RECENT ADVANCES AND FUTURE PERSPECTIVES FOR AGAVOIDEAE RESEARCH: *AGAVE*, *YUCCA* AND RELATED TAXA

EDITED BY: Luis Enrique Eguiarte, Jim Leebens-Mack and Karolina Heyduk

PUBLISHED IN: Frontiers in Plant Science and Frontiers in Agronomy

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-101-7
DOI 10.3389/978-2-88971-101-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

RECENT ADVANCES AND FUTURE PERSPECTIVES FOR AGAVOIDEAE RESEARCH: *AGAVE, YUCCA* AND RELATED TAXA

Topic Editors:
Luis Enrique Eguiarte, University of Mexico, Mexico
Jim Leebens-Mack, University of Georgia, United States
Karolina Heyduk, University of Hawaii, United States

Citation: Eguiarte, L. E., Leebens-Mack, J., Heyduk, K., eds. (2021). Recent Advances and Future Perspectives for Agavoideae Research: *Agave, Yucca* and Related Taxa. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-101-7

Table of Contents

frontiers in Plant Science

EDITORIAL
published: 10 May 2021
doi: 10.3389/fpls.2021.687596

Editorial: Recent Advances and Future Perspectives for Agavoideae Research: *Agave*, *Yucca* and Related Taxa

Luis E. Eguiarte[1]*, *James Leebens-Mack*[2]* and *Karolina Heyduk*[3]*

[1] Laboratorio de Evolución Molecular y Experimental, Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico, [2] Department of Plant Biology, University of Georgia, Athens, GA, United States, [3] School of Life Sciences, University of Hawai'i at Mãnoa, Honolulu, HI, United States

Keywords: arid and semiarid climates, biofuel, desert, hybridization, mescal, physiological adaptations, pollination mutualism, pulque

Edited by:
Stefan Wanke, Technische Universität Dresden, Germany

Reviewed by:
Carolina Granados Mendoza, National Autonomous University of Mexico, Mexico

***Correspondence:**
Luis E. Eguiarte
fruns@unam.mx
James Leebens-Mack
jleebensmack@uga.edu
Karolina Heyduk
heyduk@hawaii.edu

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *29 March 2021*
Accepted: *12 April 2021*
Published: *10 May 2021*

Citation:
Eguiarte LE, Leebens-Mack J and Heyduk K (2021) Editorial: Recent Advances and Future Perspectives for Agavoideae Research: Agave, Yucca and Related Taxa. Front. Plant Sci. 12:687596. doi: 10.3389/fpls.2021.687596

Editorial on the Research Topic

Recent Advances and Future Perspectives for Agavoideae Research: *Agave*, *Yucca* and Related Taxa

The Agavoideae (Asparagaceae) are a group of charismatic plants native to the Americas (if including the genus *Hosta*, also Asia) that are very diverse in their ecology and comprise 12 genera and ca. 445 species. Species in the Agavoideae display impressive adaptions in order to live in arid and semiarid conditions, have been historically very important for local human populations, and today are gaining increased attention due to their immense biotechnological potential (Davis et al., 2011; Cushman et al., 2015). In this collection we include 15 papers concerning a variety of cutting-edge studies on different aspects of the biology of the Agavoideae, which together illuminate their evolution and uniqueness.

Papers in this collection focus on the two largest genera in the Agavoidae: *Yucca* and *Agave*. Hybridization is not uncommon in *Yucca*, despite a long history of specialized pollination mutualism in this genus (Pellmyr and Leebens-Mack, 1999; McKain et al., 2016). Royer et al. described a hybrid zone in the *Y. brevifolia* and *Y. jaegeriana* (Joshua trees) from the Mojave desert that are pollinated by different *Tegeticula* moth species. The *Yucca* species present different sizes, with *Y. brevifolia* taller and with larger flowers than *Y. jaegeriana,* and morphological differences exist between their pollinators as well. Royer et al. analyzed a narrow (ca. 4 km in width) hybrid zone between these species in the Tikaboo Valley, southern Nevada, using RAD-seq data, microsatellite loci, chloroplast variation, vegetative and floral traits, and pollinator frequency. They found overlapping genomic and pollinator clines, consistent with a narrow hybrid zone generated by strong selection, but with wider phenotypic and a chloroplast clines.

In the Baja California peninsula, Arteaga et al. studied the populations of two closely related *Yucca* species: *Y. valida* that lives mostly in the central section of the peninsula and *Y. capensis* that is only found in the southern tip of the peninsula. They confirmed the hybrid origin of geographically and morphologically intermediate populations using nextRAD derived SNP along Principal Components and Structure analysis and Approximate Bayesian computation simulations. This conclusion was supported by the distribution models constructed using climatic data for the present and the past.

In the Atlantic coast of North America, Heyduk et al. studied the hybrid origin *of Y. gloriosa*, a putatively hybrid species between *Y. aloifolia* and *Y. filamentosa*, which is photosynthetically

intermediate (Rentsch and Leebens-Mack, 2012; Heyduk et al., 2016). The authors used whole genome shotgun data to assemble complete chloroplast genomes in the three species. They found that *Y. gloriosa* chloroplast haplotypes were nested in three separate clades: one related to *Y. filamentosa,* and two related to *Y. aloifolia,* supporting a hybrid but complex origin for *Y. gloriosa*. They also analyzed the transposons in their nuclear sequencing reads: while overall repetitive content varied between the three species, expression patterns showed little increased transcriptional activity of transposons in *Y. gloriosa*, suggesting that no transposon release occurred in the hybrid.

Jolly et al. described different morphological and physiological adaptations to deal with the dry conditions in *Y. brevifolia* and *Hesperoyucca whipplei,* two species from the Mojave desert. The authors suggested that the ability of *H. whipplei* to adjust both its vein density and stomatal density allows for higher gas exchange, thus permitting this species to grow in drier conditions than *Y. brevifolia*.

Most of the papers in this collection concentrate in studies of different species in *Agave*. Jiménez-Barron et al. analyzed the phylogeny, divergence times, and speciation rates in *Agave sensu lato*, a clade containing more than 250 species. They concluded that the genus is organized in different main clades. One clade is formed by the Striatae group, which is the sister group to the rest of the *Agave sensu lato*. Another clade is formed by the herbaceous taxa *Manfreda*, *Polianthes,* and *Prochnyanthes,* that diverged from other linages within *Agave sensu lato* ca. 3.55 Ma. Within *Agave*, they found two significant diversification shifts: one soon after the origin of *Agave sensu lato*, at ca. 6.18 Ma, and a second within *Agave sensu stricto*, ca. 2.68 Ma.

The *Agave* genus is not only fascinating because of its diversity and adaptions, but also because these plants were central for the development of the American native cultures and civilizations. Ortiz-Cano et al. reviewed their physiological adaptations to dry environments and their traditional uses, in particular for making drinks like mescal, tequila, and pulque, and their potential new applications as alternative crops for biofuel production. The authors focused on the Hohokam, a Pre-Columbian Indigenous People from the Sonoran desert that used "rock mulching" (i.e., rock piles) that helped to cultivate agaves by harvesting during rainfall and retaining soil moisture.

Cabrera-Toledo et al. analyzed the levels of morphological and genetic variation and differentiation patterns in *A. maximiliana,* used in the production of raicilla, a local kind of mescal in Western-Central Mexico. The authors found strong morphological and genetic differentiation between populations ($F_{ST} = 0.43$), and isolation by distance in the genetic markers. The authors compared plants in different management categories (monoculture, managed forestry systems, and wild populations), but they did not find differences in morphology nor in genetic diversity among these categories. Torres-García et al. studied the demography of *A. inaequidens*, another species used in the production of raicilla and other types of mescal, across four wild populations in Central Western Mexico. The study indicated that analyzed wild populations are stable, concluding that extraction rates from 10 to 30% of mature individuals for making mescal could be sustainable, but only if 200 to 300 agave plants of the younger sizes are introduced every year into the populations.

Another group of agaves used for the production of pulque–a traditional Mexican alcoholic beverage produced through open fermentation of agave sap—was studied by Trejo et al. They analyzed vegetative traits of landraces of *A. salmiana* and *A. mapisaga* in Tlaxcala, a highland state in central east Mexico, and their genetic relationships using a chloroplast region (*trnL*) and nuclear ITS sequences. Both, data on the morphology and genetic markers of the landraces aligned with the species classification, with the exception of *A. salmiana* subsp. *salmiana* "Ayoteco," which is more related to *A. mapisaga* var. *mapisaga*. They concluded that low intensity artificial selection together with gene flow and plasticity could explain the high number of phenotypically similar landraces.

Agave kerchovei has a restricted distribution, mainly in the Tehuacán -Cuicatlán Valley, in the states of Oaxaca and Puebla. Aguirre-Planter et al. studied the chloroplast variation and both current and past ecological niche models, finding high levels of total chloroplast genetic variation and very strong genetic differentiation (i.e., $F_{ST} = 0.92$). They suggest that the Pleistocene glacial cycles played a critical role in changing the distribution and population sizes of the species.

The interaction with other species seems to be very important in the evolutionary ecology of agavoid species, in particular mutualistic pollinator interactions (e.g., the *Tegeticula* moth in *Yucca* and the nectar feeding *Leptonycteris* bats in *Agave;* Pellmyr and Leebens-Mack, 1999; McKain et al., 2016: Eguiarte et al., 2021). But microbes, including both fungi and bacteria, also seem to play very important roles in the biology of the Agavoideae. In *A. lechuguilla,* in Cuatro Ciénegas, in the Chihuahuan desert, López-Lozano et al. analyzed the microbial communities in the soil and in its rhizosphere using 16S rRNA gene sequences. Actinobacteria were more abundant in the soil, while Proteobacteria dominated in the rhizosphere. The authors reported differences in bacterial diversity, community composition, potential functions of the bacteria taxa, and interaction networks between the soil and *A. lechuguilla* rhizosphere.

