007), and Hu et al. (2021) found a marked increase in LD decay in cultivated potatoes, soybean, and cassava compared to their wild conspecifics. These results are expected because the decrease in the rate of sexual reproduction and recombination ultimately will generate nonrandom associations between loci. Therefore, high linkage disequilibrium is expected to be found in clonal plants (Tibayrenc and Ayala, 1991) because clonality mimics complete physical linkage over the entire genome (Balloux et al., 2003; De Meeûs and Balloux, 2004).

Interestingly, if the levels of heterozygosity are due to the accumulation of mutations during clonal propagation rather than from demographic and reproductive history, only SNP markers will detect such mutations because microsatellites assess length variations in the number of microsatellite repeats. This difference between SNPs and microsatellites may have caused the discrepancies in diversity estimates found between studies that used different markers (i.e., Vargas-Ponce et al., 2009; Trejo et al., 2018). Therefore, the high genomewide observed heterozygosity levels detected in cultivated clonal espadín may be explained in part by decades of intense clonal propagation (Zizumbo-Villarreal and Colunga-GarcíaMarín, 2007) and accumulation of somatic mutations (Balloux et al., 2003; Foster and Aranzana, 2018). Moreover, some espadín individuals are reported to be triploid (Rivera-Lugo et al., 2018), which could have contributed to the increased levels of heterozygosity.

We suggest that the diversity previously detected in cultivated *A. angustifolia* (Cabrera-Toledo et al., 2022) using GBS could correspond, at least in part, to somatic mutations accumulated in clones (Ruiz-Mondragón et al., 2022). We also argue that levels of heterozygosity may be an inaccurate index to estimate diversity in cultivated clonal crops, and any management and conservation recommendations based on this estimate alone should be taken with caution. For example, nucleotide diversity was almost twice as high in the wild *A. angustifolia* than in the cultivated samples. Moreover, wild samples presented more private alleles and low levels of relatedness within and among sites. Furthermore, due to the extremely high LD and a high number of low-frequency alleles found in cultivated *A. angustifolia* individuals, when wild and cultivated samples were analyzed separately, more SNPs were detected in wild samples (over 20,000) compared to cultivated samples (4000). These findings are in accordance with the apparently contradictory notion that although clonal reproduction would increase heterozygosity, at the same time, it will decrease genotypic diversity (Balloux et al., 2003; Klekowski, 2003). Therefore, at least in the clonally propagated agave species and specifically when SNPs are used, the heterozygosity levels alone should never be used to decide whether the genetic diversity is high or low. Instead, we suggest using additional and complementary diversity estimates, such as unbiased nucleotide diversity.

Genetic differentiation within the *A. angustifolia* in Oaxaca and Puebla

We applied RADseq to explore the genetic relationships and structure of the intensively managed espadín and its wild relatives from Oaxaca and Puebla. It has been noted that, in general, the genetic differentiation among wild populations of *A. angustifolia* is relatively low (Eguiarte et al., 2013; Klimova et al., 2022). Moreover, a recent study showed that the genetic relationships among *A. tequilana*, wild *A. angustifolia* from Jalisco state, and *A. rhodacantha* are close and that these species do not clearly differentiate into separate genetic clusters (Cabrera-Toledo et al., 2022). These findings have been explained by extensive gene flow promoted by the primary agave pollinators, i.e., nectar-feeding bats of the genus *Leptonycteris* and other genera (Molina-Freaner and Eguiarte, 2003), and the possibility that, at least in traditional cultivation settings, farmers still readily incorporate wild plants. Consistent with the aforementioned studies, we found almost no genetic differentiation among 14 sampling sites of wild *A. angustifolia* (e.g., ADMIXTURE results indicate that the most supported value of *K* for wild samples was 1).

The population differentiation values F_{ST} for wild populations ranged from −0.0005 between nearby sites separated by 3.1 km to 0.17 between sites separated by 207 km, suggesting that there may be a spatial component, such as isolation by distance, involved in the population differentiation of *A. angustifolia*. Accordingly, the TESS analysis separated southern and northern localities into two genetic groups; nevertheless, almost all individuals presented mixed ancestry, and there was no isolation by distance according to the Mantel test. These findings are similar to those reported for *A. angustifolia* var. *pacifica* in the northern Mexican state of Sonora (Klimova et al., 2022), where population differentiation between coastal vs. inland populations was low. Future studies should analyze the population genetic structure of the wild *A. angustifolia*, covering the entire distribution of the species and including populations at different elevations, environments, and broad geographic distances.

Genetic differentiation (F_{ST}) between crops and their wild relatives varies widely among species, from as low as 0.006 for the squash *Cucurbita pepo*, to moderate differentiation for maize and grapes ($F_{ST} = 0.11$ and 0.12, respectively) and high ($F_{ST} = 0.35$) for soybean (Hufford et al., 2012; Zhou et al., 2015; Marrano et al., 2018; Martínez-González et al., 2021). For agave species, the divergence between wild and cultivated conspecifics seems to be related to time under management and management intensity (Cabrera-Toledo et al., 2022; Klimova et al., 2022). When wild samples were compared to the cultivated espadín plants, we found strong differentiation ($F_{ST} = 0.24$), with almost no admixed individuals. In contrast, virtually no genetic differentiation was found among the cultivated plants, except at one cultivated site from Puebla. There are several possible explanations for these findings. First, espadín might be closely related to cultivated *A. tequilana* or *A. rhodacantha* (Rivera-Lugo et al., 2018). It is possible that relatively recently, farmers from Oaxaca brought ramets or seeds from a few individuals collected from cultivated *A. tequilana* or other cultivated *A. angustifolia* or *A. rhodacantha* varieties from Jalisco or from elsewhere and grew them on their parcels. The cultivated populations were then maintained from generation to generation through clonal propagation, interchange among farmers, and minimal sexual reproduction.

