

RESEARCH ARTICLE

Differences in the genomic diversity, structure, and inbreeding patterns in wild and managed populations of *Agave potatorum* Zucc. used in the production of Tobalá mezcal in Southern Mexico

Karen Y. Ruiz Mondragón^{1*}, Anastasia Klimova¹, Erika Aguirre-Planter¹, Alfonso Valiente-Banuet^{2,3}, Rafael Lira⁴, Guillermo Sanchez-de la Vega¹, Luis E. Eguiarte^{1*}

1 Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, México, **2** Departamento de Ecología de la Biodiversidad, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, México, **3** Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Ciudad Universitaria, Ciudad de México, México, **4** Laboratorio de Recursos Naturales, Unidad de Biotecnología y Prototipos (UBIPRO), Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Tlalneapantla de Baz, Estado de México, México

* karen.mondragon91@gmail.com (KYRM); fruns@unam.mx (LEE)



OPEN ACCESS

Citation: Ruiz Mondragón KY, Klimova A, Aguirre-Planter E, Valiente-Banuet A, Lira R, Sanchez-de la Vega G, et al. (2023) Differences in the genomic diversity, structure, and inbreeding patterns in wild and managed populations of *Agave potatorum* Zucc. used in the production of Tobalá mezcal in Southern Mexico. PLoS ONE 18(11): e0294534. <https://doi.org/10.1371/journal.pone.0294534>

Editor: Branislav T. Šiler, Institute for Biological Research, University of Belgrade, SERBIA

Received: March 26, 2023

Accepted: November 2, 2023

Published: November 16, 2023

Copyright: © 2023 Ruiz Mondragón et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data availability. The vcf file is in Zenodo database: DOI: [10.5281/zenodo.7738332](https://doi.org/10.5281/zenodo.7738332) Raw sequences are available under BioProject PRJNA1023893 and SRA SRR26361385 in NCBI.

Funding: Ruiz Mondragón Karen Yazmin is a doctoral student from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and has received the

Abstract

Agave potatorum Zucc. locally known as Tobalá, is an important species for mezcal production. It is a perennial species that takes 10 to 15 years to reach reproductive age. Because of high demand of Tobalá mezcal and the slow maturation of the plants, its wild populations have been under intense anthropogenic pressure. The main objective of this study was to estimate the genome-wide diversity in *A. potatorum* and determine if the type of management has had any effect on its diversity, inbreeding and structure. We analyzed 174 individuals (105 wild, 42 cultivated and 27 from nurseries) from 34 sites with a reduced representation genomic method (ddRADseq), using 14,875 SNPs. The diversity measured as expected heterozygosity was higher in the nursery and wild plants than in cultivated samples. We did not find private alleles in the cultivated and nursery plants, which indicates that the individuals under management recently derived from wild populations, which was supported by higher gene flow estimated from wild populations to the managed plants. We found low but positive levels of inbreeding ($F_{IS} = 0.082$), probably related to isolation of the populations. We detected low genetic differentiation among populations ($F_{ST} = 0.0796$), with positive and significant isolation by distance. The population genetic structure in the species seems to be related to elevation and ecology, with higher gene flow among populations in less fragmented areas. We detected an outlier locus related to the recognition of pollen, which is also relevant to self-incompatibility protein (SI). Due to seed harvest and long generation time, the loss of diversity in *A. potatorum* has been gradual and artificial selection and incipient management have not yet caused drastic differences between cultivated and

CONACyT fellowship No. CVU: 739748. This study was supported by grants from PAPIIT IG200122, UNAM to Luis E. Eguiarte and Rafael Lira, and CONACyT- PRONACES project 319061 “Auge mezcalero y deudas de extinción: investigación interdisciplinaria hacia la sustentabilidad” granted to Alfonso Valiente Banuet. The funders had a role in study design, data collection, and analyses, as well as in the preparation of the manuscript

Competing interests: The authors declared that no competing interests exist.

wild plants. Also, we described an agroecological alternative to the uncontrolled extraction of wild individuals.

Introduction

Agave L., is a species-rich genus, with over 200 described species native to America’s arid lands [1–4]. Most species occur in Mexico, with many of them being endemics and microendemics [1, 3, 5]. Although *Agave* species have had enormous cultural and economic importance, currently in Mexico, the primary use of these plants is for the production of alcoholic beverages such as tequila and mezcal [2, 6, 7].

Among the alcoholic beverages production in Mexico, the tequila and mezcal industry is the second most important economic activity. Over 50% of this production is destined for export [8], which in the year 2018 represented over 53 million USD of revenue [8]. Nevertheless, to produce mezcal, large quantities of raw materials (i.e., *Agave* mature plants, water, and firewood) are required. The ever-growing demand and high requirements for raw materials have resulted in a considerable increase in land use change, with lands destined for *Agave* plant cultivation increasing from 8,663 hectares in 2003 to 21,878 ha in 2017 [8, 9].

Currently, mezcal production is considered unsustainable [10]. This is because only a few *Agave* species are cultivated in commercial monoculture plantations (i.e., *A. tequilana* Weber for tequila or *A. angustifolia* Haw. for Espadín mezcal production) [7, 11]. The recent increase in tequila and Espadín production has resulted in important ecological problems, including the removal of native vegetation, forest clearing, substitution of traditional crops for *Agave* planting, use of agrochemicals, soil contamination, and erosion. In contrast, other *Agave* species used in mezcal production are mainly extracted from the wild, resulting in overexploitation of natural populations [2, 10, 12–15]. However, as the mezcal industry keeps growing, several species of *Agave* have begun to be cultivated (i.e., *A. potatorum*, *A. marmorata* Roehl., *A. karwinskii* Zucc., *A. cupreata* Trel. & A. Berger) with seedlings derived from wild, nurseries or through clonal propagation. Furthermore, both wild and cultivated *Agave* specimens used in the production of mezcal or tequila are harvested *prior* to blossom, when the plants concentrate all their sugars and nutrients before producing their inflorescence, as these resources are used to produce tequila or mezcal. The removal of the inflorescences has consequences, it not only prevents the agave plants from producing nectar, pollen or seeds, but also impacts the populations of pollinators, causing a decline in the natural recruitment and a reduction in the genetic variation [14–18].

The increasing demand for mezcal may be already generating an “extinction debt” [19], as some *Agave* species have had a greater demand due to the unique organoleptic properties such as *A. potatorum*, known as Tobalá [16, 20].

Agave potatorum is an endemic species, restricted to the Sierra Madre del Sur, Tehuacán Valley in the state of Puebla, and part of the state of Oaxaca mountains, in Mexico [20]. It is found in evergreen sclerophyllous vegetation named Mexical [21], and in the oak and pine forest grassy slopes at an elevation between 1,760 to 2,300 m [2, 18]. It is a semelparous, monocarpic species that grows slowly (compared to *A. tequilana* and *A. angustifolia*, for instance). It takes 10 to 15 years to flower, blooming from August to November. The fruits and seeds mature from November to March. The reproductive individuals are of relatively small size compared to other Agaves that produce mezcal [20, 22]. Sexual reproduction is the primary reproductive strategy of this species; however, while rare, some vegetative propagation has

been reported in disturbed populations [23]. Each plant can produce 2,000 to 9,500 seeds [23]. *Agave potatorum* depends mainly on bats for pollination, particularly the lesser long-nosed bat (*Leptonycteris yerbabuenae* Martínez & Villa, 1940); also other bats and animals can also be involved [22, 24]. In populations characterized by a high density of flowering *Agave* species, nectarivorous bats exhibit greater foraging and pollination efficiency, due to their increased flower visitation rates, favored by the continuous nectar reward [22].

As a result of the growing demand for mezcal, high extraction rates and slow maturation, *A. potatorum* has become one of the most threatened *Agave* species [10], with IUCN Red List recognizing that populations are decreasing [25]. For instance, it was estimated that near San Luis Atolotitlán (SLA), Puebla, Mexico, 54% - 87% of the wild mature plants are extracted annually [26]. Furthermore, as this *Agave* is smaller than, for example, *A. angustifolia*, more individuals are needed to produce the same amount of mezcal [13, 14]. Due to concerns regarding the decline of wild populations, *in situ* sustainable agro-management has been proposed and is currently underway in two villages: San Luis Atolotitlán (SLA) and San Juan Raya (SJR), both in the state of Puebla in Central Mexico. In collaboration with researchers, producers have implemented collective nurseries to supply the demand of seeds and seedlings and avoid extracting wild individuals [26]. According to Delgado-Lemus [13], the cultivation of this species entails an incipient domestication process [10, 13].

Domestication is a complex evolutionary process in which human selection leads to morphological and physiological changes in plants, resulting in the modification of the genotypes and phenotypes and the differentiation of the domesticated varieties from their wild ancestors [27]. As a result of the demographic process and artificial selection, a reduction in the crop's genetic diversity through the process known as "domestication bottleneck" has been observed in many crop species [27–29]. Therefore, morphological, physiological, and genetic changes associated with domestication have been extensively studied [30–32]. However, most studies have focused on annual species, such as maize, rice, wheat, sorghum, and pumpkins [27, 33], whereas the domestication process in long-lived-perennials plants, such as *Agave*, grapes, palms, and tree nuts has received less attention [27]. Due to a prolonged juvenile period and sexual reproduction, the domestication of most perennial plants usually entails clonal propagation. This may imply an increase in diversity through the accumulation of somatic mutations in already heterozygous genotypes [27, 34, 35].

Recently several studies have focused on domestication as a model for understanding the evolutionary processes [31, 36–38]. Under this model and given the ecological, cultural, and economic importance of *A. potatorum*, we consider it is vital to understand how the management of this species is carried out in its distribution area (in the states of Oaxaca and Puebla, Mexico). Due to the conservation concerns, we employed a reduced representation genomic method, RADseq, to study the genetic variation in plants under different types of management. Our primary focus was to determine whether there was a loss of genetic diversity resulting from bottlenecks associated to the incipient domestication process, and artificial selection. To achieve this, we analyzed *A. potatorum* under different management types: wild (with no management at all), cultivated (with management), and nursery (produced in a shade house).

We believe it is crucial to understand the genetic similarities and differences between cultivated and wild plants, as wild plants can potentially serve as germplasm reservoirs for future crop improvements. Additionally, we conducted an exploratory analysis of outlier loci, as we wanted to identify agronomically or ecologically important alleles in wild populations. In cultivated localities our goal was to detect loci associated with the production of secondary metabolites that contribute to the perceived quality of this mezcal.

Given the incipient cultivation practices, we anticipated finding higher genetic diversity in undisturbed wild populations, with still low genetic differentiation from cultivated

populations, and few unique alleles in managed localities. Furthermore, we expected to find evidence of gene flow between wild and cultivated plants. Additionally, we presumed that nursery and cultivated plants would exhibit closer genetic relationships, accompanied by higher levels of inbreeding.

Materials and methods

Plant material

We collected leaves from 192 specimens of *Agave potatorum* (S1 Table in [S1 File](#) and [Fig 1](#)) covering its entire distribution range in the states of Puebla and Oaxaca, Mexico. The distribution of the sampling sites ([Fig 1](#)) was plotted using R packages *raster*, *elevatr*, *ggspatial* and *ggplot2*. A total of 34 sampling localities were visited, including different management types: wild (22 sampling sites), two nursery—sites specially dedicated to the production of seedlings—(SJRI-N and SJR2-N) and cultivated (10 sampling sites) with approximately 20 years of cultivation. In most cases cultivated plants were collected at polycultures, alternated with *A. angustifolia* or *A. karwinskii*. We only found one monoculture of *A. potatorum* in Puebla (P3-C). Apparently, due to the incipient domestication of *A. potatorum*, the number of cultivation sites (such as plantations or nurseries) for this species is still limited.

The sampled plants had different ages, the wild and cultivated plants were categorized as adults, and the nursery populations were young individuals (less than two years old). One nursery locality was in a recent set-up greenhouse, and the second one (SJRI-N) came from a shade house, where seeds have been growing for eight years; the mother plants come from a plantation under agroecological management in Cerro los Pesos, near Zapotitlán, both in San Juan Raya, Puebla (S1 Table in [S1 File](#) and [Fig 1](#)). All samples were collected and stored until DNA extraction at -80°C.