The plants of the Agavoideae have many other physiological and biochemical adaptations and unique chemical profiles. Lledías et al. isolated, characterized and analyzed the molecular function and evolution of mayahuelin, a protein that is abundant in the central spike of the rosette in *A. tequilana* var. *azul* and in other agaves, making up to 20% or even more of the total protein. The protein is a type I Ribosome inactivating protein (RIP). The authors also used this gene to identify accessions closely related to *A. tequilana* var. *azul*. Agaves store their carbohydrates in the form of fructan polymers, instead of starch or sucrose. Pérez-López and Simpson describe the fructans in *Agave* that provide a source of carbohydrates for the transition from vegetative to reproductive stages; for this reason, inflorescences are removed to avoid depletion of fructan reserves in cultivated and managed populations of *Agave*.

The striking physiological adaptations of *Agave* for living in dry areas are highlighted in the study of Jones et al. where

they conducted field experiments in *A. americana* to evaluate their ethanol yield potential, analyzing their efficiency in water use, their carbohydrates production, and the products obtained by enzymatic hydrolysis comparing with other plants—grasses–used to produce biofuels. They concluded that *A. americana* was able to produce important amounts of soluble carbohydrates by using very little water, and thus have an important potential for future biofuel production.

Finally, this collection also includes a methodological paper by González-Gutiérrez et al. that describes how to analyze the F-actinin in plant species with thick tissues, as is the case for Agavoideae plants.

This collection analyses only a few species, adaptations and uses from a very large group of fascinating plants, including agaves called "el árbol de las maravillas" (the tree of wonders) by José de Acosta (de Acosta, 1608), a Jesuits missionary and, among yuccas, Joshua Tree described as "The most repulsive tree in the Vegetable Kingdom" by American explorer Lt. John C. Fremont in 1844 (we disagree). We hope that this collection of studies will help to increase the awareness of the Agavoideae, and inspire future work on the group's striking diversity of form and ecological function.

AUTHOR CONTRIBUTIONS

LE, JL, and KH proposed the Research Topic, edited manuscripts, and wrote the editorial. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We want to acknowledge the efforts and support of all the reviewers, invited editors and authors that helped to make this collection of articles.

REFERENCES

Cushman, J. C., Davis, S. C., Yang, X., and Borland, A. M. (2015). Development and use of bioenergy feedstocks for semi-arid and arid lands. *J. Exp. Bot.* 66, 4177–4193. doi: 10.1093/jxb/erv087

Davis, S. C., Dohleman, F. G., and Long, S. P. (2011). The global potential for *Agave* as a biofuel feedstock. *Gcb Bioenergy* 3, 68–78. doi: 10.1111/j.1757-1707.2010.01077.x

de Acosta, J. (1608). *Historia Natural y Moral de las Indias*. Madrid: Alonso Martín, impresor.

Eguiarte, L. E., Jiménez Barrón, O. A., Aguirre-Planter, E., Scheinvar, E., Gámez, N., Gasca-Pineda, J., et al. (2021). Evolutionary ecology of *Agave*: distribution patterns, phylogeny, and coevolution (an homage to Howard S. Gentry). *Am. J. Bot.* 108, 216–235 doi: 10.1002/ajb2.1609

Heyduk, K., McKain, M. R., Lalani, F., and Leebens-Mack, J. (2016). Evolution of a CAM anatomy predates the origins of crassulacean acid metabolism in the Agavoideae (Asparagaceae). *Mol. Phylogenet. Evolut.* 105, 102–113. doi: 10.1016/j.ympev.2016.08.018

McKain, M. R., McNeal, J. R., Kellar, P. R., Eguiarte, L. E., Pires, J. C., and Leebens-Mack, J. (2016). Timing of rapid diversification and convergent origins of active pollination within Agavoideae (Asparagaceae). *Am. J. Bot.* 103, 1717–1729. doi: 10.3732/ajb.1600198

Pellmyr, O., and Leebens-Mack, J. (1999). Forty million years of mutualism: evidence for eocene origin of the yucca-yucca moth association. *Proc. Natl. Acad. Sic. U.S.A.* 96, 9178–9183. doi: 10.1073/pnas.96.16.9178

Rentsch, J. D., and Leebens-Mack, J. (2012). Homoploid hybrid origin of *Yucca gloriosa*: intersectional hybrid speciation in *Yucca* (Agavoideae, Asparagaceae). *Ecol. Evolut.* 2, 2213–2222. doi: 10.1002/ece3.328

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer CM declared a shared affiliation, with no collaboration, with one of the authors, LE, to the handling editor at the time of review.

frontiers in Plant Science

MINI REVIEW
published: 24 March 2020
doi: 10.3389/fpls.2020.00324

The Sweet Taste of Adapting to the Desert: Fructan Metabolism in *Agave* Species

*Arely V. Pérez-López and June Simpson**

Department of Genetic Engineering, Cinvestav Unidad Irapuato, Guanajuato, Mexico

Over 70% of *Agave* species, (159 of 206) are found in Mexico and are well adapted to survive under hot, arid conditions, often in marginal terrain, due to a unique combination of morphological and physiological attributes. In the pre-Columbian era agaves were also key to human adaptation to desert terrain. In contrast to other species such as cacti or resurrection plants, *Agaves* store carbohydrates in the form of fructan polymers rather than starch or sucrose, however, properties specific to fructans such as a strong hydration shell, the ability to be transported through phloem, variable composition throughout the *Agave* life-cycle and accumulation in succulent tissues and flowers suggest a potential for multiple functional roles. This mini-review summarizes current knowledge of molecular and biochemical aspects of fructan metabolism in *Agave* species.

Keywords: Agavaceae, agavins, signaling, metabolism, adaptation

OPEN ACCESS

Edited by:
Luis Enrique Eguiarte, National Autonomous University of Mexico, Mexico

Reviewed by:
Wim Van den Ende, KU Leuven, Belgium
Gretchen North, Occidental College, United States
Cristobal N. Aguilar, Universidad Autónoma de Coahuila, Mexico

***Correspondence:**
June Simpson
june.simpson@cinvestav.mx

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *17 December 2019*
Accepted: *05 March 2020*
Published: *24 March 2020*

Citation:
Pérez-López AV and Simpson J (2020) The Sweet Taste of Adapting to the Desert: Fructan Metabolism in Agave Species. Front. Plant Sci. 11:324. doi: 10.3389/fpls.2020.00324

INTRODUCTION

Fructan polymers, are synthesized by some bacteria and fungi and an estimated 15% of angiosperms including both monocotyledons and dicotyledons from different genera (Hendry, 1993; **Figure 1**). In plants, fructan polymers are described based on their structure and complexity (Versluys et al., 2018). Although neo type fructans have only been described in monocotyledons, no strong correlation exists between the type of fructan polymers and the genus or species in which they occur, supporting independent evolution of fructan metabolism.

Fructans are water soluble, flexible fructose based polymers synthesized from sucrose and accumulating in the vacuole. They can act as a long-term reserve carbohydrate in some plant species, alone or in combination with starch.

Fructans are an alternative to starch for long-term carbohydrate storage. Starch, composed of linear amylose or branched amylopectin glucose (hexose) polymers, accumulates in chloroplasts, whereas fructans produced by adding fructose monomers to sucrose are stored in vacuoles. Fructans are structurally flexible, highly soluble, accumulate to high levels, and have the ability to associate with cell membranes (Van den Ende, 2013). These properties are intrinsic to their roles in response to stress (Versluys et al., 2018) or developmental signals (Bolouri Moghaddam and Van den Ende, 2013). Fructans are exploited commercially as a replacement for sugar or fats, as fiber or prebiotics (Vijn and Smeekens, 1999) and have useful properties for drug delivery and cryoprotection (Audouy et al., 2011; Gupta et al., 2019).