On the other hand, the observed differentiation pattern fits into a simple scenario of domestication that considers a demographic pattern of limited sampling from specific wild populations that may be already extinct and periodic bottlenecks for managed plants. Vegetative propagation may have further increased the genetic homogeneity of espadín at the population level and increased differentiation from the wild conspecifics. Similar results were found in some cultivated varieties of *A. angustifolia* in Jalisco (Trejo et al., 2018; Cabrera-Toledo et al., 2022). To confidently answer this question, espadín samples should be compared to other cultivated *A. angustifolia* varieties and the wild *A. angustifolia* from its entire distribution range. Future studies, including samples from all of Mexico (cultivated and wild), may help to shed light on the genetic origin of the cultivated espadín.

Genetic background of cultivated espadín

Combining vegetative propagation and sexual reproduction is relatively common in angiosperms (Silvertown, 2008; Barrett, 2015). However, both propagation strategies have different ecological and genetic trade-offs (Fischer and Van Kleunen, 2001; Klekowski, 2003; Barrett, 2015). In general, clonal reproduction may help to increase reproductive assurance when sexual reproduction is difficult and preserve locally adapted genotypes while avoiding the relatively high costs and uncertainties (e.g., availability of suitable conspecific partners and adequate pollinators) of sexual reproduction (Olofsson and Lundberg, 2007; Silvertown, 2008). In the case of agave, this advantage may be sizeable considering the production costs of their very large inflorescences, many flowers, large amounts of nectar and pollen, and the dependence on bats and other animals for pollination (Arizaga and Ezcurra, 1995).

Many agave species can propagate vegetatively through aerial bulbils, ground-level basal shoots, and rhizomes or sexual reproduction by seeds (Arizaga and Ezcurra, 1995; Szarek et al., 1996). This characteristic has been considered an adaptation to life in the harsh conditions of arid lands (Gentry, 1972). Agave producers have used clonal propagation to increase the productivity of many agave species. Indeed, some agave species used for mezcal and related beverage production, in particular, *A. tequilana, A. angustifolia* var. *pacifica*, and agave espadín, are capable of clonal reproduction. Clonal reproduction offers many advantages for farmers, such as homogeneous progeny with fixed desirable characteristics,

easily detectable favorable mutations, limited or null gene flow, and easier overall propagation (McKey et al., 2010).

Based on relatedness and heterozygosity patterns, we suggest that vegetative propagation is the predominant form of reproduction in cultivated espadín, but additional field studies are still needed. This practice seems to contribute to the maintenance of high levels of heterozygosity in espadín individuals and has enabled the preservation of clones with unique and desirable traits. However, we also suggest that it has reduced the possibility of crossing and breeding new varieties. In support of this scenario, we found that over 90% of the espadín samples from 16 different plantations had the equivalent of first-degree relationships. Moreover, the pairwise IBS distribution also indicates many higher-order relationships within the sampled espadín individuals and that most individuals shared large stretches of DNA.

Our findings contrast with those for *A. angustifolia* var. *pacifica* cultivated for bacanora (a regional type of mezcal) production in Sonora state, where cultivated plants have high genetic similarity to their wild counterparts, with low relatedness. These results were attributed to a very recent start of cultivation of the variety; thus, no single clone has been extensively planted in the state (Klimova et al., 2022). Similarly, our findings contrast with those for cultivated *A. angustifolia* from Jalisco, where several genetic groups corresponding to different cultivated varieties were reported (Vargas-Ponce et al., 2009; Cabrera-Toledo et al., 2022). Nevertheless, our results are similar to those for the intensively managed, vegetatively propagated tequila agave (*A. tequilana*; Ruiz-Mondragon et al., 2022) and grape cultivars (Zhou et al., 2017).

Thus, the genetic structure of cultivated espadín in Oaxaca can be largely understood as one closely related family group. We propose that this relatedness structure resulted from a small number of individuals that have been vegetatively propagated for many decades. These findings indicate a unique genetic origin of cultivated espadín and that only a small number of the possible genetic combinations have been explored. Our findings are, therefore, consistent with a scenario of recent rapid domestication (i.e., an immediate switch to clonal propagation using favorable genotypes identified from wild populations or another related cultivated species). Under this scenario, the plant's heterozygosity would remain unchanged and later may increase with the accumulation of somatic mutations, which was apparently observed.

We, therefore, suggest that the widespread use of intensive management, clonal propagation, and prevention of flowering and pollination during espadín breeding may have a doubtful advantage. In the short term, the production of mezcal benefits from the control over genetic variability and homogeneous phenotypes that vegetative propagation offers. Nevertheless, clonal propagation has also discouraged the breeding of new varieties and may increase the susceptibility of the cultivated plants to different diseases and pests, as previously noted in agaves (Fucikovsky, 2001; Dalton, 2005; Ramírez-Ramírez et al., 2017; Rubio-Ríos et al., 2019).

On a positive note, we found three samples that presented mixed ancestry between wild and cultivated samples from managed sites in Oaxaca (OM2, OM10, and OM14). Interestingly, one of these samples was indicated to belong to the lumbre variety rather than espadín, which is of interest to some producers (personal communications with plantation owners). These findings suggest that although restricted, there may be a genetic interchange between wild and cultivated agave plants.