DNA extraction and sequencing

Genomic DNA was extracted from leaf tissue using a modified "Mini-Prep" CTAB protocol [39]. For the exact modifications of the protocol, see [11]. DNA quantity and quality were first examined using a 1% agarose electrophoresis gel. Then, samples of adequate quality were quantified using a Qubit 3.0 fluorometer and Qubit dsDNA broad-range kit. Library preparation for RADseq and sequencing were performed at the Biotechnology Center of the University of Wisconsin-Madison (<https://biotech.wisc.edu/>). Each sample was digested using two methylation-sensitive restriction enzymes (*PstI* and *MspI*). Samples were pooled in equimolar concentration after specific barcodes were ligated to them and sequenced using the Illumina NovaSeq 2x150 platform (Illumina, Inc., San Diego, CA, United States).

SNP calling and filtering

Demultiplexing of the samples, initial data filtering and SNP calling were performed with Ipyrad v.0.9.77 software [40], using the ddRAD (double digestion RAD) method analysis strategy. The transcriptome of *A. tequilana* (GAHU00000000.1; [41]) was used as a reference. Filtering of the vcf file was done using VCFtools v.0.1.15 [42]. To ensure data quality, we performed several filters retaining only bi-allelic sites, with a mean minimum depth of over 12, and no InDels. We removed SNPs with a significant deviation from Hardy-Weinberg equilibrium (function—*hwe* 1e-07), with a minor allele frequency (MAF) of <0.05, and individuals with more than 80% missing data. Using PLINK v.1.9 software [43], we excluded loci with high linkage disequilibrium (LD); thus, loci presenting r^2 of over 0.2 were removed from the further analysis.

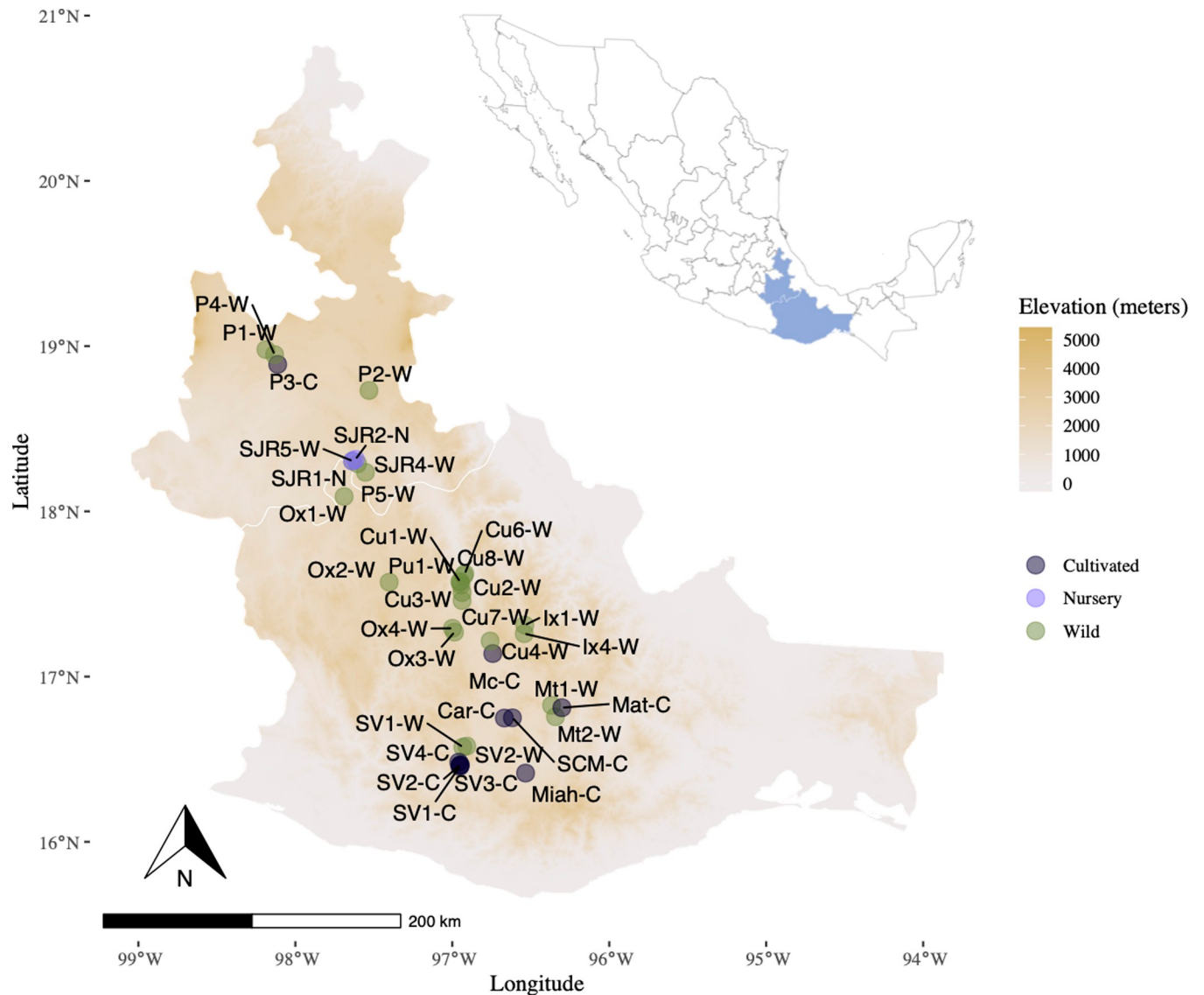


Fig 1. Localities of *Agave potatorum* in Puebla and Oaxaca, Mexico. Colors shading on the map show the elevation in the collection area, and the colored dots correspond to different types of management: wild (green), cultivated (dark purple), and nursery (purple).

<https://doi.org/10.1371/journal.pone.0294534.g001>

Genetic diversity

The number of private alleles (PA) and multilocus genotypes (mlg) were calculated with the *poppr* R package [44]. Additionally, we estimated the following genomic diversity statistics: the overall heterozygosity (H_T), observed heterozygosity (H_O), and expected heterozygosity (H_E) per locus, in three data sets: 1) per each sampling site; 2) for each management type (wild, cultivated and nursery) and 3) for each reproductive age (adults vs. juveniles). These analyses were performed with *adegenet* [45, 46] and *hierfstat* [47] packages with R [48]. We also estimated multilocus heterozygosity (MLH) with the *inbreedR* [49] package. To compare the data sets, we calculated the significant differences using a Wilcoxon test with the *ggpubr* package, and all the results were plotted using the *ggplot2* package in R.

Inbreeding and kinship

To understand how individuals within and among populations were related, we estimated the inbreeding coefficient (F_{IS}) with the *hierfstat* package. Additionally, we calculated for each sample the identity-by-descent, specifically the *Fhat3* index, based on the correlation between uniting gametes, with PLINK v.1.9 software.

With VCFtools v.0.1.15 [42], we performed a relationship analysis using the *relatedness phi* index [50] which is based on the KING algorithm (Kinship-base Inference of Genome-wide). In this test, the *relatedness* value can be interpreted as the probability of finding identical alleles by randomly sampling one allele from each heterozygous individual. The *relatedness* index equal to 0 means that individuals are unrelated, and a range from 0.0442 to 0.0884 corresponds to a relationship in 3rd degree, full siblings or 2nd degree from 0.0885 to 0.1774, parent-offspring or 1st degree from 0.1775–0.354 and a *relatedness* >0.350 corresponds to a monozygotic twin [50]. For easier visualization, all negative values were converted to 0.

Genetic structure and recent gene flow

To assess significant differences between the cultivated and wild samples, we conducted a series of complementary analyses. The first test was a principal component analysis (PCA) calculated with *adeigenet* [45, 46] package in R. Then we used F_{ST} pairwise index, to estimated confidence intervals and *p-values* according to the method proposed by Wright [51] and updated by Weir and Cockerham [52]; this analysis was done using R package *StAMPP* [53] with 1000 bootstraps and with two data set (by sampling site and by management). Additionally, we reconstructed the relationships among samples with a UPGMA tree using *poppr* R package. Also, we performed an isolation by distance test to assess the spatial component of the population structure using the *vegan* [54] R package. For the Mantel test, we used the genetic distance (F_{ST}) among sampling sites and geographic distance (km) estimated using Geographic Distance Matrix Generator [55] (https://biodiversityinformatics.amnh.org/open_source/gdmg/).

To understand the admixture between cultivated, nursery and wild plants, we performed a sampled assignment analysis using ADMIXTURE v.1.23 [56, 57], with different numbers of clusters (K 1 to 10), with three replicates for each K value and 2000 bootstraps. The best K value was determined using the likelihood and cross-validation error. The ancestry values for each sample were plotted using the *ggplot2* R package.

To trace the genetic and geographic origin of cultivated and nursery plants, we used BayesAss version 3.0.4 (BA3-SNPs). This program allows the estimation of recent migration using next-generation sequence data [58]. We performed two independent runs, the first one according to the type of management (wild, cultivated, and nursery) and the second by locality (sampling site). To optimize the analyses, we adjusted the migration parameters at 0.6 and inbreeding at 0.2, as suggested by the software's manual. The iteration number was set up at 50000000, with 1000000 additional iterations considered as *burn-in*.

Outlier loci detection and annotation

We used two complementary approaches to detect outlier loci. The data set was divided into three groups corresponding to the management type (wild, cultivated, and nursery). First, we used BayeScan v2.0 [59, 60], which identifies candidate *loci* using differences in allele frequencies among populations. The parameters were set as follow: 10 interval thinning size, 5000 length of pilot runs, and 500 000 burn-in length. Then, using R software, we applied the Empirical Cumulative Distribution Function (*ecdf*), calculated the *p-value*, and adjusted it

with $BH = 0.05$, or its alias false discovery rate (fdr), this adjustment controls the expected proportion of false discoveries [61].

The second analysis was performed using the R package *pcadapt*, which is based on a multi-dimensional approach that measures how distant each point is from the mean [62, 63]. This analysis performs principal component analysis and calculates p -values to test outliers based on the correlation between genetic variation and the first principal components [62, 63]. We estimated the first K principal components, testing different numbers of K , with the best fit in $K = 3$. Then we calculated the p -values and adjusted them with a *Bonferroni* correction setting α at 0.001.

Subsequently, we used the annotated transcriptome of *A. tequilana* (GAHU00000000.1; [41]) by performing an alignment with Blast [64] using a dataset generated for the purpose of this study, which was made up of 16 annotations of species of the class Liliopsida: *Sorghum bicolor* (GCF_000003195.3), *Brachypodium distachyon* (GCF_000005505.3), *Setaria italica* (GCF_000263155.2), *Musa acuminata subsp. malaccensis* (GCF_000313855.2), *Elaeis guineensis* (GCF_000442705.1), *Oryza sativa* (GCF_001433935.1), *Ananas comosus* (GCF_001540865.1), *Asparagus officinalis* (GCF_001876935.1), *Triticum dicoccoides* (GCF_002162155.1), *Panicum hallii* (GCF_002211085.1), *Aegilops tauschii subsp. strangulata* (GCF_002575655.1), *Setaria viridis* (GCF_005286985.1), *Phoenix dactylifera* (GCF_009389715.1), *Dioscorea cayenensis subsp. rotundata* (GCF_009730915.1), *Zea mays* B73 (GCF_902167145.1) and the model species *Arabidopsis thaliana* (GCF_000001735.4); using an e -value of 0.00001.

The loci that were detected as outliers by BayeScan and *pcadapt* were mapped to the already annotated transcriptome so we could identify their function (biological, cellular, or molecular) and the associated protein through Gene Ontology (GO) and InterProScan [65].

Results

Sequencing and genotyping

We obtained 72 GB of raw data, with an average sequencing *phred* quality of 36.54% GC content, and a total of 5,759,792 raw reads. After initial quality filtering and SNP calling with Ipyrad software, 722,166 putative SNPs were derived. Subsequently, when we performed the quality filters using VCFtools, we obtained a total of 14,875 high-quality SNPs and 29,750 alleles, with 9.1% missing data.

Eighteen plants that presented low number of reads were excluded. The final data set consisted of 174 individuals, of which 104 were collected in Oaxaca state, Mexico, and 70 in Puebla state, Mexico. By management type, the final data set was 105 individuals from the wild category, 42 cultivated, and 27 from nurseries. Moreover, 147 plants were cataloged as adults and 27 as juveniles.