Agaves evolved during the Miocene period and synthesis and storage of fructans was an important factor in adaptation to drier environments (Arakaki et al., 2011). *Agave* species range from the Canadian/United States border to the Northern region of South America (Gentry, 1982;

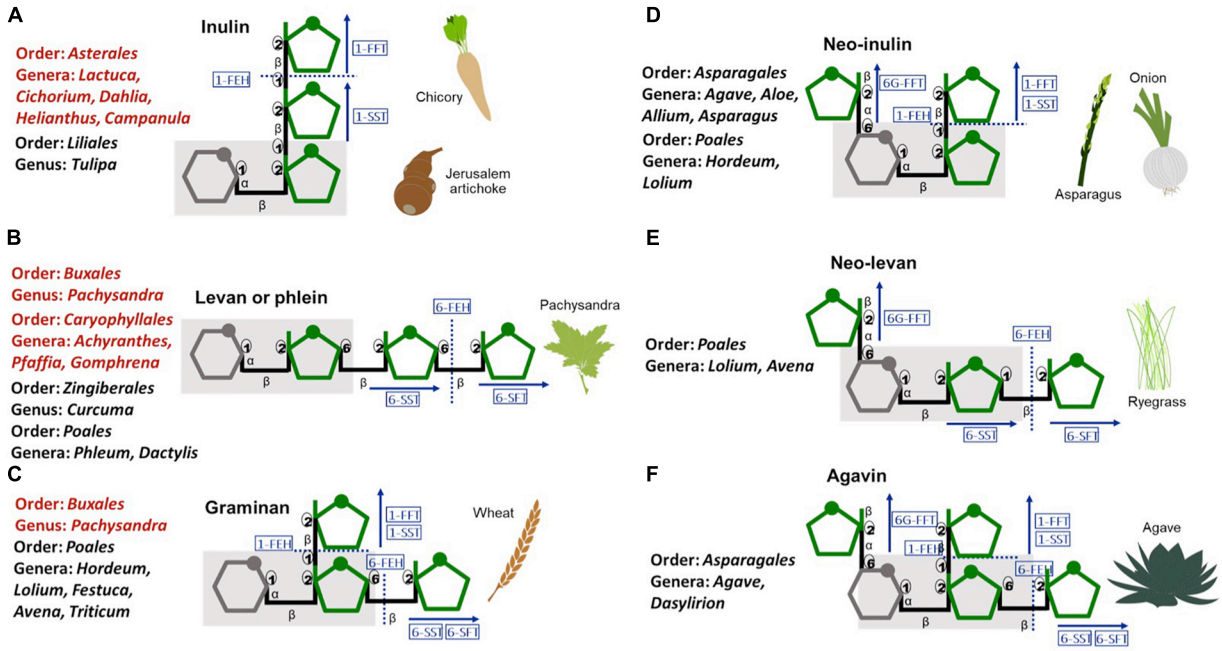


FIGURE 1 | Schematic representation of plant fructans, their structural diversity and the enzymes involved in their metabolism. **(A)** linear inulin and **(B)** levan, **(C)** branched graminan, **(D)** neo-inulin, **(E)** neo-levan, and **(F)** highly branched agavin. Gray-glucose, green-fructose, gray shadow-sucrose moiety. Blue rectangles-enzymes:1-SST-sucrose:sucrose1-fructosyltransferase, 1-FFT-fructan:fructan1-fructosyltransferase, 6-SFT-sucrose:fructan 6-fructosyltransferase, 6G-FFT-fructan:fructan 6Gfructosyltransferase, FEH-fructan exohydrolase. Red text-dicotyledons, Black text-monocotyledons.

Garcia, 2007). Whereas some species such as *A. deserti* or *A. americana* are adapted to wide temperature ranges others such as *A. tequilana* will not thrive at temperatures below −4°C or above 36°C (Nobel et al., 1998), demonstrating that tolerance mechanisms are complex.

Artificial selection of *Agaves* mainly took place in Mexico where 58% of species are endemic (Gentry, 1982; Garcia, 2007). Pre-Columbian cultures exploited these plants for food, fiber, construction and beverages and they were essential elements of nomadic life styles. *Agave* fructans provide the raw material for production of tequila and mescal, are being developed as components of treatments for diabetes and obesity (Franco-Robles et al., 2019) and as a resource for low-cost, carbon neutral production of bioenergy (Niechayev et al., 2019).

Agave FRUCTANS

The presence of fructans in *Agave* species was first recorded by Ekstrand and Johanson in 1888 as cited by Suzuki and Chatterton, 1993. In common with other members of the order Asparagales, *Agave* species synthesis inulin and neo series fructan polymers (**Figure 1**) and a new class of neofructans (subsequently known as "agavins") was first identified in *A. tequilana* (Mancilla-Margalli and Lopez, 2006). Agavins are the most complex plant fructans described to date, comprising neoseries type fructans elongated at all three possible linkages (**Figure 1**). The composition of the fructan pool in *A. tequilana* varies as plants age, with agavins increasing in abundance in relation to inulins (Mellado-Mojica and Lopez, 2012).

In *Agave* leaves starch accumulation is largely limited to stomatal guard cells with minimal accumulation in other leaf cells (Zavala-Garcia et al., 2018). The presence of oligofructans containing 3–5 fructan residues (3–5 degrees of polymerization, D.P.) in *Agave* leaves indicates that sucrose produced by photosynthesis is metabolized to produce fructans rather than starch (Wang and Nobel, 1998) have shown that these oligofructans can accumulate in vascular tissue and are transported through the phloem. Although the transport mechanism is unknown, it is most plausibly by polymer trapping (Zhang and Turgeon, 2018). However, the presence of fructans in the extracellular space (apoplast) (Raveh et al., 1998) and putative roles in defense, signaling and membrane protection indicate that an apoplastic mechanism cannot be ruled out.

Oligofructans and/or sucrose transported from leaves are either metabolized to starch that accumulates in the peripheral meristem region between the leaf base and the stem (Zavala-Garcia et al., 2018; **Figure 2**) or converted to complex fructans for long-term storage in the vacuoles of stem tissue (Mellado-Mojica et al., 2017).

FRUCTAN METABOLISM IN *Agave* SPECIES

To synthesize agavins and inulins 4 fructosyltransferase (FT) activities (1-SST, 1-FFT, 6-SFT, and 6G-FFT, **Figure 1**) are

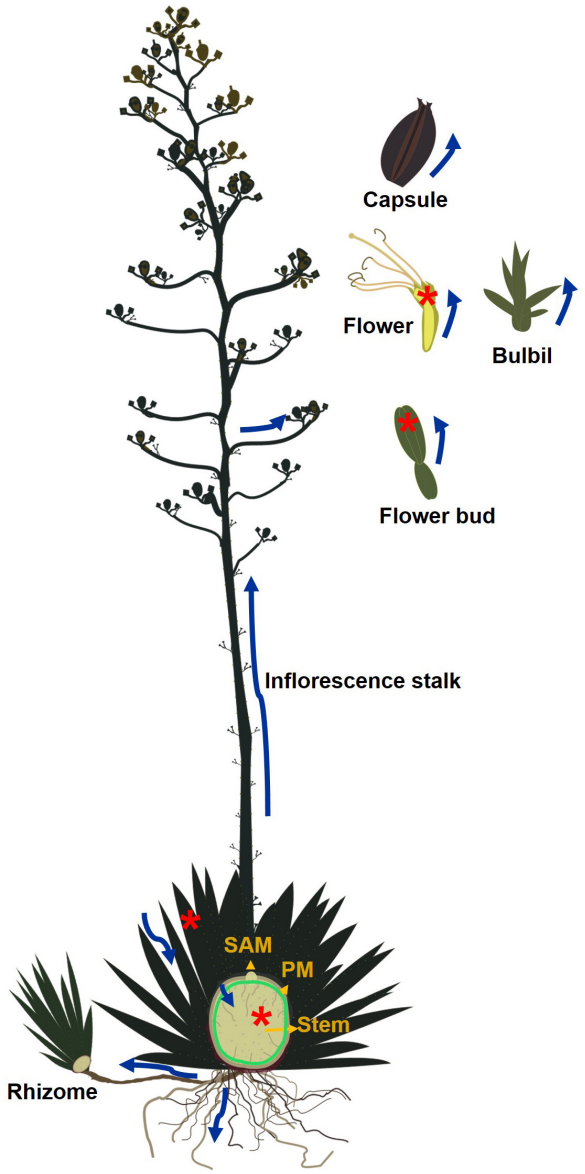


FIGURE 2 | Fructan and/or sucrose mobility throughout an *Agave* plant. Blue arrows indicate fructan movement, green circle indicates the starch layer and peripheral meristem. Red asterisks indicate tissues where agavins are synthesized. SAM-shoot apical meristem. PM-Peripheral meristem.

needed whereas degradation of fructans is carried out by fructan exohydrolases (FEH) that may be specific for β 1→2 or β 1→6 linkages or act on both. FT and FEH in common with vacuolar and cell wall invertases are members of Plant Glycoside Hydrolase Family 32 (PGHF32). The first *Agave* FT to be characterized was a 1-SST from *A. tequilana* (Avila-Fernandez et al., 2007) and by RNAseq 15 members of PGHF32 from *A. tequilana, A. deserti, and A. victoriae-reginae* were later identified (Avila De Dios et al., 2015). Sequence based predictions of enzyme activities have also been confirmed for some *A. tequilana* enzymes using the *P. pastoris* system (Cortes-Romero et al., 2012). cDNAs encoding 6-SFT or 1-FFT type enzymes have not yet been conclusively identified perhaps due to low or tissue specific expression. Alternatively, some *Agave* FT enzymes may have multiple activities as reported for a 6G-FFT from onion (Weyens et al., 2004). *In silico* modeling supports this hypothesis since (Huang et al., 2018) have shown that the predicted structure of an *A. tequilana* 6G-FFT differs from those identified in *A. deserti* and *A. sisalana.*

In silico expression patterns for genes encoding invertases and FEH across three different *Agave* species (*A. tequilana*, *A. striata* and A. *victoriae-reginae*) are consistent, whereas FT expression is highly variable (Avila De Dios et al., 2015). For example isoforms encoding 1-SST enzymes from *A. tequilana* and *A. striata* showed similar tissue specific patterns whereas those identified for *A. victoriae-reginae* varied widely and 6G-FFT encoding genes of *A. victoriae-reginae* and *A. striata* are strongly expressed in vegetative tissue in contrast to *A. tequilana*. Expression patterns for all three *Agave* species showed high levels of expression for both FT and FEH in floral tissue (Avila De Dios et al., 2015) suggesting that fructans are not only being degraded but are also being synthesized in these organs.