Conservation implications

The demand for mezcal-like beverages has been growing fast and shows no sign of decreasing. However, activities related to mezcal production are already threatening the natural population of over 40 agave species, changing traditional agricultural land-use, deforesting pristine natural lands, and decreasing biodiversity (Delgado-Lemus et al., 2014; Goettsch et al., 2021; Tetreault et al., 2021; Lira et al., 2022). These practices degrade and reduce the natural communities and endanger the traditional landraces of many species in these high diversity areas of Mexico. The mezcal boom also encourages intensive management and clonal reproduction of a few genotypes. Developing an environmentally sustainable mezcal industry requires community involvement and data-informed management of wild and cultivated plants and, eventually, perhaps law enforcement.

First, agave populations should be carefully evaluated in their whole distribution range, describing conservation status and vulnerability to climate change and human impact to ensure the long-term conservation of the diversity found in the wild *A. angustifolia* and related species. Then the genetic resources—both in situ and ex-situ—should be preserved, and large and small-scale mezcal producers should be encouraged to use germplasm collections (e.g., seeds or plant nurseries), considering their needs and cultural specificities, local characteristics, and ecological and climatic adaptation of the different accessions. Moreover, the development and standardization of the genetic methodology that can distinguish different species and varieties of agave plants under cultivated and wild conspecifics should be implemented. Finally, the collaboration between experts in genomic marker-assisted breeding approaches and mezcal producers should be carried out to generate improved cultivars better adapted to the local conditions where the plants are grown.

CONCLUSIONS

Considerable heterozygosity found in espadín has apparently been generated by decades of clonal propagation. This heterozygosity is contained within a network of highly related individuals that display high linkage disequilibrium and a high number of low-frequency alleles. We found that first-degree relationships are common among cultivated

samples, even from geographically distant regions, but nearly absent among wild *A. angustifolia*.

Our results suggest that although substantial heterozygosity is present in espadín plants, there has been limited exploration and explanation for the total genetic diversity found in wild plants. The intensive management, vegetative propagation, and prevention of flowering in espadín agave represent a doubtful advantage. Thus, although clonal propagation has provided a short-term benefit by ensuring a breeding genotype of interest, it also prevented the generation of different cultivars through crosses or an influx of diversity found in wild plants that may be better adapted to the local conditions where the plants are cultivated.

Moreover, besides the well-understood dangers of growing plants with a very narrow genetic base—and thus making the cultivar very susceptible to diseases, pests, and global climate change—the long-term sustainability of the agave and mezcal industries may be additionally jeopardized by an accumulation of deleterious alleles, and therefore both producers and researchers should consider exploiting the immense natural genetic variability of agave.

AUTHOR CONTRIBUTIONS

Conceptualization, L.E.E., and A.K.; data curation, A.K., K.Y.R.M., and E.A.P.; formal analysis, A.K., and K.Y.R.M.; funding acquisition, L.E.E., R.L., and A.V.; investigation, L.E.E., A.K., K.Y.R.M.; methodology, L.E.E., A.K., K.Y.R.M., project administration, L.E.E., E.A.P., R.L., and A.V.; resources, L.E.E., E.A.P.; supervision, L.E.E.; validation, A.K., L.E.E., and E.A.P.; writing original draft, A.K.; review and editing, L.E.E., A.K., K.Y.R.M., E.A.P. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENTS

We thank Rosalinda Tapia, and Marco Tulio Solano de la Cruz for specialized technical support in the laboratory and Guillermo Sanchez de la Vega and Tania Sánchez Ortiz for crucial help in field sampling, all from Instituto de Ecología, UNAM. We also thank Alejandra Moreno-Letelier (Instituto de Biología, UNAM) for the information on the standardization of GBS in other *Agave* species. We are particularly thankful to all the maguey and mezcal producers and families in the municipalities visited. This research was funded by PAPIIT project IG200122 awarded to Luis E. Eguiarte and R. Lira, and PAPIIT IN212321 awarded to Daniel Piñero (Instituto de Ecología, UNAM), by the National Autonomous University of Mexico, UNAM, and by the project 319061, PRONACES, CONACyT "Auge mezcalero y deudas de extinción: investigación interdisciplinaria hacia la sustentabilidad", awarded to Alfonso Valiente Banuet. A postdoctoral research fellowship for A.K. was provided by DGAPA/UNAM IECO/S.ACAD/136/2021. K.Y.R.M is a doctoral student from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and has Ph.D. CONACyT fellowship No. CVU: 739748. We thank the anonymous reviewers for their valuable comments that improved the quality of our paper. We thank Beth E. Hazen for the time and expertise devoted to editing our paper.

DATA AVAILABILITY STATEMENT

The data are deposited in the Sequence Read Archive of NCBI under the BioProject number PRJNA985067.

ORCID

Anastasia Klimova http://orcid.org/0000-0002-1502-2910
Luis E. Eguiarte http://orcid.org/0000-0002-5906-9737

REFERENCES

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655–1664.

Allendorf, F. W., C. W. Funk, S. N. Aitken, M. Byrne, and G. Luikart [eds.]. 2022. Conservation and the genomics of populations, 3rd ed. Oxford University Press, NY, NY, USA.

Andrews, K. R., J. M. Good, M. R. Miller, G. Luikart, and P. A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* 17: 81–92.

Aradhya, M. K., G. S. Dangl, B. H. Prins, J. M. Boursiquot, M. A. Walker, C. P. Meredith, and C. J. Simon. 2003. Genetic structure and differentiation in cultivated grape, *Vitis vinifera* L. *Genetical Research* 81: 179–192.