Genetic diversity

The number of multilocus genotypes was $mlg = 174$ (Table 1), indicating that all organisms had a unique (different) genotype. The total heterozygosity for *A. potatorum* was $H_T = 0.252$. At the species level, the expected heterozygosity was greater than the observed heterozygosity ($H_E = 0.232$, $H_O = 0.213$). The expected heterozygosity at the sampling site level ranged between $H_E = 0.194$ – 0.248 (S1 Fig in S1 File). The highest heterozygosity was found in two wild populations, one from Oaxaca (Ox4-W: $H_E = 0.248$) and another from Puebla (P1-W: $H_E = 0.247$). The cultivated population from Sola de Vega, Oaxaca (SV2-C) presented the lowest expected heterozygosity ($H_E = 0.194$). On the other hand, the observed heterozygosity ranged between $H_O = 0.175$ – 0.232 . Populations with the highest observed heterozygosity were the

Table 1. Genomic diversity index in *Agave potatorum* calculated with 14,875 SNPs. *mlg* = multilocus genotype; *PA* = private alleles; H_O = observed heterozygosity; H_E = expected heterozygosity; *MLH* = multilocus heterozygosity, F_{IS} and *Fhat3* = inbreeding coefficient, SD after $\pm$.

Diversity index	Management			Age		Full data set
	Wild	Cultivated	Nursery	Adults	Youth	
<i>mlg</i>	105	42	27	NA	NA	174
Private alleles	11	0	0	NA	NA	11
H_O	0.213 $\pm$ 0.012	0.209 $\pm$ 0.005	0.227 $\pm$ 0.006	0.212 $\pm$ 0.011	0.227 $\pm$ 0.006	0.213
H_E	0.234 $\pm$ 0.010	0.227 $\pm$ 0.013	0.244 $\pm$ 0.003	0.231 $\pm$ 0.011	0.244 $\pm$ 0.003	0.252
<i>MLH</i>	0.214 $\pm$ 0.026	0.209 $\pm$ 0.006	0.229 $\pm$ 0.016	0.212 $\pm$ 0.023	0.229 $\pm$ 0.016	0.215
F_{IS}	0.029 $\pm$ 0.043	0.019 $\pm$ 0.062	0.053 $\pm$ 0.021	0.026 $\pm$ 0.049	0.053 $\pm$ 0.021	0.082
<i>Fhat3</i>	0.157 $\pm$ 0.009	0.150 $\pm$ 0.021	0.121 $\pm$ 0.058	0.155 $\pm$ 0.085	0.121 $\pm$ 0.058	0.15
<i>N</i>	105	42	27	147	27	174

<https://doi.org/10.1371/journal.pone.0294534.t001>

nursery localities from San Juan Raya, Puebla (SJR1-N: H_O = 0.232), the two wild localities of the same site (SJR5-W: H_O = 0.231; SJR4-W: H_O = 0.230) and Tepalcatepec (Oax1-W: H_O = 0.227), while a wild population from Puebla (P4 -W, Valsequillo) presented the lowest observed heterozygosity (H_O = 0.175) (S1 Fig in [S1 File](#)).

The expected heterozygosity was similar in the wild and nursery groups without significant differences (Wild: H_E = 0.234, SD = 0.010; Nursery H_E = 0.244 SD = 0.001). Cultivated samples presented significantly lower genetic diversity (H_E = 0.227, SD = 0.013) compared to nursery (*Wilcoxon test*; P = 0.0303), but without significant differences between cultivated and wild groups (*Wilcoxon test*; P = 0.163) ([Fig 2A](#) and S2 Table in [S1 File](#)).

The analysis by management category showed that the number of private alleles (*PA*) in wild localities was *PA* = 11, while the nursery and the cultivated groups had no private alleles ([Table 1](#)).

In the individual-based multilocus heterozygosity (*MLH*) analysis (S1 Fig in [S1 File](#)), the most diverse sampling sites were populations from the nursery locality of San Juan Raya, Puebla (SJR1-N: *MLH* = 0.244, SD = 0.008), followed by the wild localities SJR5-W (*MLH* = 0.237, SD = 0.002) and SJR4-W (*MLH* = 0.232, SD = 0.007). A wild population of Puebla (P4-W) had the lowest value of *MLH* (*MLH* = 0.115, SD = 0.058).

When we compared management types ([Fig 2B](#) and [Table 1](#)), the highest multilocus diversity was obtained for the nursery group (*MLH* = 0.229 SD = 0.016), followed by wild samples (*MLH* = 0.214; SD = 0.014) while the plants from cultivated sites presented the lowest multilocus heterozygosity (*MLH* = 0.209; SD = 0.006). We found significant differences among all comparisons based on management type; wild vs. cultivated (*Wilcoxon test*; P = 2.76E-07), wild vs. nursery (*Wilcoxon test*; P = 3.83E-07) and cultivated vs. nursery (*Wilcoxon test*; P = 1.14E-13) (S2 Table in [S1 File](#)).

Inbreeding and kinship

The inbreeding index (F_{IS}) showed considerable variation among samples (S1 Fig in [S1 File](#)), ranging from -0.139 to 0.110 (mean 0.028; SD = 0.048). One wild population from Oaxaca “Cuicatlán” (Cu4-W), and another from Puebla (P4-W) had the highest positive inbreeding coefficient (F_{IS} = 0.110 and F_{IS} = 0.100, respectively). In contrast, one cultivated population from Oaxaca: Sola de Vega (SV2-C), and wild localities from Cuicatlán (Cu7-W, Cu3-W), Ixtlán (Ix-W) and the cultivated samples from Matatlán (Mat-C), had an excess of heterozygotes.

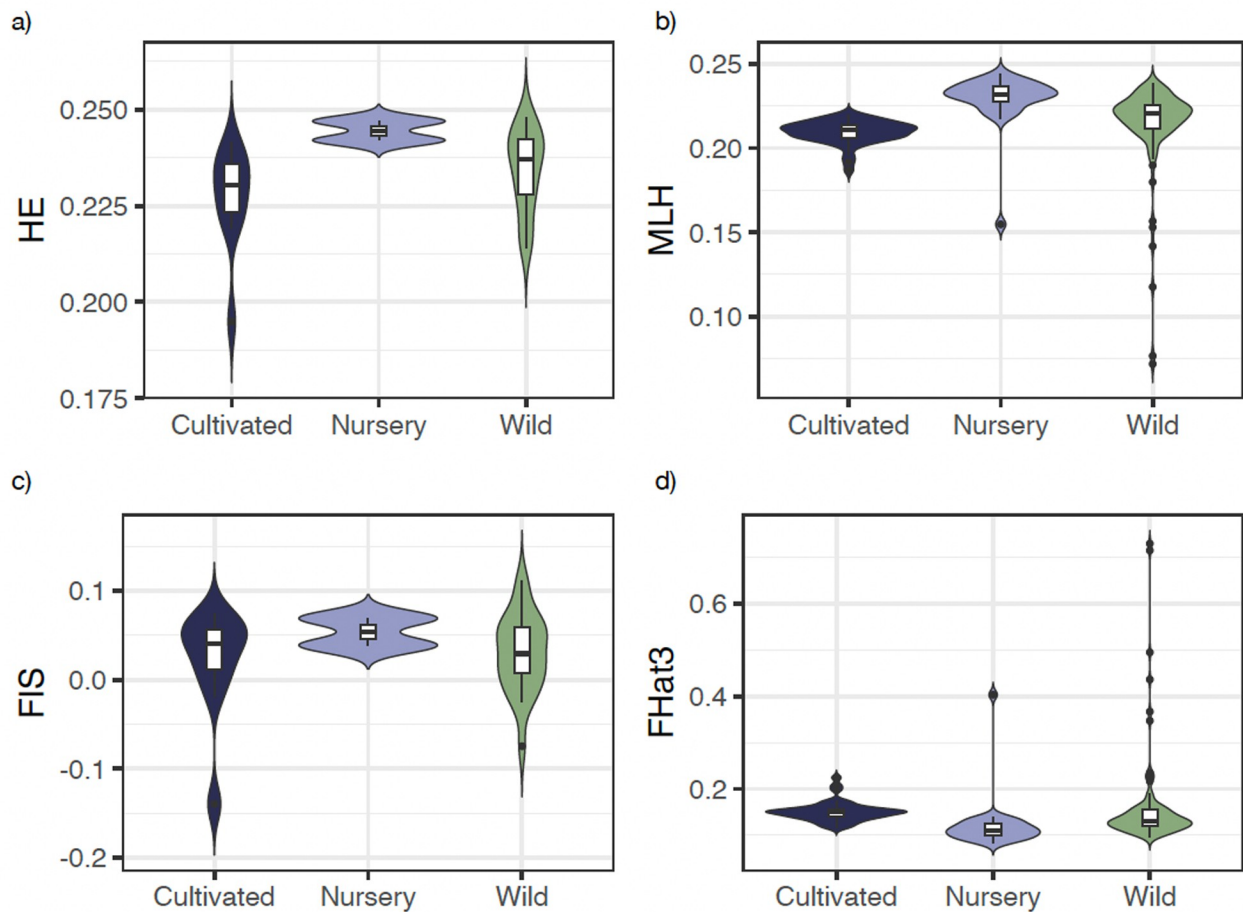


Fig 2. Genomic diversity index. Violin graph that shows the distribution of the genomic diversity; within each graph, there is a box plot showing the mean (intermediate line) and the variance of the data. a) expected heterozygosity (H_E); b) Multilocus heterozygosity (MLH); c) Inbreeding index (F_{IS}); d) F_{hat3} , per type of management: cultivated, nursery, and wild, estimated with 14 875 SNPs.

<https://doi.org/10.1371/journal.pone.0294534.g002>

When performing the analysis by management category (Fig 2A and Table 1), the highest F_{IS} values were obtained for the nursery samples ($F_{IS} = 0.053$, $SD = 0.021$), followed by the wild samples ($F_{IS} = 0.029$, $SD = 0.043$) and finally, the cultivated localities ($F_{IS} = 0.019$, $SD = 0.062$). We found significant differences (S2 Table in S1 File) between nursery and cultivated samples (Wilcoxon test; $P = 0.0303$) and nursery and wild samples (Wilcoxon test; $P = 0.029$).

All categories of *A. potatorum* presented considerable levels of inbreeding. We obtained an average inbreeding index of $F_{hat3} = 0.150$ ($SD = 0.082$). The highest coefficient of inbreeding was obtained for wild populations ($F_{hat3} = 0.157$, $SD = 0.009$), followed by cultivated populations ($F_{hat3} = 0.150$, $SD = 0.021$) and finally, nursery populations ($F_{hat3} = 0.121$, $SD = 0.058$), with significant differences between and among all types of management (Fig 2D and S2 Table in S1 File).

The kinship analysis (relatedness ϕ_i) showed that, in general, the sampled individuals were unrelated (S2 Fig in S1 File). When carrying out the analysis by management type, in wild localities we estimated an average r of 0.005, in the cultivated plants $r = 0.013$, and in the

nursery plants $r = 0.028$. In the case of the nursery group, all the seeds were collected from different mother plants, so it was expected that they had low relatedness. These estimates correspond to the categories of third-degree relationships and unrelated.

At the sampling site level (S2 Fig in [S1 File](#)), a third-degree relationship was found in different localities: 1) between individuals from the nursery of San Juan Raya, Puebla (SJRN-1 and SJRN-2); 2) in the wild populations from Puebla (SJR4 and SJR5); 3) between the wild locality from Cuicatlán (Cu7); 4) in the cultivated samples from the locality from Sola de Vega (SV4-C) and 5) a mezcal factory (Palenque “Mc-C”) from Oaxaca Valley.

Genetic structure and recent gene flow

A principal component analysis (PCA) revealed three genetic clusters in our data set. Nevertheless, the variance explained by the first two axes was not high, with EV1 at 5.2% and EV2 at 1.8%. The first group, in the upper right quadrant of the plot ([Fig 3](#)), contained individuals from the nursery site from San Juan Raya, Puebla (SJRN-1-N). The second cluster, in the lower right, was formed by the nursery localities (SJR1-N and SJR2-N), some wild samples from the same geographic region (SJR4-W and SJR5-W), and several wild localities from Oaxaca and Puebla (Oax1-W, Pu1-W, P2-W, P4-W). The last cluster was composed of all the cultivated samples from Oaxaca and Puebla, and samples from the wild localities of Cuicatlán and Matatlán, Oaxaca.

The overall genetic differentiation among sample localities was relatively low, with $F_{ST} = 0.0796$. The pairwise values between localities ranged from $F_{ST} = 0.005$ to 0.186 (S3A Fig in [S1 File](#)). The sampling sites of San Juan Raya in Puebla (SJR) had very low differentiation ($F_{ST} = 0.005$ to 0.159) among them. The site exhibiting the most divergence was SV2-C, a cultivated sample from Sola de Vega, which had the greatest difference with respect to the other samples, particularly with the cultivated site of Puebla which was the only population found in a monoculture (P3-C) ($F_{ST} = 0.159$).