Transcriptome analysis of the vegetative to reproductive transition in *A. tequilana* revealed no differential expression for starch metabolism related genes (Zavala-Garcia et al., 2018) whereas fructan related genes are highly expressed in SAM tissue in comparison to leaf tissue. In particular a 6G-FFT isoform is specifically and strongly expressed at the initial stage of the reproductive phase (Avila De Dios et al., 2019).

BIOLOGICAL FUNCTIONS OF *Agave* FRUCTANS

In *Agave* species fructans provide a source of carbohydrates for the vegetative to reproductive transition. Inflorescences can grow at a rate of 4–10 cms per day to reach 10 m or more (Valenzuela, 2003) and produce thousands of flowers, capsules, and seeds (Escobar-Guzman et al., 2008). Under cultivation, inflorescences are removed to avoid depletion of fructan reserves. Delgado Sandoval et al., 2012, showed that as the reproductive stage initiates, development of photosynthetically active leaves is suppressed and the SAM differentiates. Genes encoding FEH and invertases increase their expression during bolting (Avila De Dios et al., 2019) and leaves and stems senesce indicating that carbohydrate reserves are being harnessed for flowering.

Fructan reserves are also exploited during asexual reproduction since suckers produced from rhizomes or bulbils produced on inflorescences (**Figure 2**) also benefit from carbohydrates stored in the mother plant and may not survive if detached too early (Szarek et al., 1996). To accomplish these functions fructans must be mobilized over significant distances. Active fructan metabolism in floral tissue suggests carbohydrate availability could be limited by the rate of turnover or transport. Fructans may act as precursors to nectar production in floral tissue since *A. palmeri* produces 74 mg of nectar/flower composed mainly of glucose and fructose (Riffell et al., 2008).

Alternatively fructans may be involved in generating osmolarity fluxes that lead to flower opening as described for *Campanula rapunculoides* (Vergauwen et al., 2000).

Agaves are perennial, monocarpic species with life cycles of 5 to over 50 years. They remain unresponsive to cues such as photoperiod or temperature, which induce flowering in annual or polycarpic species and probably respond to age-determined signals involving carbohydrate regulation. It could be speculated that accumulation of specific agavins produced by the 6G-FFT isoform described above may serve as age related molecular signals (Salinas et al., 2016) have also shown that neofructan levels increase in drought stressed *A. barbadensis* suggesting that neofructans play important functional roles.

The natural habitat of *Agave* species is in marginal desert terrain. Localization of fructans in hydrenchyma tissue in succulent *A. victoria-reginae* leaves (Singh et al., 2020) supports the evolution of fructan accumulation as an adaptation of Agavaceae to arid conditions. Consistent with these observations (Morales-Hernandez et al., 2019) showed that *Agave* fructans have a higher hydration shell in comparison to inulin and have predicted bioprotectant properties equivalent to trehalose. Suarez-Gonzalez et al., 2014 have shown that *A. tequilana* and *A. inaequidens* respond to cold and elicitors by increased FT expression and fructan production, consistent with roles in stress tolerance mechanisms.

DISCUSSION AND PERSPECTIVES

Biochemical analysis has shown the presence of fructans in all organs of different *Agave* species and the quantity and complexity of these polymers varies depending on specific tissue and plant age. The monocarpic, perennial life cycle, large genome size and lack of molecular tools for *Agave* species have hampered molecular/genetic analysis, however, transcriptome data has allowed preliminary classification, and characterization of cDNAs and enzymes involved in fructan metabolism. The failure to identify 2 key enzymes may reflect low or transient gene expression or multiple enzyme activities. Functional genetic analysis in *Agave* is inefficient but heterologous systems such as *A. thaliana* and *P. pastoris* are being exploited and development of a genome sequence will resolve questions regarding isoforms, gene structure and regulatory elements. Comparisons of fructan metabolism on an evolutionary level between related taxa such as *Yucca* and *Aloe spp.* and aspects of coevolution with nectar feeding pollinators will also be feasible. Sub-cellular localization of FT or FEH enzymes, detailed gene expression patterns and aspects of fructan mobility must also be addressed to provide insights to roles in signaling and stress tolerance.

Agaves represent an invaluable resource in relation to development of agricultural systems on marginal land with resilience to climate change. However, indiscriminate collection of wild plants leads to decimation of natural populations and their pollinators. Fructans are the basis for the commercial exploitation of *Agaves*, therefore, understanding *Agave* fructan metabolism, its multiple roles in the *Agave* life-cycle and in adaptation to different habitats will facilitate strategies for exploitation and conservation. The current challenge in Mexico is how to exploit *Agave* fructans under a profitable, sustainable and socially pertinent agricultural system.

AUTHOR CONTRIBUTIONS

JS developed the outline of the manuscript. JS and AP-L wrote the manuscript and designed the figures. AP-L prepared the figures.

FUNDING

Funding for this research was provided by CONACyT grant CB 2013-220339 and SEP-CINVESTAV Grant# 131. AP-L received postgraduate fellowship 566173 from CONACyT.

ACKNOWLEDGMENTS

We are grateful to Katia Gil Vega for technical assistance.

REFERENCES

Arakaki, M., Christin, P. A., Nyffeler, R., Lendel, A., Eggli, U., Ogburn, R. M., et al. (2011). Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8379–8384. doi: 10.1073/pnas.1100628108

Audouy, S. A., Van Der Schaaf, G., Hinrichs, W. L., Frijlink, H. W., Wilschut, J., and Huckriede, A. (2011). Development of a dried influenza whole inactivated virus vaccine for pulmonary immunization. *Vaccine* 29, 4345–4352. doi: 10.1016/j.vaccine.2011.04.029

Avila De Dios, E., Delaye, L., and Simpson, J. (2019). Transcriptome analysis of bolting in A. *tequilana* reveals roles for florigen. MADS, fructans and gibberellins. *BMC Genomics* 20:473. doi: 10.1186/s12864-019-5808-5809

Avila De Dios, E., Gomez Vargas, A. D., Damian Santos, M. L., and Simpson, J. (2015). New insights into plant glycoside hydrolase family 32 in Agave species. *Front. Plant Sci.* 6:594. doi: 10.3389/fpls.2015.00594

Avila-Fernandez, A., Olvera-Carranza, C., Rudino-Pinera, E., Cassab, G. I., Nieto-Sotelo, J., and Lopez-Munguia, A. (2007). Molecular characterization of sucrose: sucrose 1-fructosyltransferase (1-SST) from *Agave tequilana* Weber var. *azul*. *Plant Sci.* 173, 478–486. doi: 10.1016/j.plantsci.2007.07.009

Bolouri Moghaddam, M. R., and Van den Ende, W. (2013). Sugars, the clock and transition to flowering. *Front. Plant Sci.* 4:22. doi: 10.3389/fpls.2013.00022

Cortes-Romero, C., Martinez-Hernandez, A., Mellado-Mojica, E., Lopez, M. G., and Simpson, J. (2012). Molecular and functional characterization of novel fructosyltransferases and invertases from *Agave tequilana*. *PLoS One* 7:e35878. doi: 10.1371/journal.pone.0035878

Delgado Sandoval, S. C., Juárez, M. J. A., and Simpson, J. (2012). *Agave tequilana* MADS genes show novel expression patterns in meristems, developing bulbils and floral organs. *Sex. Plant Reprod.* 25, 11–26. doi: 10.1007/s00497-011-0176-x

Escobar-Guzman, R. E., Hernandez, F. Z., Vega, K. G., and Simpson, J. (2008). Seed production and gametophyte formation in *Agave tequilana* and *Agave americana*. *Botany-Botanique* 86, 1343–1353. doi: 10.1139/B08-099

Franco-Robles, E., Ramirez-Emiliano, J., and Lopez, M. G. (2019). Agave fructans and oligofructose decrease oxidative stress in brain regions involved in learning and memory of overweight mice. *Nat. Prod. Res.* 33, 1527–1530. doi: 10.1080/14786419.2017.1423297

Garcia, J. (2007). Los Agaves de México. *Red de Revistas Científicas de América Latina Y El Caribe, España Y Portugal* 87, 14–23.

Gentry, H. S. (1982). *Agaves of Continental North America*. Tucson: University of Arizona Press.

Gupta, N., Jangid, A. K., Pooja, D., and Kulhari, H. (2019). Inulin: a novel and stretchy polysaccharide tool for biomedical and nutritional applications. *Int. J. Biol. Macromol.* 132, 852–863. doi: 10.1016/j.ijbiomac.2019.03.188

Hendry, G. A. F. (1993). Evolutionary origins and natural functions of fructans - a climatological, biogeographic and mechanistic appraisal. *New Phytol.* 123, 3–14.