Arellano-Plaza, M., J. B. Paez-Lerma, N. O. Soto-Cruz, M. R. Kirchmayr, and A. Gschaedler Mathis. 2022. Mezcal production in Mexico: between tradition and commercial exploitation. *Frontiers in Sustainable Food Systems* 6: 832532.

Arizaga, S., and E. Ezcurra. 1995. Insurance against reproductive failure in a semelparous plant: bulbil formation in *Agave macroacantha flowering stalks*. *Oecologia* 101: 329–334.

Arizaga, S., and E. Ezcurra, E. 2002. Propagation mechanisms in Agave macroacantha (Agavaceae), a tropical arid-land succulent rosette. *American Journal of Botany* 89: 632–641.

Arnaud-Haond, S., F. Alberto, S. Teixeira, G. Procaccini, E. A. Serrão, and C. M. Duarte. 2005. Assessing genetic diversity in clonal organisms: low diversity or low resolution? Combining power and cost efficiency in selecting markers. *Journal of Heredity* 96: 434–440.

Azhar, M. T., and S. H. Wani [eds.]. 2021. Wild germplasm for genetic improvement in crop plants. Academic Press, Cambridge, MA, USA.

Bai, Y., and P. Lindhout. 2007. Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? *Annals of Botany* 100: 1085–1094.

Balloux, F., L. Lehmann, and T. de Meeûs. 2003. The population genetics of clonal and partially clonal diploids. *Genetics* 164: 1635–1644.

Barrett, S. C. H. 2015. Influences of clonality on plant sexual reproduction. *Proceedings of the National Academy of Sciences, USA* 112: 8859–8866.

Bitocchi, E., D. Rau, E. Bellucci, M. Rodriguez, M. L. Murgia, T. Gioia, D. Santo, et al. 2017. Beans (Phaseolus ssp.) as a Model for Understanding Crop Evolution. *Frontiers in Plant Science* 8: 722.

Bohra, A., B. Kilian, S. Sivasankar, M. Caccamo, C. Mba, S. R. McCouch, and R. K. Varshney. 2022. Reap the crop wild relatives for breeding future crops. *Trends in Biotechnology* 40: 412–431.

Cabrera-Toledo, D., E. Mendoza-Galindo, N. Larranaga, A. Herrera-Estrella, M. Vásquez-Cruz, and T. Hernández-Hernández. 2022. Genomic and morphological differentiation of spirit producing *Agave angustifolia* traditional landraces cultivated in Jalisco, Mexico. *Plants* 11: 2274.

Cabrera-Toledo, D., O. Vargas-Ponce, S. Ascencio-Ramírez, L. M. Valadez-Sandoval, J. Pérez-Alquicira, J. Morales-Saavedra, and O. F. Huerta-Galván. 2020. Morphological and genetic variation in monocultures, forestry systems and wild populations of *Agave maximiliana* of

western Mexico: implications for its conservation. *Frontiers in Plant Science* 11: 817.

Callen, E. 1967. Analysis of the Tehuacan coprolites. *In* D. S. Byers [ed.], The prehistory of the Tehuacán Valley, 1, 261–289. University of Texas Press, Austin, TX, USA.

Callen, E. O. 1965. Food habits of some pre-Columbian Mexican Indians. *Economic Botany* 19: 335–343.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* 22: 3124–3140.

Caye, K., T. M. Deist, H. Martins, O. Michel, and O. François. 2015. TESS3: fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources* 16: 540–548.

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4: s13742-015-0047-8.

Charlesworth, D., and J. Willis. 2009. The genetics of inbreeding depression. *Nature Reviews Genetics* 10: 783–796.

Clement, C. R., A. Casas, F. A. Parra-Rondinel, C. Levis, N. Peroni, N. Hanazaki, L. Cortés-Zárraga, et al. 2021. Disentangling domestication from food production systems in the neotropics. *Quaternary* 4: 4.

Dalton, R. 2005. Saving the agave. *Nature* 438: 1070–1071.

COMERCAM. 2022. Council for Mescal Regulation. El Mezcal: La Cultura Líquida de México. Website: https://comercam-dom.org.mx [accessed on 1 February 2022].

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

De Meeûs, T., and F. Balloux. 2004. Clonal reproduction and linkage disequilibrium in diploids: a simulation study. *Infection, Genetics and Evolution* 4: 345–351.

Delgado-Lemus, A., I. Torres, J. Blancas, and A. Casas. 2014. Vulnerability and risk management of *Agave* species in the Tehuacán Valley, México. *Journal of Ethnobiology and Ethnomedicine* 10: 53.

Dempewolf, H., G. Baute, J. Anderson, B. Kilian, C. Smith, and L. Guarino. 2017. Past and future use of wild relatives in crop breeding. *Crop Science* 57: 1070–1082.

Denham, T., H. Barton, C. Castillo, A. Crowther, E. Dotte-Sarout, S. A. Florin, J Pritchard, et al. 2020. The domestication syndrome in vegetatively propagated field crops. *Annals of Botany* 125: 581–597.

Doebley, J. F., B. S. Gaut, and B. D. Smith. 2006. The molecular genetics of crop domestication. *Cell* 127: 1309–1321.

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

Eaton, D. A. R., and I. Overcast. 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* 36: 2592–2594.

Eguiarte, L. E., E. Aguirre-Planter, X. Aguirre, R. Colín, A. González, M. Rocha, E. Scheinvar, et al. 2013. From isozymes to genomics: population genetics and conservation of *Agave* in México. *Botanical Review* 79: 483–506.