When we analyzed the genetic differentiation considering the different types of management, we found the highest genetic difference between cultivated and nursery, $F_{ST} = 0.080$ (p -value = 0). Wild vs. nursery $F_{ST} = 0.031$ (p -value = 0) was intermediate. The lowest differentiation was found between cultivated and wild samples $F_{ST} = 0.023$ (p -value = 0).

In the UPGMA analysis (S3B Fig in [S1 File](#)), we found three main groups. The most divergent groups contained the wild individuals from Puebla (P4-5W) and Cuicatlán (Cu6-7, CU4-10, Cu1-3 Cu4-2), Oaxaca. The second group was the largest, comprising the cultivated localities of Puebla (P3-C) and cultivated (Palenque) localities of Oaxaca (Mc-C, PCon, PH, Car, SCM), the wild and cultivated populations of Sola de Vega (SV-C, SV-W), and the wild localities of Matatlán (Mt-W), Cuicatlán (Cu-W), Ixtlán (Ix-W) and Oaxaca (Oax-W). The last group was almost entirely formed by wild populations from Oaxaca, wild and nursery localities of San Juan Raya (SJR) and wild populations of Puebla (P-W). An analysis of isolation by distance showed a strong correlation ($r = 0.48$, $p < 0.001$), suggesting the importance of distance in structuring *A. potatorum* populations (S4 Fig in [S1 File](#)).

Admixture analysis showed different levels of assignment ([Fig 4A and 4B](#)). $K = 3$ was the most probable K (CV = 0.49875), followed by $K = 2$ (CV = 0.49945) (S3 Table in [S1 File](#)). At $K = 2$ ([Fig 4A](#)), the wild localities were divided into two genetic groups, each corresponding to a state (i.e., Oaxaca or Puebla). The nursery localities were in the same genetic pool (pale blue in $K = 2$, [Fig 4A](#)) as the wild samples from Puebla. In contrast, the cultivated individuals shared alleles with wild samples from Oaxaca (in dark blue in $K = 2$, [Fig 4A](#)). In $K = 3$, we identified a similar pattern to $K = 2$, only with more detail; the nursery and cultivated localities present

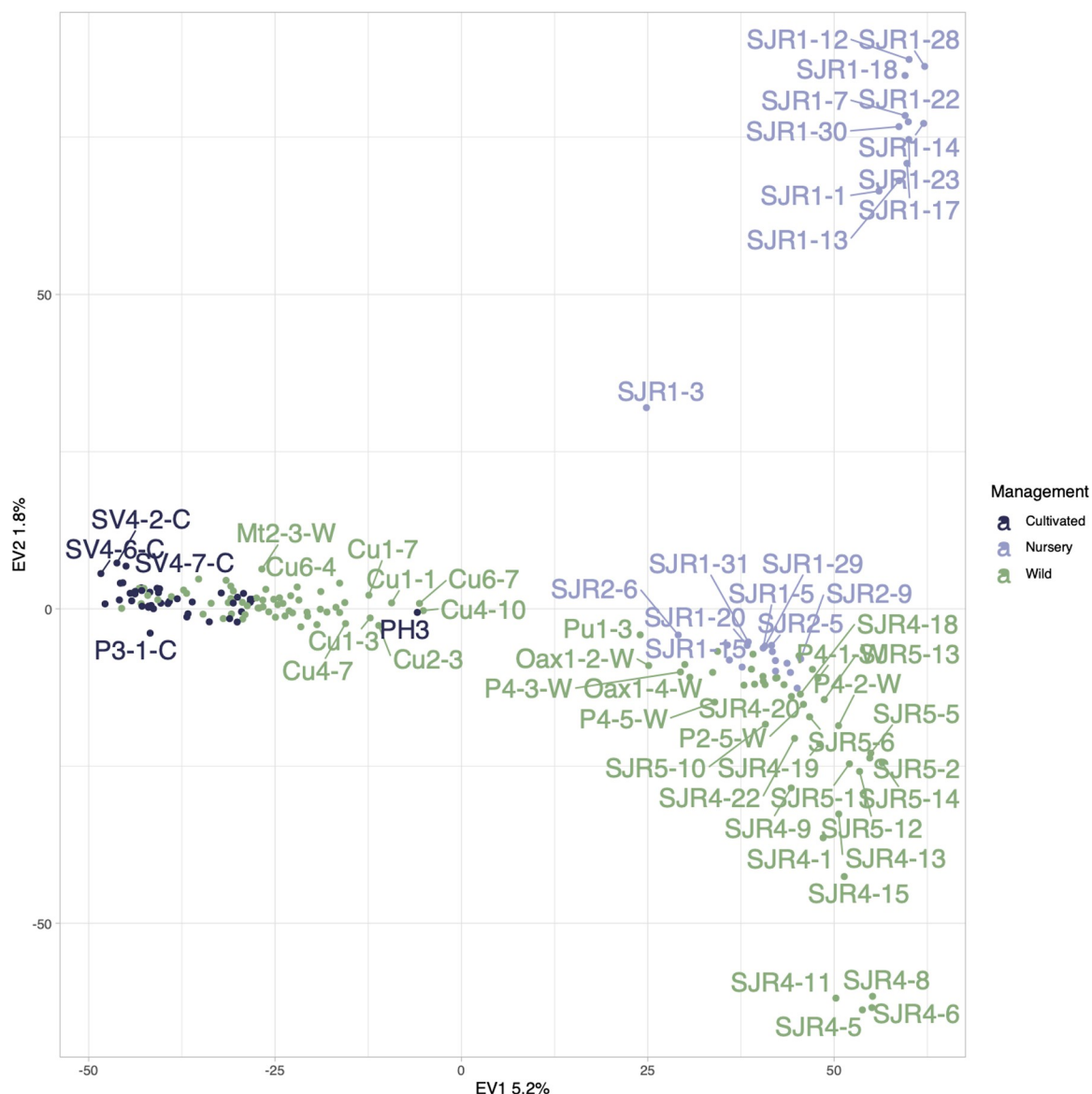


Fig 3. Principal component analysis (PCA) of *Agave potatorum* individuals, cultivated samples in dark purple, nursery in purple and wild in green estimated with 14 875 SNPs.

<https://doi.org/10.1371/journal.pone.0294534.g003>

their gene pool (pale blue for cultivated, dark blue for nursery, $K = 3$ Fig 4A), and each group with different derived alleles from the wild populations.

To understand better the ancestry, we made the plots of $K = 2$ and $K = 3$ by locality (Fig 4B). In $K = 2$, all the localities from Sola de Vega grouped in their own genetic pool (in dark blue in Fig 4B); likewise for the populations of San Juan Raya (SJR) and the wild localities of Puebla (P1-W, P2-W, P3-W, P4-W) (in pale blue in the Fig 4B). In contrast, in $K = 2$ the populations from wild localities of Oaxaca (Cuicatlán, Ixtlán, Matatlán, Ox2-W, Ox3-W, and Ox4-W) and the cultivated localities (Mc-C, Matatlán, Miahuatlán, and Santa Catarina Minas) had alleles shared between the two large gene pools.

In $K = 3$ (Fig 4B) we observed that the wild localities from Cuicatlán presented their own genetic pool (in green in this figure), forming a group among themselves and with a

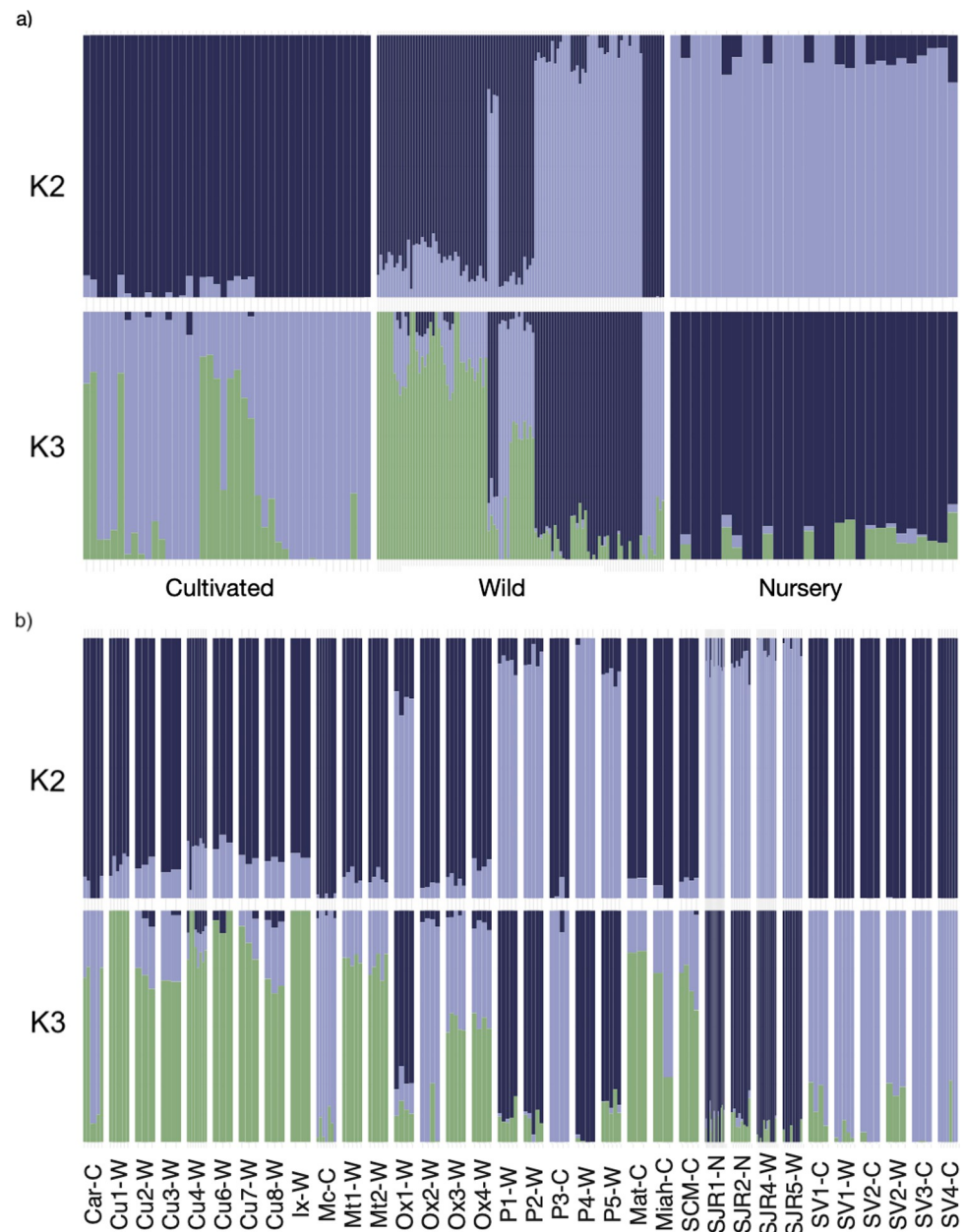


Fig 4. Admixture analysis showing the assignment probability of individuals from *Agave potatorum* ($K = 2$ and $K = 3$) a) by type of management: (cultivated, nursery, and wild), and b) by locality.

<https://doi.org/10.1371/journal.pone.0294534.g004>

geographically close locality from Ixtlán (Ix-W), with alleles shared with the wild (Ox1-W, Ox2-W, Ox3-W, Ox4-W) and cultivated (Mat-C, Miah-C, SCM-C) populations of Oaxaca, while the localities of Sola de Vega (in pale blue in this figure) and San Juan Raya (in dark blue) had their own gene pools.