Huang, X., Wang, B., Xi, J., Zhang, Y., He, C., Zheng, J., et al. (2018). Transcriptome comparison reveals distinct selection patterns in domesticated and wild *Agave* species, the important CAM plants. *Int. J. Genomics* 2018:5716518. doi: 10.1155/2018/5716518

Mancilla-Margalli, N. A., and Lopez, M. G. (2006). Water-soluble carbohydrates and fructan structure patterns from *Agave* and *Dasylirion* species. *J. Agric. Food Chem.* 54, 7832–7839. doi: 10.1021/jf060354v

Mellado-Mojica, E., Gonzalez De La Vara, L. E., and Lopez, M. G. (2017). Fructan active enzymes (FAZY) activities and biosynthesis of fructooligosaccharides in the vacuoles of *Agave tequilana* weber blue variety plants of different age. *Planta* 245, 265–281. doi: 10.1007/s00425-016-2602-2607

Mellado-Mojica, E., and Lopez, M. G. (2012). Fructan metabolism in *A. tequilana* Weber Blue variety along its developmental cycle in the field. *J. Agric. Food Chem.* 60, 11704–11713. doi: 10.1021/jf303332n

Morales-Hernandez, J. A., Singh, A. K., Villanueva-Rodriguez, S. J., and Castro-Camus, E. (2019). Hydration shells of carbohydrate polymers studied by calorimetry and terahertz spectroscopy. *Food Chem.* 291, 94–100. doi: 10.1016/j.foodchem.2019.03.132

Niechayev, N. A., Jones, A. M., Rosenthal, D. M., and Davis, S. C. (2019). A model of environmental limitations on production of *Agave americana* L. grown as a biofuel crop in semi-arid regions. *J. Exp. Bot.* 70, 6549–6559. doi: 10.1093/jxb/ery383

Nobel, P. S., Castaneda, M., North, G., Pimienta-Barrios, E., and Ruiz, A. (1998). Temperature influences on leaf CO2 exchange, cell viability and cultivation range for *Agave tequilana*. *J. Arid Environ.* 39, 1–9. doi: 10.1006/Jare.1998.0374

Raveh, E., Wang, N., and Nobel, P. S. (1998). Gas exchange and metabolite fluctuations in green and yellow bands of variegated leaves of the monocotyledonous CAM species *Agave americana*. *Physiologia Plantarum* 103, 99–106. doi: 10.1034/j.1399-3054.1998.1030112.x

Riffell, J. A., Alarcon, R., Abrell, L., Davidowitz, G., Bronstein, J. L., and Hildebrand, J. G. (2008). Behavioral consequences of innate preferences and olfactory learning in hawkmoth-flower interactions. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3404–3409. doi: 10.1073/pnas.0709811105

Salinas, C., Handford, M., Pauly, M., Dupree, P., and Cardemil, L. (2016). Structural modifications of fructans in aloe barbadensis miller (Aloe Vera) grown under water stress. *PLoS One* 11:e0159819. doi: 10.1371/journal.pone.0159819

Singh, A. K., Pérez-López, A. V., Simpson, J., and Castro-Camus, E. (2020). Three-dimensional water mapping of succulent *Agave victoriae-reginae* leaves by terahertz imaging. *Sci. Rep.* 10:1404. doi: 10.1038/s41598-020-58277-z

Suarez-Gonzalez, E. M., Lopez, M. G., Delano-Frier, J. P., and Gomez-Leyva, J. F. (2014). Expression of the 1-SST and 1-FFT genes and consequent fructan accumulation in *Agave tequilana* and *A. inaequidens* is differentially induced by diverse (a)biotic-stress related elicitors. *J. Plant Physiol.* 171, 359–372. doi: 10.1016/j.jplph.2013.08.002

Suzuki, M., and Chatterton, N. J. (1993). *Science and Technology of Fructans*. Boca Raton: CRC Press.

Szarek, S. R., Driscoll, B., Shohet, C., and Priebe, S. (1996). Bulbil production in *Agave* (Agavaceeae) and related genera. *Southwest. Nat.* 41, 465–469.

Valenzuela, A. (2003). *El agave Tequilero: Su Cultivo e Industria de México*. México: Mundiprensa Editores.

Van den Ende, W. (2013). Multifunctional fructans and raffinose family oligosaccharides. *Front. Plant Sci.* 4:247. doi: 10.3389/fpls.2013.00247

Vergauwen, R., Van den Ende, W., and Van Laere, A. (2000). The role of fructan in flowering of *Campanula* rapunculoides. *J. Exp. Bot.* 51, 1261–1266. doi: 10.1093/Jexbot/51.348.1261

Versluys, M., Kirtel, O., Toksoy Oner, E., and Van den Ende, W. (2018). The fructan syndrome: evolutionary aspects and common themes among plants and microbes. *Plant Cell Environ.* 41, 16–38. doi: 10.1111/pce.13070

Vijn, I., and Smeekens, S. (1999). Fructan: more than a reserve carbohydrate? *Plant Physiol.* 120, 351–360.

Wang, N., and Nobel, P. S. (1998). Phloem transport of fructans in the crassulacean acid metabolism species *Agave deserti*. *Plant Physiol.* 116, 709–714.

Weyens, G., Ritsema, T., Van Dun, K., Meyer, D., Lommel, M., Lathouwers, J., et al. (2004). Production of tailor-made fructans in sugar beet by expression of onion fructosyltransferase genes. *Plant Biotechnol. J.* 2, 321–327. doi: 10.1111/j.1467-7652.2004.00074.x

Zavala-Garcia, L. E., Sanchez-Segura, L., Avila De Dios, E., Perez-Lopez, A., and Simpson, J. (2018). Starch accumulation is associated with active growth in *A. tequilana*. *Plant Physiol. Biochem.* 130, 623–632. doi: 10.1016/j.plaphy.2018.08.011

Zhang, C., and Turgeon, R. (2018). Mechanisms of phloem loading. *Curr. Opin. Plant Biol.* 43, 71–75. doi: 10.1016/j.pbi.2018.01.009

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

CORRECTION
published: 28 May 2020
doi: 10.3389/fpls.2020.00659

Corrigendum: The Sweet Taste of Adapting to the Desert: Fructan Metabolism in *Agave* Species

*Arely V. Pérez-López and June Simpson**

Department of Genetic Engineering, Cinvestav Unidad Irapuato, Guanajuato, Mexico

Keywords: Agavaceae, agavins, signaling, metabolism, adaptation

A Corrigendum on

The Sweet Taste of Adapting to the Desert: Fructan Metabolism in *Agave* Species
by Pérez-López, A. V., and Simpson, J. (2020). Front. Plant Sci. 11:324. doi: 10.3389/fpls.2020.00324

OPEN ACCESS

Approved by:
Frontiers Editorial Office, Frontiers Media SA, Switzerland

****Correspondence:***
June Simpson
june.simpson@cinvestav.mx

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *23 April 2020*
Accepted: *28 April 2020*
Published: *28 May 2020*

Citation:
Pérez-López AV and Simpson J (2020) Corrigendum: The Sweet Taste of Adapting to the Desert: Fructan Metabolism in Agave Species. Front. Plant Sci. 11:659. doi: 10.3389/fpls.2020.00659

In the original article, there was a mistake in the legend of **Figure 1**. The correct legend appears below.

"**Figure 1**. Schematic representation of plant fructans, their structural diversity and the enzymes involved in their metabolism. **(A)** linear inulin and **(B)** levan, **(C)** branched graminan, **(D)** neo-inulin, **(E)** neo-levan and **(F)** highly branched agavin. Gray-glucose, green-fructose, gray shadow-sucrose moiety. Blue rectangles-enzymes:1-SST-sucrose:sucrose1-fructosyltransferase, 1-FFT-fructan:fructan1-fructosyltransferase, 6-SFT-sucrose:fructan 6-fructosyltransferase, 6G-FFT-fructan:fructan 6G fructosyltransferase, FEH-fructan exohydrolase. Red text-dicotyledons, Black text-monocotyledons."

In addition, the error in the legend of **Figure 1** was carried over into the text as the word "pentose" should have been removed. A correction has been made to the **Introduction**, paragraph 3:

"Fructans are an alternative to starch for long-term carbohydrate storage. Starch, composed of linear amylose or branched amylopectin glucose (hexose) polymers, accumulates in chloroplasts, whereas fructans produced by adding fructose monomers to sucrose are stored in vacuoles. Fructans are structurally flexible, highly soluble, accumulate to high levels, and have the ability to associate with cell membranes (Van den Ende, 2013). These properties are intrinsic to their roles in response to stress (Versluys et al., 2018) or developmental signals (Bolouri Moghaddam and Van den Ende, 2013). Fructans are exploited commercially as a replacement for sugar or fats, as fiber or prebiotics (Vijn and Smeekens, 1999) and have useful properties for drug delivery and cryoprotection (Audouy et al., 2011; Gupta et al., 2019)."

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

REFERENCES

Audouy, S. A., Van Der Schaaf, G., Hinrichs, W. L., Frijlink, H. W., Wilschut, J., and Huckriede, A. (2011). Development of a dried influenza whole inactivated virus vaccine for pulmonary immunization. *Vaccine* 29, 4345–4352. doi: 10.1016/j.vaccine.2011.04.029

Bolouri Moghaddam, M. R., and Van den Ende, W. (2013). Sugars, the clock and transition to flowering. *Front. Plant Sci.* 4:22. doi: 10.3389/fpls.2013.00022

Gupta, N., Jangid, A. K., Pooja, D., and Kulhari, H. (2019). Inulin: a novel and stretchy polysaccharide tool for biomedical and nutritional applications. *Int. J. Biol. Macromol.* 132, 852–863. doi: 10.1016/j.ijbiomac.2019.03.188

Van den Ende, W. (2013). Multifunctional fructans and raffinose family oligosaccharides. *Front. Plant Sci.* 4:247. doi: 10.3389/fpls.2013.00247

Versluys, M., Kirtel, O., Toksoy Oner, E., and Van den Ende, W. (2018). The fructan syndrome: evolutionary aspects and common themes among plants and microbes. *Plant Cell Environ.* 41, 16–38. doi: 10.1111/pce.13070

Vijn, I., and Smeekens, S. (1999). Fructan: more than a reserve carbohydrate? *Plant Physiol.* 120, 351–360.

frontiers in Plant Science

METHODS
published: 09 April 2020
doi: 10.3389/fpls.2020.00384

Simple Whole-Mount Staining Protocol of F-Actin for Studies of the Female Gametophyte in Agavoideae and Other Crassinucellate Ovules