Eguiarte, L. E., O. A. Jiménez Barrón, E. Aguirre-Planter, E. Scheinvar, N. Gámez, J. Gasca-Pineda, G. Castellanos-Morales, et al. 2021. Evolutionary ecology of *Agave*: distribution patterns, phylogeny, and coevolution (an homage to Howard S. Gentry). *American Journal of Botany* 108: 216–235.

Ersts, P. J. 2022. Geographic distance matrix generator (v.1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation [online]. Website: http://biodiversityinformatics.amnh.org/open_source/gdmg [accessed 19 November 2022].

Favre, F., C. Jourda, M. Grisoni, Q. Piet, R. Rivallan, J. B. Dijoux, J. Hascoat, et al. 2022. A genome-wide assessment of the genetic diversity, evolution and relationships with allied species of the clonally propagated crop *Vanilla planifolia* Jacks. ex Andrews. *Genetic Resources and Crop Evolution* 69: 2125–2139.

Félix-Valdez, L. I., O. Vargas-Ponce, D. Cabrera-Toledo, A. Casas, A. Cibrian-Jaramillo, and L. de la Cruz-Larios. 2015. Effects of traditional management for mescal production on the diversity and genetic structure of *Agave potatorum* (Asparagaceae) in central Mexico. *Genetic Resources and Crop Evolution* 63: 1255–1271.

Figueredo, C. J., A. Casas, A. González-Rodríguez, J. M. Nassar, P. Colunga-GarcíaMarín, and V. Rocha-Ramírez. 2015. Genetic structure of coexisting wild and managed agave populations: implications for the evolution of plants under domestication. *AoB Plants* 7: plv114.

Figueredo-Urbina, C. J., G. D. Álvarez-Ríos, M. A. García-Montes, and P. Octavio-Aguilar. 2021. Morphological and genetic diversity of traditional varieties of agave in Hidalgo State, Mexico. *PLoS One* 16: e0254376.

Fischer, M., and M. Van Kleunen. 2001. On the evolution of clonal plant life histories. *Evolutionary Ecology* 15: 565–582.

Foster, T. M., and M. J. Aranzana. 2018. Attention sports fans! The far-reaching contributions of bud sport mutants to horticulture and plant biology. *Horticulture Research* 5: 44.

Fucikovsky, L. 2001. "Tristeza" and death of *Agave tequilana* Weber var. *Blue* [sic]. *In* S. H. De Boer [ed.], Plant pathogenic bacteria. Springer, Dordrecht, Netherlands. Website: https://doi.org/10.1007/978-94-010-0003-1_81

Gaut, B. S., C. M. Díez, and P. L. Morrell. 2015. Genomics and the contrasting dynamics of annual and perennial domestication. *Trends in Genetics* 31: 709–719.

Gaut, B. S., D. K. Seymour, Q. Liu, and Y. Zhou. 2018. Demography and its effects on genomic variation in crop domestication. *Nature Plants* 4: 512–520.

Gentry, H. S. 1972. The agave family in Sonora (Agriculture handbook no. 399). U.S. Agricultural Research Service, Washington, D.C., USA.

Gentry, H. S. 2004. Agaves of continental North America, 3rd ed. University of Arizona Press, Tucson, AZ, USA.

Gil-Vega, K. G., M. G. Chavira, O. M. De La Vega, J. Simpson, and G. J. Vandemark. 2001. Analysis of genetic diversity in Agave tequilana var. Azul using RAPD markers. *Euphytica* 119: 335–341.

Gil-Vega, K., C. Díaz, A. Nava-Cedillo, and J. Simpson. 2006. AFLP analysis of *Agave tequilana* varieties. *Plant Science* 170: 904–909.

Goettsch, B., T. Urquiza-Haas, P. Koleff, F. Acevedo Gasman, A. Aguilar-Meléndez, V. Alavez, G. Alejandre-Iturbide, et al. 2021. Extinction risk of Mesoamerican crop wild relatives. *Plants, People, Planet* 3: 775–795.

Gross, S. M., J. A. Martin, J. Simpson, M. J. Abraham-Juarez, Z. Wang, and A. Visel. 2013. De novo transcriptome assembly of drought tolerant CAM plants, *Agave deserti* and *Agave tequilana*. *BMC Genomics* 14: 563.

Gutaker, R. M., C. L. Weiß, D. Ellis, N. L. Anglin, S. Knapp, J. Luis Fernández-Alonso, S. Prat, and H. A. Burbano. 2019. The origins and adaptation of European potatoes reconstructed from historical genomes. *Nature Ecology and Evolution* 3: 1093–1101.

Hu, W., C. Ji, Z. Liang, J. Ye, W. Ou, Z. Ding, G. Zhou, et al. 2021. Resequencing of 388 cassava accessions identifies valuable loci and selection for variation in heterozygosity. *Genome Biology* 22(1), 1–23. https://doi.org/10.1186/s13059-021-02524-7

Hübner, S., and M. B. Kantar. 2021. Tapping diversity from the wild: from sampling to implementation. *Frontiers in Plant Science* 12: 626565.

Hufford, M. B., X. Xu, J. Van Heerwaarden, T. Pyhäjärvi, J. M. Chia, R. A. Cartwright, R. J. Elshire, et al. 2012. Comparative population genomics of maize domestication and improvement. *Nature Genetics* 44: 808–811.

Hyten, D. L., I. Choi, Q. Song, R. C. Shoemaker, R. L. Nelson, J. M. C. Da Costa, J. E. Specht, and P. B. Cregan. 2007. Highly variable patterns of linkage disequilibrium in multiple soybean populations. *Genetics* 175: 1937–1944.

Hyten, D. L., Q. Song, Y. Zhu, I. Y. Choi, R. L. Nelson, J. M. Costa, J. E. Specht, et al. 2006. Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences, USA* 103: 16666–16671.