Gene flow analysis revealed an interesting pattern (Fig 5). In the site-level analysis (Fig 5B), we estimated a low proportion of migrants within each site, ranging from 0.007 to 0.098 (S4 Table in S1 File). For instance, the most considerable fraction of migrants ($m = 0.099$; $SD = 0.042$) was found from populations of Sola de Vega (SVC) to different locations, for

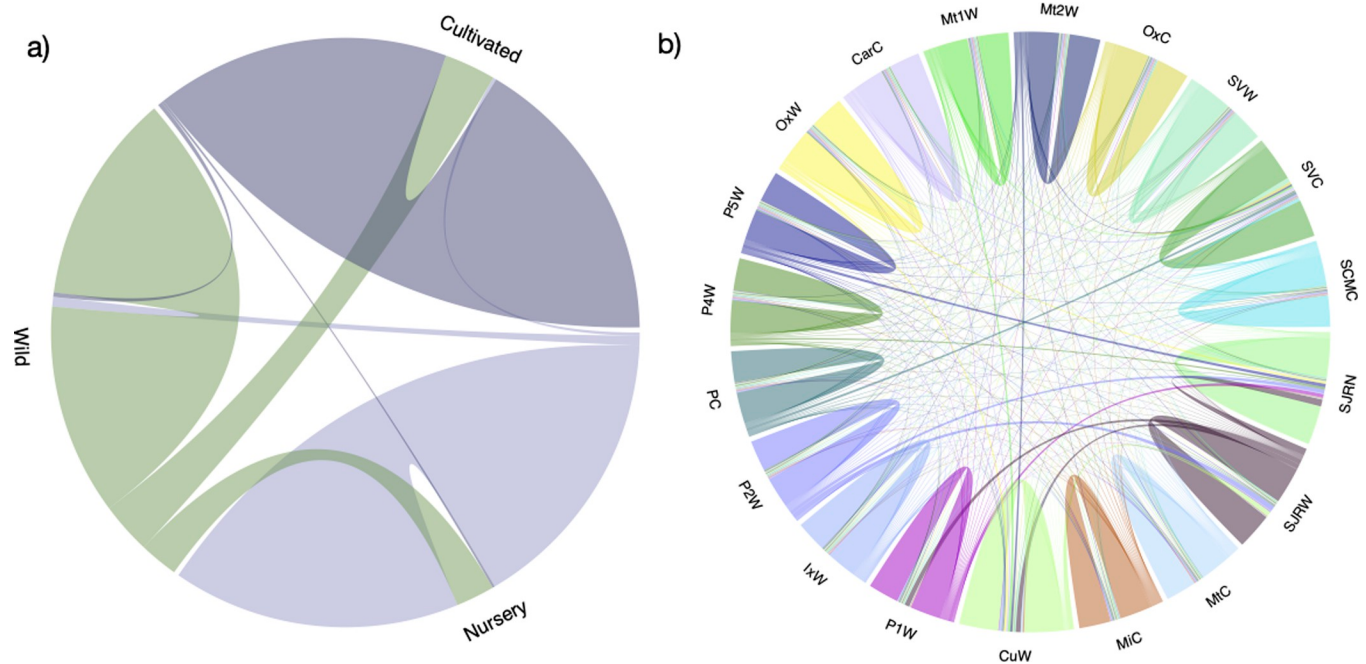


Fig 5. Migration rates estimated with BA3 software and 14,875 SNPs. a) migration rate per management type: cultivated, nursery, and wild; b) Migration rates per locality.

<https://doi.org/10.1371/journal.pone.0294534.g005>

example, the cultivated populations of San Dionisio Ocotlán (CarC: $m = 0.065$, $SD = 0.027$) and Oaxaca (OxC: $m = 0.060$, $SD = 0.028$).

Gene flow analysis at a management type level (wild, cultivated, and nursery), revealed a higher proportion of migration from wild populations ($m = 0.171$; $SD = 0.016$) towards cultivated localities, followed by migration from wild populations to nursery sites ($m = 0.129$; $SD = 0.016$) (Fig 5A and S5 Table in S1 File).

Outlier loci detection and annotation

Our analysis of outliers only identified one significant SNP with BayeScan software, with $BH = 0.05$ and a p -value of 0.01. (S5 Fig in S1 File) Nevertheless, it was not possible to annotate it.

On the other hand, we identified 32 candidate outlier loci using *pcadapt* (Fig 6 and S6 Table in S1 File). The functional activity was annotated for the candidate loci, and 12 of these outliers belonged to the “No GO terms” (terms of Gene Ontology) category. Most annotated candidate loci associated to molecular functions (75%), followed by biological processes (21.875%), and the rest to cellular components (3.125%). We found outliers related to methyltransferase activity (GO:0008168), protein kinase activity (GO:0004672), protein (GO:0005515) adenosine triphosphate (ATP) (GO:0005524) binding, zinc ion binding (GO:0008270), protein phosphorylation (GO:0006468), flavin adenine dinucleotide (FAD) binding (GO:0006633), acyltransferase activity (GO:0016746), recognition of pollen (GO:0048544) and membrane components (GO:0016020).

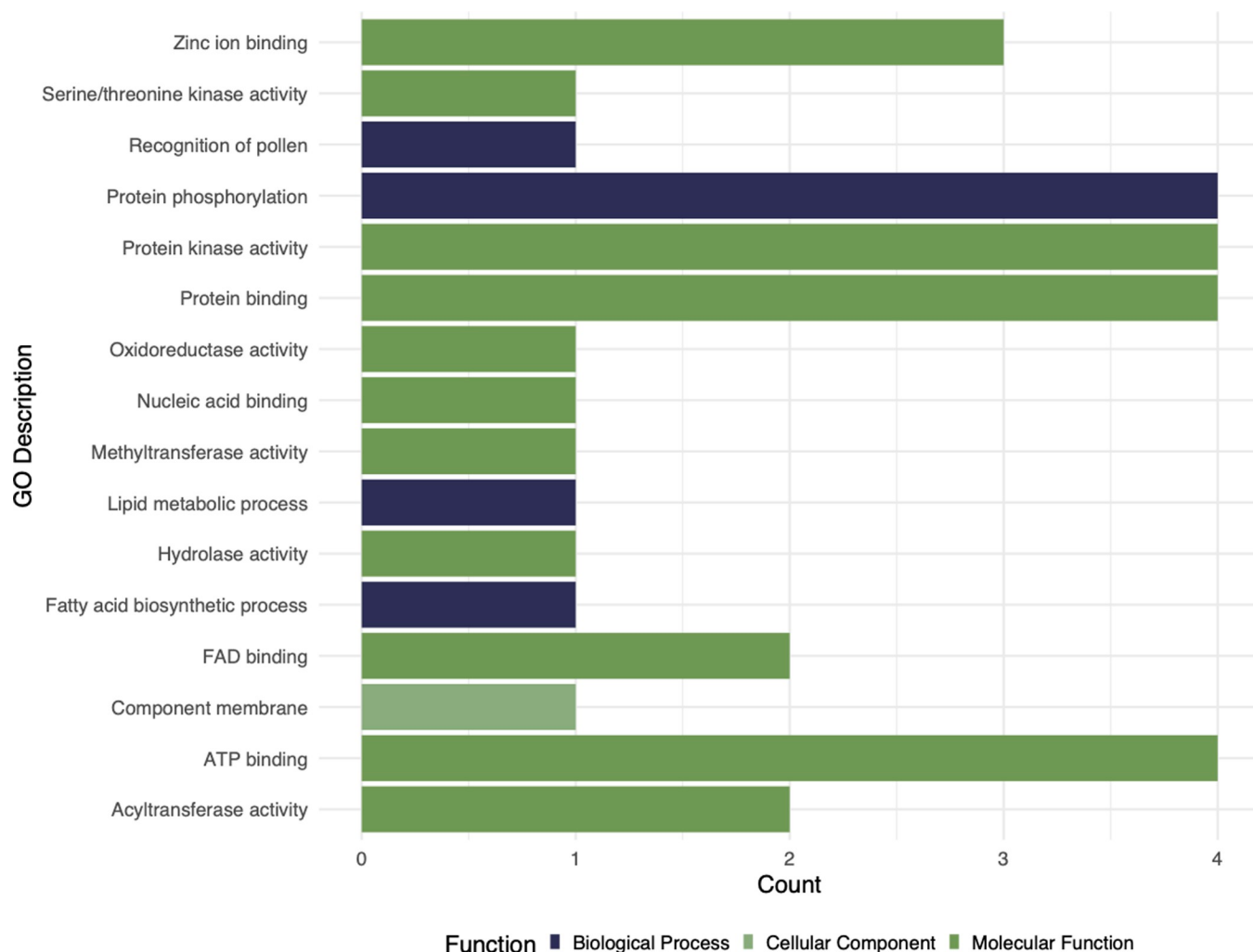


Fig 6. Annotation with Gene Ontology of outlier loci detected with pcadapt with possible selection signs in *Agave potatorum* from the Mexican states of Oaxaca and Puebla.

<https://doi.org/10.1371/journal.pone.0294534.g006>

Discussion

Genetic diversity

For plants with a long generation time and under incipient domestication, such as *Agave* species, understanding the evolution process and estimating their genetic diversity is essential for their conservation and management. Artificial selection creates bottlenecks in populations, and plants under cultivation generally represent only a limited portion of number of wild plants, which reduces the effective population size decreasing in turn their levels of genetic diversity. Therefore, it is crucial to set a baseline for diversity and differentiation metrics within these species.

In the case of *A. potatorum*, we found that the genetic diversity measured as expected heterozygosity was higher in the nursery and wild plants than in the cultivated samples ($H_E = 0.244$, $H_E = 0.234$, $H_E = 0.227$, respectively). We did not find private alleles in the cultivated or in the nursery plants, indicating that the plants under management came very recently from wild populations.

Expected heterozygosity values in *A. potatorum* ($H_T = 0.252$) were similar but lower than those previously reported for this species with dominant ISSR markers (i.e., $H_T = 0.302$ [16] and $H_T = 0.294$ [66]), and much lower than results based on microsatellites (SSRs), where very high genetic diversity was reported (wild: $H_T = 0.87$, populations under extraction, $H_T = 0.72$, and samples of a germplasm bank $H_T = 0.69$) [15, 67]. These differences may be due to differences in the mutation rates of the molecular markers used. Nevertheless, contrary to previous studies that did not detect a significant difference in genetic diversity between different management categories [15, 66, 67], we found that the cultivated samples of *A. potatorum* presented significant lower genetic diversity.

Similar diversity levels to the ones we obtained were reported in recent *Agave* studies using the same RADseq methodology. For instance, in *A. angustifolia* from Sonora, used to produce a type of mezcal called Bacanora, the value of genetic diversity ($H_T = 0.250$) was similar, with no differences between wild and cultivated plants [11]. In this species, similar management practices to Tobalá are used (i.e., incipient management and the introduction of plants and seeds from the wild). Our heterozygosity results were also similar to those reported using RADseq for the wild *A. angustifolia* from Oaxaca ($H_E = 0.24$), but lower than those founded in the cultivated *A. angustifolia* known as “Espadín” ($H_E = 0.29$) [68]. This last difference may result from different type of management practices between *A. angustifolia* “Espadín” and *A. potatorum*. *Agave angustifolia* has a long history of intensive clonal propagation as monoculture, with no genetic exchange with wild conspecifics. *Agave potatorum* management is recent and relies on wild plants and seeds. In contrast, we obtained a higher expected heterozygosity than the value estimated for the intensively cultivated, *A. tequilana* ($H_T = 0.120$) using RADseq [7].

Agave potatorum is a perennial species that takes 10 to 15 years to reach reproductive age. Due to this long generation time and seed harvest, the loss of diversity has apparently been gradual. Therefore, we argue that this species’ artificial selection and incipient management have not yet caused drastic differences between cultivated and wild individuals. In addition, *A. potatorum* has cross-pollination and usually reproduces only by seed and not clonally. Thus, we did not observe an excess of heterozygosity related to long-term clonal propagation found in other *Agave* species such as *A. tequilana* [7] and *A. angustifolia* [68].

In other perennial species, such as grapes [69], apples [70, 71], and cherries [27, 72] slight loss of diversity has been documented, suggesting that they maintain ~95% of the neutral variation found in wild populations [27], which seems to be the case for *A. potatorum*.

Inbreeding and kinship

Inbreeding in cultivated crops may be common, it can result from either selection, self-pollination, or mating with closely related individuals within the same cultivated plot [73]. The genetic consequence of inbreeding is an increase in homozygosity [74] and, in consequence, the expression of deleterious recessive alleles known as inbreeding depression.

In previous studies, a high level of inbreeding was reported in *A. potatorum* by using microsatellites ($F_{IS} = 0.267$ [15]). In contrast, we estimated lower inbreeding at the species level ($F_{IS} = 0.0827$ and $F_{hat3} = 0.150$), but we obtained significant differences between different types of management with F_{hat3} index (wild = 0.157; cultivated = 0.150; nursery = 0.121). We found positive levels of inbreeding (i.e., F_{IS} and F_{hat3} index) in several *A. potatorum* populations. This inbreeding is probably related to habitat fragmentation and isolation among populations, causing mating among relatives to be more frequent.