Alejandra G. González-Gutiérrez[1], Jorge Verdín[2] and Benjamín Rodríguez-Garay[1]**

[1] Unidad de Biotecnología Vegetal, CIATEJ, Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco, A.C., Zapopan, Mexico, [2] Unidad de Biotecnología Industrial, CIATEJ, Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco, A.C., Zapopan, Mexico

OPEN ACCESS

Edited by:
Karolina Heyduk, University of Hawaii, United States

Reviewed by:
Tomokazu Kawashima, University of Kentucky, United States
Jie Le, Institute of Botany, The Chinese Academy of Sciences, China

***Correspondence:**
Jorge Verdín
jverdin@ciatej.mx
Benjamín Rodríguez-Garay
agavero01@hotmail.com

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *11 December 2019*
Accepted: *17 March 2020*
Published: *09 April 2020*

Citation:
González-Gutiérrez AG, Verdín J and Rodríguez-Garay B (2020) Simple Whole-Mount Staining Protocol of F-Actin for Studies of the Female Gametophyte in Agavoideae and Other Crassinucellate Ovules. Front. Plant Sci. 11:384. doi: 10.3389/fpls.2020.00384

During plant sexual reproduction, F-actin takes part in the elongation of the pollen tube and the movement of sperm cells along with it. Moreover, F-actin is involved in the transport of sperm cells throughout the embryo sac when double fertilization occurs. Different techniques for analysis of F-actin in plant cells have been developed: from classical actin-immunolocalization in fixed tissues to genetically tagged actin with fluorescent proteins for live imaging of cells. Despite the implementation of live cell imaging tools, fixed plant tissue methods for cytoskeletal studies remain an essential tool for genetically intractable systems. Also, most of the work on live imaging of the cytoskeleton has been conducted on cells located on the plant's surface, such as epidermal cells, trichomes, and root hairs. In cells situated in the plant's interior, especially those from plant species with thicker organ systems, it is necessary to utilize conventional sectioning and permeabilization methods to allow the label access to the cytoskeleton. Studies about the role of F-actin cytoskeleton during double fertilization in plants with crassinucellate ovules (e.g., *Agave, Yucca, Polianthes, Prochnyantes,* and *Manfreda*) remain scarce due to the difficulties to access the female gametophyte. Here, we have developed a straightforward method for analysis of F-actin in the female gametophyte of different Agavoideae sub-family species. The procedure includes the fixation of whole ovules with formaldehyde, followed by membrane permeabilization with cold acetone, a prolonged staining step with rhodamine-phalloidin, and Hoechst 33342 as a counterstain and two final steps of dehydration of samples in increasing-concentration series of cold isopropanol and clarification of tissues with methyl salicylate. This technique allows the analysis of a large number of samples in a short period, cell positioning relative to neighbor cells is maintained, and, with the help of a confocal microscope, reconstruction of a single 3D image of F-actin structures into the embryo sac can be obtained.

Keywords: double fertilization, ovular apparatus, central cell nucleus, cytoskeleton, F-actin staining, fixed-tissue staining, confocal microscopy

INTRODUCTION

The actin cytoskeleton is a complex structure present in all eukaryotic cells (Povarova et al., 2012). In plants, actin is an important research target since it is involved in key cellular processes such as cell polarity, division plane determination, organogenesis, and intracellular signaling (Higaki et al., 2007). During reproduction of higher plants, actin filaments also play an important role; they are involved in pollen tube elongation (Vidali et al., 2001), vesicle and organelle transport (Drøbak et al., 2004; Cai and Cresti, 2009) and self-incompatibility responses (Roldán et al., 2012). Moreover, actin is involved in the female gametophyte development (Huang et al., 1999; Kawashima and Berger, 2015), in double fertilization (Huang and Sheridan, 1998; Kawashima et al., 2014) and the subsequent processes of endosperm (Świerczyńska and Bohdanowicz, 2003; Barranco-Guzmán et al., 2019) and embryo development in the seed (Kimata et al., 2016).

Due to its relevance, different techniques for visualization and analysis of F-actin have been developed: from classical actin-immunolocalization (Lazarides and Weber, 1974; Andersland et al., 1994) and phalloidin-based labeling (Wulf et al., 1979; Vandekerckhove et al., 1985) in fixed tissues, to genetically tagged actin with fluorescent proteins for live imaging (Kost et al., 1998). Among the latter, Lifeact, a short peptide consisting of the first 17 amino acids of *Saccharomyces cerevisiae* Abp14p, has revolutionized the study of F-actin physiology in eukaryotic cells (Sheahan et al., 2004; Era et al., 2009). Despite such progress, live-cell imaging is limited to genetically tractable systems. Also, most cytoskeleton's live imaging in plants has been conducted on surface cells (Blancaflor and Hasenstein, 2000) such as pollen tubes (Cheung et al., 2008), trichomes (Chang et al., 2019), and root hairs (McCurdy and Gunning, 1990; Colling and Wasteneys, 2005). However, the study of some biological processes, such as female gametophyte development and fertilization, requires the observation of the interior of the plant, which has specific technical challenges (Blancaflor and Hasenstein, 2000; Cheng, 2006).

The major technical challenge for female gametophyte imaging studies is the thickness of the sporogenous layers that cover it (Schneitz et al., 1995). These layers of nucellar tissue lead to poor quality observations or even access prevention of chemical and immunological dyes to their targets. The latter is particularly true for crassinucellate ovules (e.g., *Agave, Yucca, Polianthes, Prochnyantes,* and *Manfreda*) (Rudall, 1997), where one or more layers of hypodermic tissues are found between the meiocyte and the apex of the nuclei (Reddy, 2007; Endress, 2011). A first choice to solve this problem is two-photon confocal microscopy (Diaspro and Robello, 2000; Feijó and Moreno, 2004; Kimata et al., 2016) or, a cheaper alternative, microtome sectioning (Stelly et al., 1984). However, in microtomy techniques, the positioning of cells concerning neighbor cells are often lost, and the resulting sectioned planes are difficult to reconstruct in a single three-dimensional (3D) image (Haseloff, 2003; Barrell and Grossniklaus, 2005). Tissue permeabilization and clearing is an option to overcome those obstacles (Cheng, 2006). Under this strategy, thick tissue masses are made translucent through chemical treatments with substances with a high refractive index such as xylene, chloral hydrate, and methyl salicylate (Herr, 1993), reducing the problems of light scattering and spherical aberration, allowing high image resolution (Haseloff, 2003).

Here, we report an improved whole-mount technique to label F-actin in the female gametophyte of thick crassinucellate ovules of some genera of the Agavoideae sub-family and *Petunia hybrida* (Rezanejad, 2008) as an example of a different plant family. This technique combines classical tissue fixation, chemical staining, and a tissue clarification step that significantly improves image quality. This protocol allows the analysis of a large number of samples in a short period, cell positioning relative to neighboring cells is maintained, and 3D images of the cytoskeleton in deep tissues can be obtained.

MATERIALS AND EQUIPMENT

Reagents

PIPES, 1,4-piperazinediethanesulfonic acid (Sigma, Cat. No. P1851)

EGTA, Ethylene glycol-bis(2-aminoethylether)-N,N,N′,N′-tetraacetic acid (Sigma, Cat. No. E3889)

Magnesium chloride hexahydrate (Sigma, Cat. No. M2670)

Potassium hydroxide (Sigma, Cat. No. 221473)

37% Formaldehyde solution (Sigma, Cat. No. 252549)

Acetone (Sigma, Cat. No. 270725)

BSA, bovine serum albumin fraction V (Sigma, Cat. No. 10735078001)

Rhodamine-phalloidin (Molecular Probes, Cat. No. R415)

Hoechst 33258 pentahydrate (Molecular Probes, Cat. No. H21491)

2-Propanol (Sigma, Cat. No. 190764)

Methyl salicylate (Sigma, Cat. No. M6752)

Leica immersion oil type F (Leica, Cat. No. 11513859)

Latrunculin B from *Latruncula magnifica* (Sigma, Cat. No. L5288).

Materials

0.2–0.6 ml microcentrifuge tubes

Glass Pasteur pipettes and bulbs

Insulin needles and syringes

Glass slides, 75 mm × 25 mm (Corning, Cat. No. 2947)

Glass coverslips, 24 mm × 40 mm (Thermo Fisher Scientific, Cat. No. C7931)

Straight fine point tweezers.

Equipment

TCS SPE Confocal microscope (Leica Microsystems)

EZ4 HD Dissecting stereomicroscope (Leica Microsystems)

LAS X software® (Leica Microsystems).

SOLUTIONS RECIPES

ASB (Actin-stabilizing buffer) (Płachno and Świątek, 2012)

50 mM PIPES, 10 mM EGTA, and 1 mM $MgCl_2$, pH 6.8 adjusted with 10M KOH. It is important to previously dissolve EGTA and PIPES in a few drops of 10M KOH.

Fixative solution

3.7% formaldehyde in ASB. It is preferable to use the fixative solution just after preparation; however, it can be stored at 4°C for up to 5 days.

Blocking solution

1% BSA in ASB. BSA solution can be stored at 4°C.

Rhodamine-phalloidin stock solution

6.6 μM rhodamine-phalloidin in methanol. Store the solution at −20°C in darkness.

Hoechst 33258 pentahydrate stock solution

10 mg/ml Hoechst 33258 pentahydrate in distilled water. Prepare 2 ml aliquots, store them protected from light at −20°C.