Kamvar, Z. N., J. F. Tabima, and N. J. Grünwald. 2014. poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2: e281.

Kantar, M. B., A. R. Nashoba, J. E. Anderson, B. K. Blackman, and L. H. Rieseberg. 2017. The genetics and genomics of plant domestication. *BioScience* 67: 971–982.

Khoury, C. K., S. Brush, D. E. Costich, H. A. Curry, S. Haan, J. M. M. Engels, L. Guarino, et al. 2021. Crop genetic erosion: understanding and responding to loss of crop diversity. *New Phytologist* 233: 84–118.

Klekowski, E. J. 2003. Plant clonality, mutation, diplontic selection and mutational meltdown. *Biological Journal of the Linnean Society* 79: 61–67.

Klimova, A., K. Y. Ruiz Mondragón, F. Molina Freaner, E. Aguirre-Planter, and L. E. Eguiarte. 2022. Genomic analyses of wild and cultivated bacanora agave (*Agave angustifolia* var. *pacifica*) reveal inbreeding, few signs of cultivation history and shallow population structure. *Plants* 11: 1426.

Korunes, K. L., and K. Samuk. 2021. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources* 21: 1359–1368.

Laucou, V., A. Launay, R. Bacilieri, T. Lacombe, A. F. Adam-Blondon, A. Bérard, A. Chauveau, et al. 2018. Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. *PLoS One* 13: e0192540.

Li, Y., C. Colleoni, J. Zhang, Q. Liang, Y. Hu, H. Ruess, R. Simon, et al. 2018. Genomic analyses yield markers for identifying agronomically important genes in potato. *Molecular Plant* 11: 473–484.

Lian, Q., D. Tang, Z. Bai, J. Qi, F. Lu, S. Huang, and C. Zhang. 2019. Acquisition of deleterious mutations during potato polyploidization. *Journal of Integrative Plant Biology* 61: 7–11.

Linck, E., and C. J. Battey. 2019. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources* 19: 639–647.

Lira, M. G., I. J. Davidson-Hunt, and J. P. Robson. 2022. Artisanal products and land-use land-cover change in indigenous communities: The case of mezcal production in Oaxaca, *Mexico. Land* 11: 387.

MacNeish, R. S. 1967. A summary of the subsistence. *In* D. S. Byers [ed.], The prehistory of the Tehuacán Valley, 1. Environment and subsistence. University of Texas Press, Austin, TX, USA.

Malinsky, M., E. Trucchi, D. J. Lawson, and D. Falush. 2018. RADpainter and fineRADstructure: Population inference from RADseq data. *Molecular Biology and Evolution* 35: 1284–1290.

Manichaikul, A., J. C. Mychaleckyj, S. S. Rich, K. Daly, M. M. Sale, and W. Chen. 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics* 26: 2867–2873.

Manrique-Carpintero, N. C., J. J. Coombs, G. M. Pham, F. P. E. Laimbeer, G. T. Braz, J. Jiang, R. E. Veilleux et al. 2018. Genome reduction in tetraploid potato reveals genetic load, haplotype variation, and loci associated with agronomic traits. *Frontiers in Plant Science* 9: 944.

Marrano, A., D. Micheletti, S. Lorenzi, D. B. Neale, and M. S. Grando. 2018. Genomic signatures of different adaptations to environmental stimuli between wild and cultivated *Vitis vinifera* L. *Horticulture Research* 5: 34.

Martínez-González, C., G. Castellanos-Morales, J. Barrera-Redondo, G. S. La Vega, H. S. Hernández-Rosales, J. Gasca-Pineda, E. Aguirre-Planter, et al. 2021. Recent and historical gene flow in cultivars, landraces, and a wild taxon of *Cucurbita pepo* in Mexico. *Frontiers in Ecology and Evolution* 9: 616051.

McKey, D., M. Elias, B. Pujol, and A. Duputié. 2010. The evolutionary ecology of clonally propagated domesticated plants. *New Phytologist* 186: 318–332.

Molina-Freaner, F., and L. E. Eguiarte. 2003. The pollination biology of two paniculate agaves (Agavaceae) from northwestern Mexico: contrasting roles of bats as pollinators. *American Journal of Botany* 90: 1016–1024.

Moreno Salazar, S. F., M. Esqueda, J. Martínez, and G. Palomino. 2007. Tamaño del genoma y cariotipo en Agave angustifolia y A. rhodacantha de Sonora, México. *Revista Fitotecnia Mexicana* 30: 13–23.

Myles, S., A. R. Boyko, C. L. Owens, P. J. Brown, F. Grassi, M. K. Aradhya, B. Prins, et al. 2011. Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences, USA* 108: 3530–3535.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGill, P. R. Minchin, et al. 2017. Vegan: Community ecology package. *R package version* 2.4-2.

Olofsson, H., and P. Lundberg. 2007. The twofold cost of sex unfolded. *Evolutionary Ecology Research* 9: 1119–1129.

Patterson, N., A. L. Price, and D. Reich. 2006. Population structure and eigenanalysis. *PLoS Genetics* 2: e190.

Pembleton, L. W., N. O. Cogan, and J. W. Forster. 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources* 13: 946–952.

Pickersgill, B. 2016. Domestication of plants in Mesoamerica: an archaeological review with some ethnobotanical interpretations. *In* R. Lira, A. Casas, and J. Blancas [eds.], Ethnobotany of Mexico. Springer, NY, NY, USA.