During our fieldwork, we observed extensive deforested areas with few or no plants, and only a handful of individuals at the reproductive stage. These factors, including habitat fragmentation, isolation, and inbreeding, can lead to reduced effective population sizes (N_e),

resulting in a loss of genetic variation [75–77]. Additionally, these plants thrive in specific climates, such as pine-oak forests and xerophytic scrublands; which further isolate them and limit their ability to reproduce with geographically distant populations.

In addition to the genetic problems that isolation and possible fragmentation generate, demographic and ecological issues [78–80] lead to a disruption in biotic interactions [77] (for example, plant-pollinator, or plant-microbiome relationships, especially association with mycorrhizas). For instance, in *Agave* spp., a density-dependent effect has been documented with their main pollinators, that are nectarivorous bats. Pollinators are more efficient and abundant in populations with a higher density of flowering plants [22]. A similar effect has been observed in the interaction with mycorrhizae, with greater efficiency in obtaining nutrients, minerals, and protection against root pathogens if the plants have a healthy microbiome [81].

We found a negative inbreeding index, indicating an excess of heterozygotes in two cultivated populations (SV2-C y Mat-C), and five wild localities (Cu7-W, Cu3-W, Ix-W, SJR5-W, SV1-W). In previously published works on the *Agave* genus, negative F_{IS} values have been found, especially in cultivated individuals of intensively managed species, such as *A. tequilana* $F_{IS} = -0.042$ [7] and *A. angustifolia* “Espadín” $F_{IS} = -0.07$ [68]. These values have been related to artificial selection for hybrid vigor and an accumulation of somatic mutations in the genome. However, due to the type of reproduction in *A. potatorum* (usually by seed) and incipient cultivation, it is possible that these negative values of the Inbreeding Coefficient (F_{IS}) are related to the fact that individuals with different genotypes or phenotypes tend to mate, resulting in excess of heterozygotes [74], or that selection eliminates the more inbred or homozygote individuals.

In the kinship analysis, we obtained very low values, most of them corresponding to unrelated individuals, which we consider to be the result of sexual reproduction and probably long-distance gene flow within this species.

Genetic structure and recent gene flow

The genetic structure reflects the balance between population cohesion and the process of divergence [82]. In general terms, the genetic structure is related to mating and migration since the probability that two individuals mate depends mainly on the geographical location of the potential mate and the dispersal capabilities of the species in question [74].

Agave species typically exhibit low genetic structure and differentiation [2, 3], a characteristic attributed to their long generation time, reproduction and pollination by highly mobile pollinators, in particular by bats. Our results confirmed this pattern as we found low genetic differentiation among populations of *A. potatorum* ($F_{ST} = 0.0796$). Similar values have been reported in the same species in two studies using ISSRs ($F_{ST} = 0.099$ [16], $F_{ST} = 0.079$ [66]). These authors attributed this low genetic structure to the exchange of pollen driven by the nectar-feeding bat *L. yerbabuena*. Nevertheless, higher genetic differentiation values were reported for the species based on the microsatellites study (SSRs: $F_{ST} = 0.389$); these results were attributed to ecological, geological, and climatic effects of the sampling area [15, 67].

In our study, the paired lowest values of genetic differentiation (F_{ST}) were found among San Juan Raya (SJR) populations in the south of the state of Puebla, and Cuicatlán, Oaxaca; both localities belonging to the Tehuacán-Cuicatlán Biosphere Reserve. This is a protected natural area located in the southeast of Mexico characterized by its great floristic diversity and high number of endemic species [83]. These localities are geographically close, and their distribution seems continuous. In contrast, we observed habitat fragmentation and isolation among populations in the Central Valley of Oaxaca. These patterns were also observed in the UPGMA cluster tree, where all the samples from San Juan Raya form a unique cluster.

Furthermore, isolation by distance analysis was high and significant ($r = 0.48$, $p < 0.001$), indicating that geographically close populations in general tend to be more genetically similar. We suggest that the low genetic differentiation is indeed related to pollinators' great mobility, mainly nectar-feeding bats. Indeed, we found genetic structure associated with geographic patterns, using the Admixture software, where the wild populations presented two gene pools (divided by Oaxaca and Puebla), with derived alleles towards nursery and cultivated localities.

At the finer scale, we found that genetic structure seems to be related to geography and ecology. For example, samples from Sola de Vega and Cuicatlán were found at a higher elevation (700–3,000 and 400–2,600 ms. n.m, respectively) along pine-oak forests, opposite to San Juan Raya, Puebla (average 1,700 ms. n.m), where the plants were collected in a xeric scrub. On the other hand, populations of the Central Valley, Oaxaca (700 a 1,600 ms.n.m.) were found at lower altitudes with a greater degree of deforestation and, therefore, less continuity and lower connectivity among populations.

Gene flow among populations limits genetic structure [82], helps to maintain genetic diversity and prevents inbreeding, which is especially important for small and fragmented populations [84]. Nevertheless, it requires that migrating individuals successfully contribute alleles to the mating pool of the population [74]. We estimated a higher gene flow from wild populations to cultivated sites, indicating that the plants under cultivation have been extracted from wild populations, or that the seed that originated them came from wild plants. However, at the population level, we found that gene flow is generally low. The highest inferred migration rate was in the localities of Sola de Vega and San Juan Raya; these populations are the best conserved and have less habitat fragmentation. As mentioned above, the main pollinators of *A. potatorum* are nectarivorous bats, including not only *Leptonycteris yerbabuenae*, also *Leptonycteris nivalis* (Saussure, 1860), and *Choeronycteris mexicana* (Tschudi, 1844) [22, 85]. The long-distance dispersal can counteract the effects of genetic differentiation, further providing a possible mechanism for introducing new alleles [86]. In these localities with less fragmentation, bats are usually more common, due to the dependent density effect.

Outlier loci detection and annotation

We consider it is relevant to generate information about the adaptive processes and identify potentially valuable genes for the cultivation and management of *A. potatorum* and other *Agave* species. This research can aid in future management, genetic conservation, and possible improvement in the face of global change. This data would be particularly relevant since a decrease in genetic diversity and inbreeding related to domestication may eventually cause reductions in adaptive potential and plasticity [77]. Additionally, local genetic adaptation is crucial to maintain phenotypic diversity in wild populations [87].

We found that the highest number of outlier loci with possible selection signals were related to molecular processes, of which we think the most noteworthy was related to protein phosphorylation, which plays an essential role in signal transduction and has a vital role in the life cycle of plants [88–91]. We also detected outlier loci related to protein kinase activity. This is the most important protein in signaling pathways, and can transfer phosphate groups with a significant role in enzyme regulation, gene expression, and transduction [92]. It is important to highlight that we found a protein family membership with S-receptor-like serine/threonine-protein kinase (SRK) (IPR024171). This protein has an important function in plants, as described in *Brassica* [93–95], where it is involved in the self-incompatibility mechanism that prevents inbreeding in and thereby increases genetic diversity [95–97].

Additionally, we also found an outlier locus related to the recognition of pollen, which is also relevant to Self-incompatibility protein (SI) and has the inability to produce zygotes after

self-pollination in a hermaphrodite plant [97]. It is important to remember that *Agave* species have protandrous hermaphrodite flowers, female and male sex are found in the same plant. However, the female gonads mature first than the male ones, limiting self-pollination [22].

Given the significance of organoleptic properties in *A. potatorum* during distillation, our focus was to identifying outliers related to these traits or those associated with sugar production. However, we did not detect any outliers related to these functions. Instead, we found outlier loci with potential signs of selection linked to genes involved in reproduction and response to abiotic stress that may be relevant to the local adaptation of both wild and cultivated plants.

The outlier analyses should be taken with caution, as they are based on genetic differences among populations. We suggest that in the future it would be beneficial to establish a database with gene annotations and functions for *Agave*, similar to databases available for other species like rice (<http://rapdb.dna.affrc.go.jp/>; <http://rice.plantbiology.msu.edu/>). This will require complete, well-resolved at chromosome levels genomes, transcriptomes, proteomes, and annotation of the proteins, at least in some *Agave* species, which for the moment have not yet been published.

Conservation implications and breeding strategies

The uncontrolled extraction of wild individuals and seeds increases the extinction risk of the populations [19] and, eventually, of the complete species, so we suggest some urgent steps to minimize these possibilities. First of all, it will be important to reduce future habitat fragmentation and the loss of wild populations of the species. It is also essential to increase the number of nurseries and shade houses for *A. potatorum*. This has been done in San Juan Raya (SJR), Puebla, which is the case of a successful collaborative effort between researchers and peasant communities initiated eight years ago to mitigate the overexploitation of wild *Agave* plants. As a tangible result, ca. 4000 plants were transplanted to natural vegetation after two years of growing in a shade house. Plants were obtained from seeds from a local *A. potatorum* population that were collected from spatially distant individuals, separated for more than 100 meters. After four years, 75% of transplanted plants survived [Valiente-Banuet et al., unpublished data]. Although the agroecological management at the SJR nursery can be considered a success, other *A. potatorum* producers should be aware of the importance of reducing the relatedness among cultivated individuals. Seeds should be collected from different mother plants and localities. These practices will decrease inbreeding and maintain the genetic diversity of the cultivated individuals. Nevertheless, it is also essential to maintain local adaptation and avoid outcrossing depression; therefore, the ecological characteristics of the habitat should also be considered. We believe that a global breeding program can also be implemented for *A. potatorum* where the wild samples' origin, initial genetic diversity and differentiation are registered. These practices will reduce inbreeding, inbreeding depression and serve as a gene bank for future breeding necessities and for adapting to the global changes.

On the other hand, successful *in situ* conservation efforts hinge on a careful consideration of the demography patterns of the wild populations, which can effectively serve as germplasm banks [26]. Additionally, the connectivity of these populations and the reproductive success of the plants relies mainly on the nectar feeding bats, making it essential to protect these primary pollinators.

Given the absence of significant genetic differences between wild and cultivated populations, we propose the reintroduction of a percentage of nursery-raised plants into the wild populations (as it has been done in SJR and SLA). Also, where appropriate, implementing agroecological management for this species *in situ*, [26] is recommend. This should take into

account the consideration for biotic interactions, such as promoting establishment and survival with nurse plants [98], as well as facilitating seedling establishment [85].

Conclusions

We found higher genetic diversity in the wild and nursery than in cultivated samples. Moreover, no private alleles related to management type (under cultivation and nursery) were found, meaning that most of the managed samples were recently derived from natural populations. Inbreeding levels varied widely among studied samples, apparently associated with isolation and fragmentation, but unrelated to human selection and management. Genetic differentiation at the species level was low. However, we found three genetic groups. Also, we identified higher gene flow in populations growing in less fragmented habitats. Furthermore, elevated gene flow from wild populations to the other two types of management was detected. Finally, we found loci with possible selection signals mainly related to biological processes and apparently to self-incompatibility.

As a conservation proposal, we suggest following the examples of SJR and SLA, with the production of plants in nurseries to avoid over-exploitation of wild plants and with the use of agroecological management, together with analysis of breeding and genetics, in order to trace the lineage of the plants and prevent inbreeding processes in populations of *A. potatorum*.

Supporting information

S1 File. Supporting information file containing multiple supporting figures (S1-S5 Fig) and tables (S1-S6 Tables).
(DOCX)

Acknowledgments

We are greatly thankful to the specialized technical support of Drs. Rosalinda Tapia and Marco Tulio Solano De la Cruz. We also thank Dra. Alejandra Moreno-Letelier (Institute of Biology, UNAM) for the information on the standardization of the RADseq in other *Agave* species. Finally, we acknowledge M. Sc. Tania Sánchez Ortiz and M. Sc. Noé Flores for their crucial help in field sampling.

We are particularly thankful to all the maguey and mezcal producers from Palenques “Convite”, “Medios Chiles”, “Hermogenes”, “Tierra Blanca”, “Hacienda Carreño”, and “Gota Fría” for plant access and general field work support.

Author Contributions

Conceptualization: Karen Y. Ruiz Mondragón, Luis E. Eguiarte.

Data curation: Karen Y. Ruiz Mondragón.

Formal analysis: Karen Y. Ruiz Mondragón, Anastasia Klimova.

Funding acquisition: Erika Aguirre-Planter, Alfonso Valiente-Banuet, Rafael Lira, Luis E. Eguiarte.

Investigation: Karen Y. Ruiz Mondragón, Anastasia Klimova, Luis E. Eguiarte.

Methodology: Karen Y. Ruiz Mondragón, Anastasia Klimova.

Project administration: Erika Aguirre-Planter, Alfonso Valiente-Banuet, Rafael Lira, Luis E. Eguiarte.