Latrunculin B stock solution

20 μM latrunculin B in ethanol. Store at −20°C in darkness.

METHODS

Sample Collection

Flower buds of different sizes, mature flowers and immature fruits (collected at a distinct time after pollination) are collected and processed as follows to visualize the F-actin cytoskeleton at different stages of the female gametophyte development and early embryogenesis (an overview of the protocol described below is shown in **Figure 1**).

Dissection of Ovules

The dissection of ovules and immature seeds is performed with the help of straight fine-point tweezers and an insulin needle under the stereoscope.

Ovules Collection and Fixation

Ovules of the same ovary are collected in a 0.2–0.6 ml microtube containing ASB (N.B.1) at 25°C (room temperature). Once enough ovules have been collected (keep in mind that a fraction of ovules is lost during the staining process), they are incubated in ASB at 55°C for 5 min (N.B.2). Afterward, ovules are fixed with a fixative solution for 7–10 min at 25°C. Small-sized ovules require less fixation time than larger ones (e.g., *Agave* ovules are fixed for 10 min, while *Petunia* ovules are fixed for 7 min). After fixation, rinse ovules twice with ASB. If needed, previously fixed and washed ovules can be stored up to five days at 4°C protected from light. Afterward, continue the technique in section "Cuticle Solubilization and Membrane Permeabilization."

N.B.1 unlike similar protocols, we have used ASB instead of MBS. EGTA contained in ASB binds Ca^{2+} ions, which prevents actin filaments severing (Yin et al., 1981; Hepler, 2016).

N.B.2 pretreatment at 55°C allows more efficient fixative penetration. For Asparagales species, warm buffer incubation does not affect the structure of neither the ovule nor actin filaments.

Cuticle Solubilization and Membrane Permeabilization

After completing the fixation step, quickly rinse twice ovules with acetone at −20°C, and afterward keep them in fresh cold acetone for 5 min. Finally, wash ovules three times with ASB or until it remains crystalline.

Blocking and Staining

Pre-incubate ovules in blocking solution (1% BSA in ASB) for 20 min at room temperature. Then, stain ovules overnight at 4°C with 0.33 μM rhodamine-phalloidin for labeling F-actin and 3 μg/ml Hoechst 33258 to counterstain cell nuclei (diluted in blocking solution).

Dehydration

After the staining period (N.B.3), dehydrate ovules in isopropanol (N.B.4) increasing-concentration solutions (75, 85, 95, and 100%) for 7 min each at 4°C, and finally, in 100% isopropanol for 10–12 min also at 4°C. All isopropanol solutions must be continuously renewed, and samples should be gently shaken to homogenize the exposure of tissues to isopropanol.

N.B.3 it is not necessary to wash the stain with a buffer since the next dehydration series work also as a washing step.

N.B.4 in this protocol we use isopropanol instead of ethanol or methanol to dehydrate samples since dehydration with isopropanol is faster and produces better quality images.

Clarification

For tissue clarification, remove isopropanol and add a 1:1 methyl salicylate-isopropanol solution for 30–60 min. In the beginning, ovules will remain on top of the solution, but eventually, they will sink to the bottom of the microcentrifuge tube. Incubation time in methyl salicylate-isopropanol concludes when all ovules precipitate. Before observation, ovules are incubated in 100% methyl salicylate for at least 30 min. During this time, ovules get completely clear. Ovules can be kept in this solution in darkness at 4°C for about a week.

Mounting and Microscopy

Mount treated ovules directly on glass slides with 100% methyl salicylate. Observe the samples under the confocal microscope using a 532 nm laser for rhodamine-phalloidin (ex/em = 540/556 nm) and a 405 nm laser for Hoechst 33258 observation (ex/em = 352/461 nm). Analyze images with LAS X® software or any other appropriate software.

Control of the Specificity of Rhodamine-Phalloidin F-Actin Staining

To confirm the specificity of rhodamine-phalloidin F-actin staining, ovules of *Agave* sp. were treated with latrunculin-B, which prevents G-actin polymerization (Spector et al., 1983). Inhibition assays were conducted following the protocol of Yuan et al. (2002) with some modifications; in short, dissected ovules of *Agave* sp. were collected in microcentrifuge tubes containing culture medium (5 mM HEPES, 1 mM KCl, 1 mM $MgCl_2$, 0.1 mM $CaCl_2$, 3% w/v sucrose). Once enough ovules were

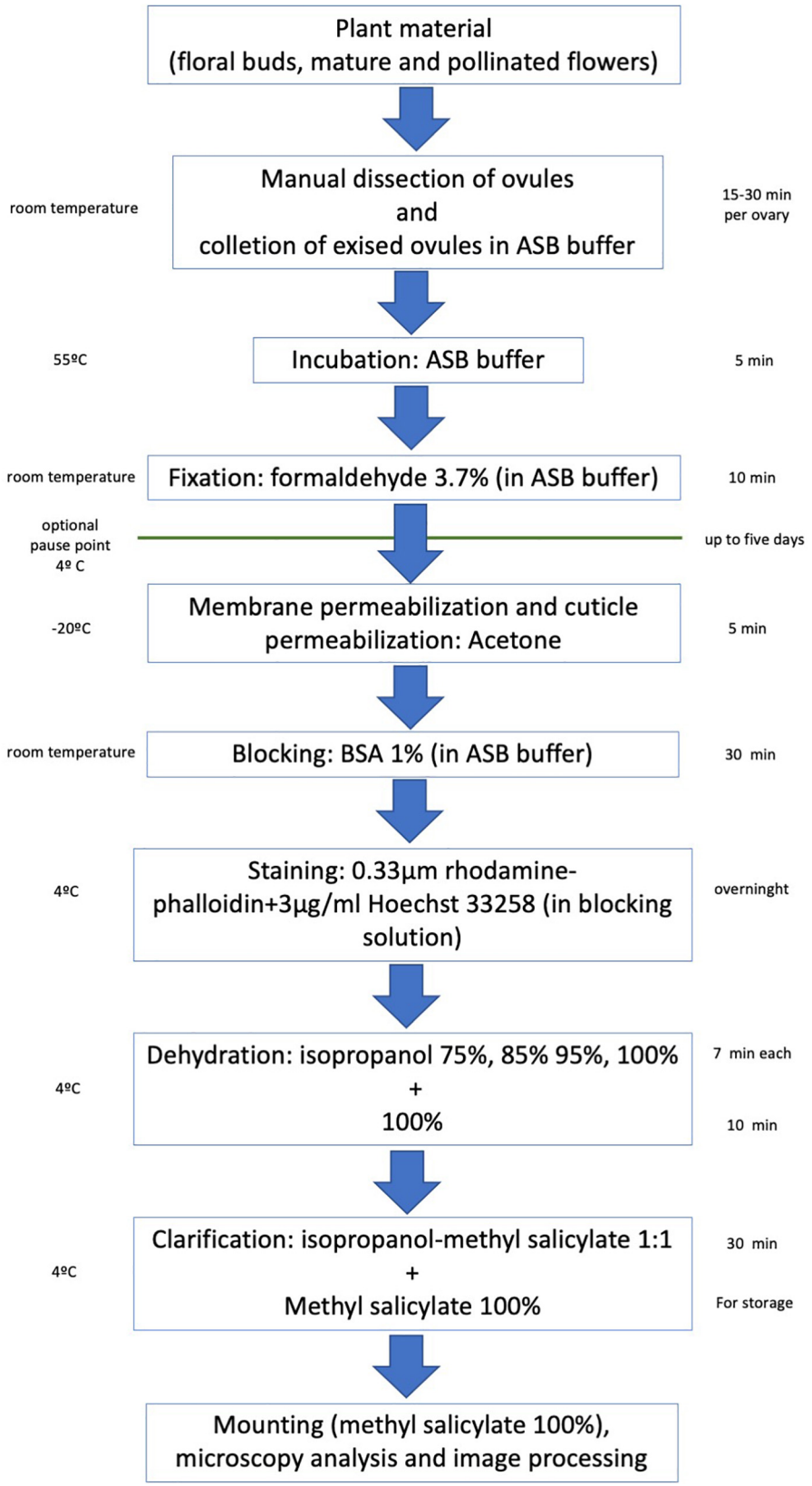


FIGURE 1 | Overview of the main steps of the rhodamine-phalloidin staining and methyl salicylate clarification of crassinucellate ovules.