Potato Genome Sequencing Consortium. 2011. Genome sequence and analysis of the tuber crop potato. *Nature* 475: 189–195.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J Maller, et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81: 559–575.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website: https://www.R-project.org/

Ramírez-Ramírez, M.J., N. A. Mancilla-Margalli, L. Meza-Álvarez, R. Turincio-Tadeo, D. Guzmán-de Peña, and M. E. Avila-Miranda ME. 2017. Epidemiology of *Fusarium* agave wilt in Agave tequilana Weber var. Azul. *Plant Protection Science* 53: 144–152.

Ramu, P., W. Esuma, R. Kawuki, I. Y. Rabbi, C. Egesi, J. V. Bredeson, R. S. Bart, et al. 2017. Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. Nature Genetics 49: 959–963.

Racimo, F., J. Schraiber, F. Casey, and E. Huerta-Sanchez. 2016. Directional selection and adaptation. In R. M. Kliman [ed.], Encyclopedia of evolutionary biology, 444–452. Academic Press, Waltham, MA, USA.

Rivera-Lugo, M., A. García-Mendoza, J. Simpson, E. Solano, and K. Gil-Vega. 2018. Taxonomic implications of the morphological and genetic variation of cultivated and domesticated populations of the *Agave angustifolia* complex (Agavoideae, Asparagaceae) in Oaxaca, Mexico. *Plant Systematics and Evolution* 304: 969–979.

Robertson, A. 1964. The effect of non-random mating within inbred lines on the rate of inbreeding. *Genetics Research* 5: 164–167.

Rodríguez-Garay, B., J. Lomelí-Sención, E. Tapia-Campos, A. Gutiérrez-Mora, J. García-Galindo, J. Rodríguez-Domínguez, D. Urbina-López, and I. Vicente-Ramírez. 2009. Morphological and molecular diversity of *Agave tequilana* Weber var. Azul and *Agave angustifolia* Haw. var. Lineño. *Industrial Crops and Products* 29: 220–228.

Ross-Ibarra, J., P. L. Morrell, and B. S. Gaut. 2007. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences, USA* 104: 8641–8648.

Rubio-Ríos J. R., R. A. Guzmán-Plazola, V. Ayala-Escobar, and R. Rubio-Cortés. 2019. Spatial and temporal dynamics of blue agave (Agave tequilana Weber var. Azul) wilt in Jalisco Mexico. *Journal of Phytopathology* 167: 299–311.

Ruiz-Mondragon, K. Y., E. Aguirre-Planter, J. Gasca-Pineda, A. Klimova, R. Trejo-Salazar, M. A. R. Guerra, R. A. Medellin, et al. 2022. Conservation genomics of *Agave tequilana* Weber var. Azul: low genetic differentiation and heterozygote excess in the tequila agave from Jalisco, Mexico. *PeerJ* 10: e14398.

Schoen, D. J., and S. T. Schultz. 2019. Somatic mutation and evolution in -plants. *Annual Review of Ecology, Evolution, and Systematics* 50: 49–73.

Shi, J., and J. Lai. 2015. Patterns of genomic changes with crop domestication and breeding. *Current Opinion in Plant Biology* 24: 47–53.

Shiferaw, B., B. M. Prasanna, J. Hellin, and M. Bänziger. 2011. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Security* 3: 307–327.

Silvertown, J. 2008. The evolutionary maintenance of sexual reproduction: evidence from the ecological distribution of asexual reproduction in clonal plants. *International Journal of Plant Sciences* 169: 157–168.

Slate, J., P. David, K. G. Dodds, B. A. Veenvliet, B. C. Glass, T. E. Broad, and J. C. McEwan. 2004. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity* 93: 255–265.

Smith, B. D. 2001. Documenting plant domestication: the consilience of biological and archaeological approaches. *Proceedings of the National Academy of Sciences, USA* 98: 1324–1326.

Stoffel, M. A., M. Esser, M. Kardos, E. Humble, H. Nichols, P. David, and J. I. Hoffman. 2016. inbreedR: anRpackage for the analysis of inbreeding based on genetic markers. *Methods in Ecology and Evolution* 7: 1331–1339.

Swarup, S., E. J. Cargill, K. Crosby, L. Flagel, J. Kniskern, and K. C. Glenn. 2021. Genetic diversity is indispensable for plant breeding to improve crops. *Crop Science* 61: 839–852.

Szarek, S. R., B. Driscoll, C. Shohet, and S. Priebe. 1996. Bulbil production in *Agave* (Agavaceae) and related genera. *Southwestern Naturalist* 41: 465–469.

Tetreault, D., C. McCulligh, and C. Lucio. 2021. Distilling agro-extractivism: agave and tequila production in Mexico. *Journal of Agrarian Change* 21: 219–241.

Tian, F., N. M. Stevens, and E. S. Buckler. 2009. Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. *Proceedings of the National Academy of Sciences, USA* 106: 9979–9986.

Tibayrenc, M., and F. J. Ayala. 1991. Towards a population genetics of microorganisms: The clonal theory of parasitic protozoa. *Parasitology Today* 7: 228–232.

Trame, A. M., A. J. Coddington, and K. N. Paige. 1995. Field and genetic studies testing optimal outcrossing in *Agave schottii*, a long-lived clonal plant. *Oecologia* 104: 93–100.

Trejo, L., V. Limones, G. Peña, E. Scheinvar, O. Vargas-Ponce, D. Zizumbo -Villarreal, and P. Colunga-GarcíaMarín. 2018. Genetic variation and relationships among agaves related to the production of tequila and mezcal in Jalisco. *Industrial Crops and Products* 125: 140–149.