Resources: Karen Y. Ruiz Mondragón, Erika Aguirre-Planter, Alfonso Valiente-Banuet, Rafael Lira, Luis E. Eguiarte.

Software: Karen Y. Ruiz Mondragón, Anastasia Klimova.

Supervision: Karen Y. Ruiz Mondragón, Anastasia Klimova, Erika Aguirre-Planter, Luis E. Eguiarte.

Validation: Karen Y. Ruiz Mondragón, Anastasia Klimova.

Visualization: Karen Y. Ruiz Mondragón.

Writing – original draft: Karen Y. Ruiz Mondragón.

Writing – review & editing: Karen Y. Ruiz Mondragón, Anastasia Klimova, Erika Aguirre-Planter, Alfonso Valiente-Banuet, Rafael Lira, Guillermo Sanchez-de la Vega, Luis E. Eguiarte.

References

1. García-Mendoza A. J. Agavaceae. Pp 159–170. In: García Mendoza A. J., Ordoñez M. J., and Briones Salas M. (eds). Biodiversidad de Oaxaca. Instituto de Biología, UNAM, Fondo Oaxaqueño para la conservación de la Naturaleza, World Wildlife Fund, Mexico. 2004.
2. Eguiarte LE, Aguirre-Planter E, Aguirre X, Colín R, González A, Rocha M, Souza V. From Isozymes to genomics: population genetics and conservation of *Agave* in Mexico. The Botanical Review. 2013; 79 (4):483–506 <https://doi.org/10.1007/s12229-013-9123-x>
3. Eguiarte LE, Jiménez Barrón OA, Aguirre-Planter E, Scheinvar E, Gámez N, Gasca-Pineda J, et al. Evolutionary ecology of *Agave*: distribution patterns, phylogeny, and coevolution (an homage to Howard S. Gentry). American Journal of Botany. 2021; 108(2):216–235 <https://doi.org/10.1002/ajb2.1609> PMID: 33576061
4. Alducin-Martínez C, Ruiz Mondragón KY, Jiménez-Barrón O, Aguirre-Planter E., Gasca-Pineda J, Eguiarte LE, et al. Uses, Knowledge and Extinction Risk Faced by *Agave* Species in Mexico. Plants (Basel). 2022; 12(1):124. Published 2022 Dec 27. <https://doi.org/10.3390/plants12010124> PMID: 36616253
5. García-Mendoza A. J. Distribution of *Agave* (Agavaceae) in Mexico. Cactus and Succulent Journal. 2002; 74: 177–186.
6. Álvarez-Ríos GD, Pacheco-Torres F, Figueredo-Urbina CJ, Casas A. 2020. Management, morphological and genetic diversity of domesticated Agaves in Michoacán, Mexico. Journal of Ethnobiology and Ethnomedicine 16(1):1–17 <https://doi.org/10.1186/s13002-020-0353-9> PMID: 31948439
7. Ruiz Mondragon KY, Aguirre-Planter E, Gasca-Pineda J, Klimova A, Trejo-Salazar RE, Reyes Guerra MA, et al. Conservation genomics of *Agave tequilana* Weber var. azul: low genetic differentiation and heterozygote excess in the tequila *Agave* from Jalisco, Mexico vol. 10 e14398. 2022; <https://doi.org/10.7717/peerj.14398> PMID: 36415865
8. Instituto Nacional de Estadística y Geografía. INEGI. Conociendo la Industria del Tequila y el Mezcal. Colección de estudios sectoriales y regionales. 2019. <https://books.google.com.mx/books?id=gWD5DwAAQBAJ>
9. Servicio de Información Agroalimentaria y Pesquera, SIAP. Panorama agroalimentario 2020. Available at <https://www.gob.mx/siap>.
10. Torres I, Casas A, Vega E, Martínez-Ramos M, and Delgado-Lemus A. Population Dynamics and Sustainable Management of Mescal Agaves in Central Mexico: *Agave potatorum* in the Tehuacán- Cuicatlán Valley. Economic Botany. 2015; 69: 26–41. <https://doi.org/10.1007/s12231-014-9295-2>
11. Klimova A, Ruiz Mondragon KY, Molina Frenier F, Aguirre-Planter E, Eguiarte LE. Genomic analyzes of wild and cultivated bacanora *Agave* (*Agave angustifolia* var. *pacifica*) reveal inbreeding, few signs of cultivation history and shallow population structure. Plants. 2022; 11(11):1426 <https://doi.org/10.3390/plants11111426> PMID: 35684199
12. Casas A, Rangel-Landa S, Torres I, Pérez-Negrón E, Solís L, Parra F, et al. *In situ* management and conservation of plant resources in the Tehuacán-Cuicatlán Valley, Mexico: An ethnobotanical and ecological approach. In: Current topics in ethnobotany, eds. De Albuquerque U. P. and Ramos M., 1–25. Kerala, India: Research Signpost. 2008.

13. Delgado-Lemus A. Aprovechamiento y disponibilidad espacial de *Agave potatorum* en el Valle de Tehuacán-Cuicatlán. Mexico: Master thesis. Universidad Nacional Autónoma de México; 2008.
14. Delgado-Lemus A, Casas A, and Tellez O. Distribution, abundance and traditional management of *Agave potatorum* in the Tehuacán Valley, Mexico: Bases for sustainable use of non-timber forest products. *Journal of Ethnobiology and Ethnomedicine* 2014; 10(63).
15. Félix-Valdez LI, Vargas-Ponce O, Cabrera-Toledo D, Casas A, Cibrian-Jaramillo A, from the Cruz-Larios L. 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*. 2016; 63:1255–1271 <https://doi.org/10.1007/s10722-015-0315-6>
16. Aguirre Dugua X and Eguiarte LE. Genetic diversity, conservation and sustainable use of wild *Agave cupreata* and *Agave potatorum* extracted for mezcal production in Mexico. *Journal of Arid Environments*. 2013; 90:36–44.
17. Colunga-GarcíaMarín P, Zizumbo-Villarreal D, and Martínez-Torres J. Tradiciones en el aprovechamiento de los Agaves mexicanos: Una aportación a su protección legal y conservación biológica y cultural. In: En lo ancestral hay futuro: Del tequila, los mezcales y otros Agaves, eds. Colunga-GarcíaMarín P., Eguiarte L., Larqué A., and Zizumbo-Villarreal D., 229–248. Mexico: CICY-CONA-CYTONABIO-INE. 2007.
18. Gentry HS. Agaves of continental North America. The University of Arizona Press, Tucson, USA. 1982.
19. Valiente-Banuet A. Mezcal Boom and Extinction Debts. In: Jones RW, Ornelas-García CP, Pineda-López R, Álvarez F. Mexican Fauna in the Anthropocene. Springer, Cham. 2023. https://doi.org/10.1007/978-3-031-17277-9_14
20. García-Mendoza AJ. Revisión taxonómica del complejo *Agave potatorum* Zucc. (Agavaceae): nuevos taxa y neotipificación. *Acta Botánica Mexicana*. 2010; 91: 71–93.
21. Valiente-Banuet A. and Verdú M. Mexical Shrubland. In: Goldstein M.I Dellasala D.A. Encyclopedia of the World Biomes. 2020. Vol. 3. Elsevier, pp. 532–545. USA.
22. Estrella-Ruiz P. Efecto de la explotación humana en la biología de la polinización de *Agave salmiana* y *Agave potatorum* en el Valle de Tehuacán- Cuicatlán. Tesis de maestría. Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México. Mexico, D. F. 2008.
23. Torres I. Dinámica poblacional de dos morfos de *Agave potatorum* Zucc. en el Valle de Tehuacán: bases para su manejo sustentable. Tesis de Maestría, Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México. Mexico, D.F. 2009.
24. Ober HK, and Steidl R. Foraging rates of *Leptoncyrtis curasoae* Vary with characteristics of *Agave palmeri*. *Southwestern Naturalist*. 2004; 49:68–74.
25. García-Mendoza, A.J., Sandoval-Gutiérrez, D., Torres-García, I. & Casas, A. 2019. *Agave potatorum*. The IUCN Red List of Threatened Species 2019: e.T115690109A116354368. <https://doi.org/10.2305/IUCN.UK.2019-3.RLTS.T115690109A116354368.en>. Accessed on March 13, 2023.
26. Torres I, Casas A, Delgado-Lemus A, Rangel-Landa S. Aprovechamiento, demografía y establecimiento de *Agave potatorum* en el Valle de Tehuacán, Mexico: aportes etnobiológicos y ecológicos para su manejo sustentable. *Zo. Arid*. 2013. 15, 92–109.
27. Gaut BS, Díez CM, and Morrell PL. Genomics and the Contrasting Dynamics of Annual and Perennial Domestication. *Trends in Genetics*. 2015; 31(12), 709–719. <https://doi.org/10.1016/j.tig.2015.10.002> PMID: 26603610
28. Hancock JF. Contributions of domesticated plant studies to our understanding of plant evolution. *Ann. Bot. (Lond.)* 2005; 96, 953–963. <https://doi.org/10.1093/aob/mci259> PMID: 16159942
29. Purugganan MD, and Fuller DQ. The nature of selection during plant domestication. *Nature*. 2009; 457, 843–848 <https://doi.org/10.1038/nature07895> PMID: 19212403
30. Salamini F, Özkan H, Brandolini A. Brandolini A, Schäfer-Pregl R, and Martin W. Genetics and geography of wild cereal domestication in the near east. *Nat Rev Genet* 3. 2002; 429–441. <https://doi.org/10.1038/nrg817>
31. Doebley JF, Gaut BS, and Smith BD. The molecular genetics of crop domestication. *Cell*. 2006; 127(7), 1309–1321. <https://doi.org/10.1016/j.cell.2006.12.006> PMID: 17190597
32. Vaughan DA, Balázs E, and Heslop-Harrison JS. From crop domestication to super-domestication. *Annals of botany*. 2007; 100(5), 893–901. <https://doi.org/10.1093/aob/mcm224> PMID: 17940074
33. Eguiarte LE, Helena S. Hernández-Rosales, Josué Barrera-Redondo, Gabriela Castellanos-Morales, Leslie M. Paredes-Torres, Guillermo Sánchez-de la Vega, Karen Y. Ruiz-Mondragón, Alejandra Vázquez-Lobo, Salvador Montes-Hernández, Erika Aguirre-Planter, Valeria Souza, Rafael Lira. Domesticación, diversidad y recursos genéticos y genómicos de México: el caso de las calabazas. *TIP Revista Especializada en Ciencias Químico Biológicas*. 2018; 21, 85–101. (<https://doi.org/10.22201/fesz.23958723e.2018.0.159>)