Valenzuela, A. 2010. A new agenda for blue agave landraces: food, energy and tequila. *GCB Bioenergy* 3: 15–24.

Vargas-Ponce, O., D. Zizumbo-Villarreal, J. Martínez-Castillo, J. Coello-Coello, and P. Colunga-GarcíaMarín. 2009. Diversity and structure of landraces of *Agave* grown for spirits under traditional agriculture: a comparison with wild populations of *A. angustifolia* (Agavaceae) and commercial plantations of *A. tequilana*. *American Journal of Botany* 96: 448–457.

Vavilov, N. I. 1931. Mexico and Central America as the principal centre of origin of cultivated plants of the New World. *Bulletin of Applied Botany, of Genetics and Plant-Breeding* 26: 179–199.

Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher. 2011. GCTA: A Tool for genome-wide complex trait analysis. *American Journal of Human Genetics* 88: 76–82.

Yu, L., C. Boström, S. Franzenburg, T. Bayer, T. Dagan, and T. B. H. Reusch. 2020. Somatic genetic drift and multilevel selection in a clonal seagrass. *Nature Ecology and Evolution* 4: 952–962.

Zeller, U., and T. Göttert. 2019. The relations between evolution and domestication reconsidered: implications for systematics, ecology, and nature conservation. *Global Ecology and Conservation* 20: e00756.

Zhang, H., N. Mittal, L. J. Leamy, O. Barazani, and B. H. Song. 2016. Back into the wild—Apply untapped genetic diversity of wild relatives for crop improvement. *Evolutionary Applications* 10: 5–24.

Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28: 3326–3328.

Zheng, Z., H. Hu, W. Lei, J. Zhang, M. Zhu, Y. Li, X. Zhang, et al. 2022. Somatic mutations during rapid clonal domestication of *Populus alba* var. *pyramidalis*. *Evolutionary Applications* 15: 1875–1887.

Zhou, Y., M. Massonnet, J. S. Sanjak, D. Cantu, and B. S. Gaut. 2017. Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proceedings of the National Academy of Sciences, USA* 114: 11715–11720.

Zhou, Z., Y. Jiang, Z. Wang, Z. Gou, J. Lyu, W. Li, Y. Yu, et al. 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology* 33: 408–414.

Zhu, M., Y. Cheng, S. Wu, X. Huang, and J. Qiu. 2022. Deleterious mutations are characterized by higher genomic heterozygosity than other genic variants in plant genomes. *Genomics* 114: 110290.

Zizumbo-Villarreal, D., and P. Colunga-GarcíaMarín. 2007. Early coconut distillation and the origins of mezcal and tequila spirits in west-central Mexico. *Genetic Resources and Crop Evolution* 55: 493–510.

SUPPORTING INFORMATION

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

APPENDIX S1. Collection information for 94 individuals of wild and cultivated *Agave angustifolia* from the states of Oaxaca and Puebla.

APPENDIX S2. Quality information for the filtered SNPs for 94 wild and cultivated individuals of *Agave angustifolia* from the states of Oaxaca and Puebla, Mexico, and 8616 SNPs.

APPENDIX S3. Boxplots of the individual-based diversity estimates and inbreeding index for wild and cultivated samples of *Agave angustifolia* from the states of Oaxaca and Puebla, Mexico.

APPENDIX S4. Barplots of the individual-based inbreeding estimates (F_{IS}) for wild *Agave angustifolia* and cultivated espadín samples from the states of Oaxaca and Puebla, Mexico.

APPENDIX S5. Heat map of relatedness among (A) 42 wild samples genotyped with 20,342 SNPs and (B) 45 cultivated samples genotyped with 4096 SNPs.

APPENDIX S6. Distribution of pairwise identity-by-state (IBS) allele sharing values amongst wild *Agave angustifolia* and cultivated espadín samples determined by the analysis of the combined data set and 8616 SNPs.

APPENDIX S7. Pairwise F_{ST} differences among all the sampling sites of wild *Agave angustifolia* and cultivated "espadín" from the states of Oaxaca and Puebla, Mexico.

APPENDIX S8. A plot of ADMIXTURE cross-validation error and respective standard deviation based on five repetitions for each K-value, from $K = 1$ through $K = 10$, based on all samples (cultivated and wild) and 8616 SNPs.

APPENDIX S9. Relationships among wild and cultivated samples of *Agave angustifolia* individuals from the states of Oaxaca and Puebla, Mexico, as shown by principal component analysis (PCA) using 8616 genomewide SNPs.

APPENDIX S10. Plot of ADMIXTURE cross-validation error and respective standard deviations based on five repetitions for each *K* value from 1 to 10 and on 42 wild *Agave angustifolia* samples genotyped with 20,342 SNPs.

APPENDIX S11. Population genetic structure of the wild *Agave angustifolia* samples collected in Puebla and Oaxaca states, Mexico, based on 20,342 SNPs.

APPENDIX S12. Mantel test results show the relationships between geographic distance and genetic distance as estimates with pairwise F_{ST} among sampling sites of wild *Agave angustifolia* from Puebla and Oaxaca states.

APPENDIX S13. Spatial genetic structure of wild *Agave angustifolia* individuals from Oaxaca and Puebla states.

How to cite this article: Klimova, A., K. Y. Ruiz Mondragón, E. Aguirre-Planter, A. Valiente, R. Lira, and L. E. Eguiarte. 2023. Genomic analysis unveils reduced genetic variability but increased proportion of heterozygotic genotypes of the intensively managed mezcal agave, *Agave angustifolia*. *American Journal of Botany* 110(8): e16216.
https://doi.org/10.1002/ajb2.16216