34. Miller AJ, and Gross BL. From forest to field: perennial fruit crop domestication. *Am. J. Bot.* 2011; 98, 1389–1414 <https://doi.org/10.3732/ajb.1000522> PMID: 21865506
35. McClure KA, Sawler J, Gardner KM, Money D, and Myles S. Genomics: a potential panacea for the perennial problem. *American journal of botany.* 2014; 101(10), 1780–1790. <https://doi.org/10.3732/ajb.1400143> PMID: 25326620
36. Gepts P, Papa R. Evolution during domestication. In: *Encyclopedia of Life Sciences.* 2003. Macmillan Publishers <https://doi.org/10.1038/npg.els.0003071>
37. Hardigan MA, Laimbeer FPE, Newton L, Crisovan E, Hamilton JP, Vaillancourt B, et al. Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc Natl Acad Sci U S A.* 2017 Nov 14; 114(46):E9999–E10008. <https://doi.org/10.1073/pnas.1714380114> Epub 2017 Oct 30. PMID: 29087343; PMCID: PMC5699086.
38. Kantar MB, Bruford MW, Rieseberg LH. The genomics of domestication special issue editorial. *Evol Appl.* 2019; 12: 3–5. <https://doi.org/10.1111/eva.12693>
39. Doyle JJ, Doyle JL. A rapid DNA isolation produce for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 1987; 19(1):11–15.
40. Eaton D, Overcast I. Ipyrad: interactive assembly and analysis of RADseq datasets. *Bioinformatics.* 2020; 36(8):2592–2594 <https://doi.org/10.1093/bioinformatics/btz966> PMID: 31904816
41. Gross SM, Martin JA, Simpson J, Abraham-Juarez MJ, Wang Z, Visel A. De novo transcriptome assembly of drought tolerant CAM plants, *Agave deserti* and *Agave tequilana*. *BMC Genomics.* 2013; 14 (1):563 <https://doi.org/10.1186/1471-2164-14-563> PMID: 23957668
42. Danecek P, Auton A, Abecasis G, Cornelis A, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011; 27(15):2156–2158 <https://doi.org/10.1093/bioinformatics/btr330> PMID: 21653522
43. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics.* 2007; 81(3):559–575 <https://doi.org/10.1086/519795> PMID: 17701901
44. Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ.* 2014; 2:e281 <https://doi.org/10.7717/peerj.281> PMID: 24688859
45. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics.* 2008; 24(11):1403–1405 <https://doi.org/10.1093/bioinformatics/btn129> PMID: 18397895
46. Jombart T, Ahmed I. adegenet 1.3–1: new tools for the analysis of genome-wide SNP data. *Bioinformatics.* 2011; 27(21):3070–3071 <https://doi.org/10.1093/bioinformatics/btr521> PMID: 21926124
47. Goudet J. hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes.* 2005; 5(1):184–186 <https://doi.org/10.1111/j.1471-8286.2004.00828.x>
48. R Core Team R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. 2020. Available at <https://www.R-project.org/>.
49. Stoffel MA, Esser M, Kardos M, Humble E, Nichols H, David P, et al. inbreedR: an R package for the analysis of inbreeding based on genetic markers. *Methods in Ecology and Evolution.* 2016; 7 (11):1331–1339 <https://doi.org/10.1111/2041-210X.12588>
50. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics* (Oxford, England). 2010; 26(22), 2867–2873. <https://doi.org/10.1093/bioinformatics/btq559> PMID: 20926424
51. Wright S. The genetic structure of populations. *Annals of Eugenics.* 1951; 15(4):323–354 <https://doi.org/10.1111/j.1469-1809.1949.tb02451.x> PMID: 24540312
52. Weir BS, Cockerham CC. Estimating F Statistics for the ANalysis of Population Structure. *Evolution.* 1984; 38, 1358–1370. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x> PMID: 28563791
53. Pembleton LW, Cogan NO, Forster JW. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources.* 2013; 13(5):946–952 <https://doi.org/10.1111/1755-0998.12129> PMID: 23738873
54. Oksanen, F.J., et al., *Vegan: Community Ecology Package.* R package Version 2.4–3. 2017.
55. Ersts, P.J. Geographic Distance Matrix Generator (v.1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation [online]. 2016. Website http://biodiversityinformatics.amnh.org/open_source/gdmg [accessed 19-11-2022].
56. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research.* 2009; 19(9):1655–1664 <https://doi.org/10.1101/gr.094052.109> PMID: 19648217
57. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics.* 2011; 12(1):246 <https://doi.org/10.1186/1471-2105-12-246> PMID: 21682921

58. Musmann SM, Douglas MR, Chafin TK, Douglas ME. BA3-SNPs: contemporary migration reconFigd in BayesAss for next-generation sequence data. *Methods in Ecology and Evolution*. 2019; 10(10):1808–1813 <https://doi.org/10.1111/2041-210X.13252>
59. Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*. 2008 Oct; 180(2):977–93. <https://doi.org/10.1534/genetics.108.092221> Epub 2008 Sep 9. PMID: 18780740; PMCID: PMC2567396.
60. Foll M, Fischer MC, Heckel G and Excoffier L. Estimating population structure from AFLP amplification intensity. *Molecular Ecology* 2010; 19: 4638–4647 <https://doi.org/10.1111/j.1365-294X.2010.04820.x> PMID: 20874760
61. Benjamini Y, and Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*. 1995; 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
62. Luu K, Bazin E, and Blum M GB. pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*. 2017; 17(1), 67–77. <https://doi.org/10.1111/1755-0998.12592> PMID: 27601374
63. Privé F, Luu K, Vilhjálmsson BJ, and Blum M GB. Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. *Molecular Biology and Evolution*. 2020.
64. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: Architecture and applications. *BMC Bioinformatics*, 2009; 10 (421). <https://doi.org/10.1186/1471-2105-10-421> PMID: 20003500
65. Zdobnov EM, Apweiler R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*. 2001 Sep; 17(9):847–8. <https://doi.org/10.1093/bioinformatics/17.9.847> PMID: 11590104.
66. Scheinvar Gottdiener E. Genética de poblaciones silvestres y cultivadas de dos especies Mezcaleras: *Agave cupreata* y *Agave potatorum*. Tesis Maestría en Ciencias Biológicas, Instituto de Ecología, UNAM. 2008
67. Félix Valdez LI. Patrones de Diversidad y Estructura Genética de *Agave potatorum* Zucc. E implicaciones en su manejo in-situ. Tesis de Maestría. Centro Universitario de Ciencias Biológicas y Agropecuarias. Universidad De Guadalajara. 2014.
68. Klimova A, Ruiz Mondragón KY, Aguirre-Planter E, Valiente A, Lira R, Eguiarte L. Genomic analysis unveils reduced genetic variability but increased proportion of heterozygotic genotypes of the intensively managed mezcal agave, *Agave angustifolia*. *Am J Bot*. 2023. <https://doi.org/10.1002/ajb2.16216> Epub 2023 Aug 25. PMID: 37478873.
69. Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, et al. Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108(9), 3530–3535. <https://doi.org/10.1073/pnas.1009363108> PMID: 21245334
70. Cornille A, Giraud T, Smulders MJ, Roldán-Ruiz I, Gladieux P. The domestication and evolutionary ecology of apples. *Trends in genetics: TIG*. 2014; 30(2), 57–65. <https://doi.org/10.1016/j.tig.2013.10.002> PMID: 24290193
71. Gross BL, Henk AD, Richards CM, Fazio G, and Volk GM. 2014. Genetic diversity in *Malus × domestica* (Rosaceae) through time in response to domestication. *American journal of botany*. 2014; 101(10), 1770–1779. <https://doi.org/10.3732/ajb.1400297> PMID: 25326619
72. Mariette S, Tavaud M, Arunyawat U, Capdeville G, Millan M, and Franck Salin. Population structure and genetic bottleneck in sweet cherry estimated with SSRs and the gametophytic self-incompatibility locus. *BMC Genet*. 2010; 11, 77. <https://doi.org/10.1186/1471-2156-11-77> PMID: 20727153
73. Nielsen R. and Slatkin M. *An Introduction to Population Genetics: Theory and Applications*. Sinauer Associates is an imprint of Oxford University. 2013.
74. Hamilton MB. *Population genetics*. Chichester, UK: Wiley-Blackwell; 2009.
75. Ellstrand NC, Elam DR. Population genetic consequences of small population size: implications for plant conservation. *Annu. Rev. Ecol. Syst*. 1993; 24: 217–243.
76. Young A, Boyle T, Brown T. The population genetic consequences of habitat fragmentation for plants. *Trends Ecol. Evol*. 1996; 11: 413–418. [https://doi.org/10.1016/0169-5347\(96\)10045-8](https://doi.org/10.1016/0169-5347(96)10045-8) PMID: 21237900
77. Leimu R, Vergeer P, Angeloni F, Ouborg NJ. Habitat fragmentation, climate change, and inbreeding in plants. *Ann N Y Acad Sci*. 2010 May; 1195:84–98. <https://doi.org/10.1111/j.1749-6632.2010.05450.x> PMID: 20536818.
78. Aguilar R L, Ashworth L, Galetto M, Aizen A. Plant reproductive susceptibility to habitat fragmentation: review and synthesis through a meta-analysis. *Ecol. Lett*. 2006; 9: 968–980. <https://doi.org/10.1111/j.1461-0248.2006.00927.x> PMID: 16913941

79. Rathcke BJ, Jules ES. Habitat fragmentation and plant-pollinator interactions. *Curr. Sci.* 1993; 65: 273–278.
80. Olesen JM, Jain SK. Fragmented plant populations and their lost interactions. In: Conservation Genetics. Loeschcke V, Tomiuk J. and Jain S.K., Eds.: 1994; 417–426. Birkhauser.
81. Hernández-Morales JL, López-Sánchez C, Palma-Cruz FJ. Caracterización morfológica de micorriza arbuscular asociada a *Agave potatorum* Zucc. con potencial de uso agronómico. *Revista Mexicana de Agroecosistemas*. Vol. 1(2): 82–93, 2014
82. Duminil J, Hardy OJ, Petit RJ. Plant traits correlated with generation time directly affect inbreeding depression and mating system and indirectly genetic structure. *BMC Evol Biol.* 2009 Jul 27; 9:177. <https://doi.org/10.1186/1471-2148-9-177> PMID: 19635127; PMCID: PMC2728730.
83. Valiente-Banuet A, Solís L, Dávila P, Arizmendi M, Del C, Silva C, Ortega-Ramírez J, et al. Guía de vegetación del Valle de Tehuacán- Cuicatlán. Universidad Nacional Autónoma de México, Fundación Cuicatlán A.C, Ediciones Margen Rojo, México. 2009.
84. Ksiazek-Mikenas K, Fant JB and Skogen KA. Pollinator-Mediated Gene Flow Connects Green Roof Populations Across the Urban Matrix: A Paternity Analysis of the Self-Compatible Forb *Penstemon hirsutus*. *Front. Ecol. Evol.* 2019; 7:299. <https://doi.org/10.3389/fevo.2019.00299>
85. Valiente-Banuet A, Verdú M. Human impacts on multiple ecological networks act synergistically to drive ecosystem collapse. *Frontiers in Ecology and the Environment*. 2013; 11: 408–413. <https://doi.org/10.1890/130002>
86. Trejo-Salazar RE, Eguarte LE, Suro-Piñera D, Medellín RA. Save our bats, save our tequila: industry and science join forces to help bats and Agaves. *Natural Areas Journal*. 2016; 36(4):523–530 <https://doi.org/10.3375/043.036.0417>
87. Merilä J, Crnokrak P. Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*. 2001; 14: 892–903. <https://doi.org/10.1046/j.1420-9101.2001.00348.x>
88. Damaris RN, Yang P. Protein Phosphorylation Response to Abiotic Stress in Plants. In: Wu X.N. (eds) *Plant Phosphoproteomics. Methods in Molecular Biology*. vol 2358. Humana, New York, NY. 2021. https://doi.org/10.1007/978-1-0716-1625-3_2
89. Seet BT, Dikic I, Zhou MM, and Pawson T. Reading protein modifications with interaction domains. *Nature reviews. Molecular cell biology*. 2006; 7(7), 473–483. <https://doi.org/10.1038/nrm1960>
90. Pawson T, Scott JD. Signaling through scaffold, anchoring, and adaptor proteins. *Science*. 1997; 278 (5346):2075–2080. <https://doi.org/10.1126/science.278.5346.2075> PMID: 9405336
91. Lin S, Chen L, Tao H, Huang J, Xu C, Li L, et al. Impact of SNPs on Protein Phosphorylation Status in Rice (*Oryza sativa* L.). *Int J Mol Sci*. 2016 Nov 11; 17(11):1738. <https://doi.org/10.3390/ijms17111738> PMID: 27845739; PMCID: PMC5133773.
92. Taiz L, Zeiger E. *Plant physiology*. 4th Edition, Sinauer Associates, Inc., Sunderland. 2006.
93. Stein JC, Howlett B, Boyes DC, Nasrallah ME, Nasrallah JB. Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. *Proceedings of the National Academy of Sciences of the United States of America*. 1991 Oct; 88(19):8816–8820. <https://doi.org/10.1073/pnas.88.19.8816> PMID: 1681543; PMCID: PMC52601.
94. Takasaki T, Hatakeyama K, Suzuki G, Watanabe M, Isogai A, Hinata K. The S receptor kinase determines self-incompatibility in *Brassica stigma*. *Nature*. 2000; 403(6772):913–916. <https://doi.org/10.1038/35002628> PMID: 10706292
95. Hiscock SJ, Tabah DA. The different mechanisms of sporophytic self-incompatibility. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. 2003 Jun; 358 (1434):1037–1045. <https://doi.org/10.1098/rstb.2003.1297> PMID: 12831470; PMCID: PMC1693206
96. De Nettancourt D. *Incompatibility and Incongruity in Wild and Cultivated Plants*, 2nd ed., Springer, Berlin, Heidelberg, New York. 2001.
97. Watanabe M, Suwabe K, Suzuki G. Molecular genetics, physiology and biology of self-incompatibility in Brassicaceae. *Proc Jpn Acad Ser B Phys Biol Sci*. 2012; 88(10):519–35. <https://doi.org/10.2183/pjab.88.519> PMID: 23229748; PMCID: PMC3552045.
98. Rangel- Landa S. Establecimiento de *Agave potatorum* en el Valle de Tehuacán, Puebla. Tesis de maestría. Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México. México, D.F. 2009.