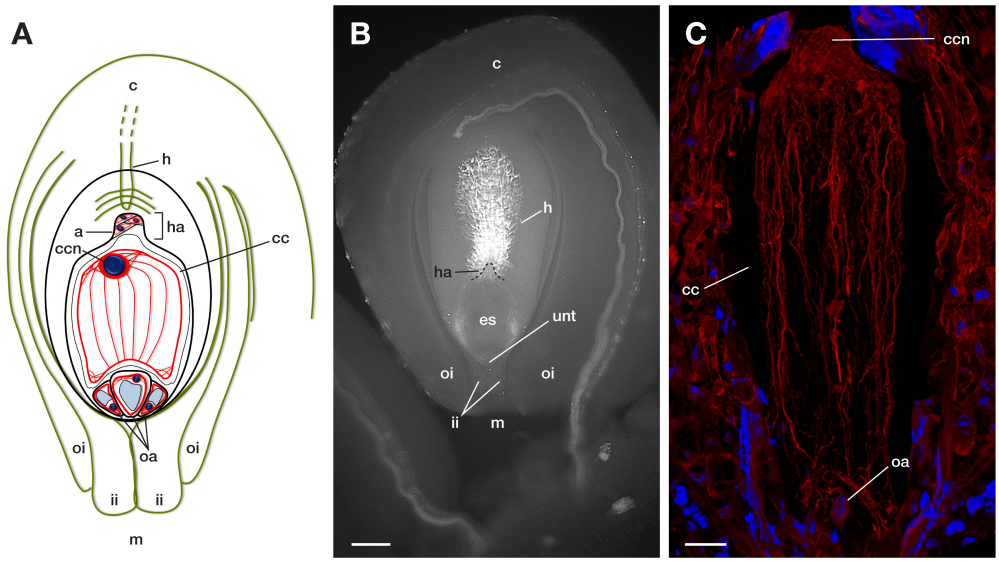


FIGURE 2 | Main components of a mature crassinucellate ovule of the Agavoideae sub-family. **(A)** Schematic representation of an anatropous, bitegmic, crassinucellate ovule of *Agave*. **(B)** Representative microphotography and **(C)** F-actin cytoskeleton of a mature embryo sac of *Agave.* c, chalaza; m, micropyle; oi, outer integument; ii, inner integument; unt, uniseriate nucellar tissue; es, embryo sac; h, hypostase; ha, haustorium; oa, ovular apparatus; a, antipodal cells; cc, central cell; ccn, central cell nucleus. The blue color in A and C indicate nuclei of cells (Hoechst 33258) and red color (phalloidin) represents F-actin filaments. Bar in **(B)**, 90 μm, in **(C)**, 40 μm.

collected, they were incubated in culture medium with (20 nM, final concentration) or without (control) latrunculin B for 4 h, at 22°C. After completing this incubation, ovules were quickly washed three times with culture medium and, finally, fixed and stain-cleared as described above.

RESULTS

The protocol described here can be performed in 48 h, which includes an overnight staining incubation (**Figure 1**). It enables us to perform microscopy observations of whole embryo sacs and determine the 3D allocation of the F-actin cytoskeleton inside them (**Figures 2**, **3A,C,E**, **4**, **5**). Up to 25–30 μm thick ovules could be observed without microtome sectioning (**Supplementary Movie 1**).

Fixation time should be optimized for each plant species and the sample developmental stage. In general, smaller ovules and ovules in early development stages need shorter fixation times. Cuticle solubilization and dehydration steps are also critical; they require constant solutions renewal and gentle hand-shaking to homogenize components. Rhodamine-phalloidin and Hoechst 33258 fluorophores maintain their fluorescence stable up to 10 days on samples treated with this stain-clearing technique when they are stored at 4°C. Moreover, co-staining with Hoechst 33258 provides information on the spatial position of nuclei within the cell and its relationship with actin filaments (**Figures 4C,F,I,L**, **5C,F,I,L**).

Latrunculin B inhibition assays were performed on *Agave* embryo sacs (**Figure 3**) to confirm the specificity of rhodamine-phalloidin F-actin staining. In the presence of latrunculin B, actin filaments appeared fragmented or completely disappeared (**Figures 3B,D,F**) while in the control treatment, intact actin filaments were observed (**Figures 3A,C,E**).

The methyl salicylate clearing step is critical for the protocol's success since it allows us to get over the physical barriers that usually impede imaging of the whole embryo sac. Sample observation needs a minimum of 30 min incubation in methyl salicylate after mounting; longer incubation times usually improve image quality.

Following this protocol, we managed to visualize the F-actin cytoskeleton in the female gametophyte of different genera of the Agavoideae sub-family (*Agave, Manfreda, Yucca,* and *Prochnyanthes*) (**Figures 4**, **5A–I**) and other non-related species with crassinucellate ovules such as *P. hybrida* (**Figures 5J–L**).

This protocol is useful for the analysis of different female developmental stages of crassinucellate ovules, from the differentiation of the megaspore mother cell, the megasporogenesis, megagametogenesis, and the double fertilization, to early stages of embryo and endosperm development (**Figures 4**, **5**). The rhodamine-phalloidin staining followed by methyl salicylate clarification allows identifying dense actin cables as well as thin actin filaments (**Figures 4B,E,H,K**, **5B,E,H,K**).

In the mature embryo sac, the cytoskeleton of each cell type (central cell, synergid, antipodal cells, and egg cell) located beneath the membrane could be observed; similarly, the F-actin coat around the nuclei of the cells could be appreciated with great detail (**Figures 4**, **5**). F-actin strands that run parallel along the chalazal-micropylar axis of the large central cell vacuole were detected without spherical aberration. This technique allowed the transmission of the microscope laser through the thicker tissues

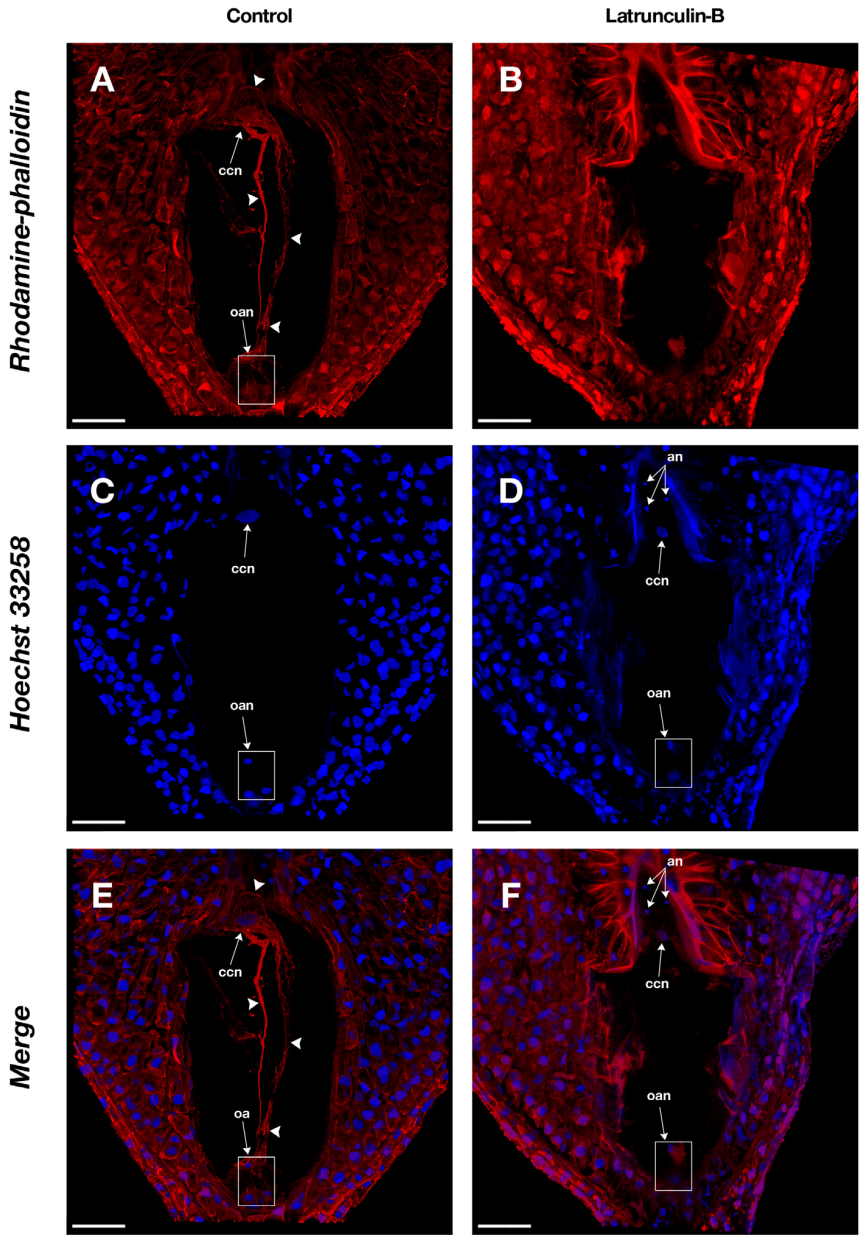


FIGURE 3 | Actin filaments in the embryo sac of *Agave* sp. **(A,C,E)** and the effect of latrunculin B **(B,D,F)**. **(A,B)** show rhodamine-phalloidin staining of the control and latrunculin B treated cells, respectively. **(C,D)** show Hoechst 33258 counterstaining. **(E,F)** are the merge of rhodamine-phalloidin and Hoechst 33258 channels. Arrowheads indicate F-actin filaments and cables that form part of the cytoskeleton of each cell type in the female gametophyte. an, antipodal nuclei; oan, ovular apparatus nuclei; oa, ovular apparatus; ccn, central cell nucleus. Bars, 40 μm.

that are found in the immature seeds; thus, actin cables that connect free nuclei of the endosperm in the embryo sac could be registered (**Figures 5A–C, G–I**).

DISCUSSION

Despite the great progress of fluorescent protein-tagging of cellular targets for live-cell imaging, phalloidin conjugated with any fluorochrome remains the gold standard for actin filament visualization (Melak et al., 2017). Immunofluorescence- and phalloidin-based techniques are useful for the structural analysis of cytoskeleton, especially in fixed cells, and, even when they share some critical steps like fixation and permeabilization (Blancaflor and Hasenstein, 2000), each one presents its advantages and drawbacks.

Some researchers have shown that phalloidin, fluorescent proteins and antibodies give different imaging results (Tang et al., 1989; Le et al., 2003; Thomas et al., 2009; Zhang et al., 2018; Flores et al., 2019). They claim that phalloidin and fluorescent

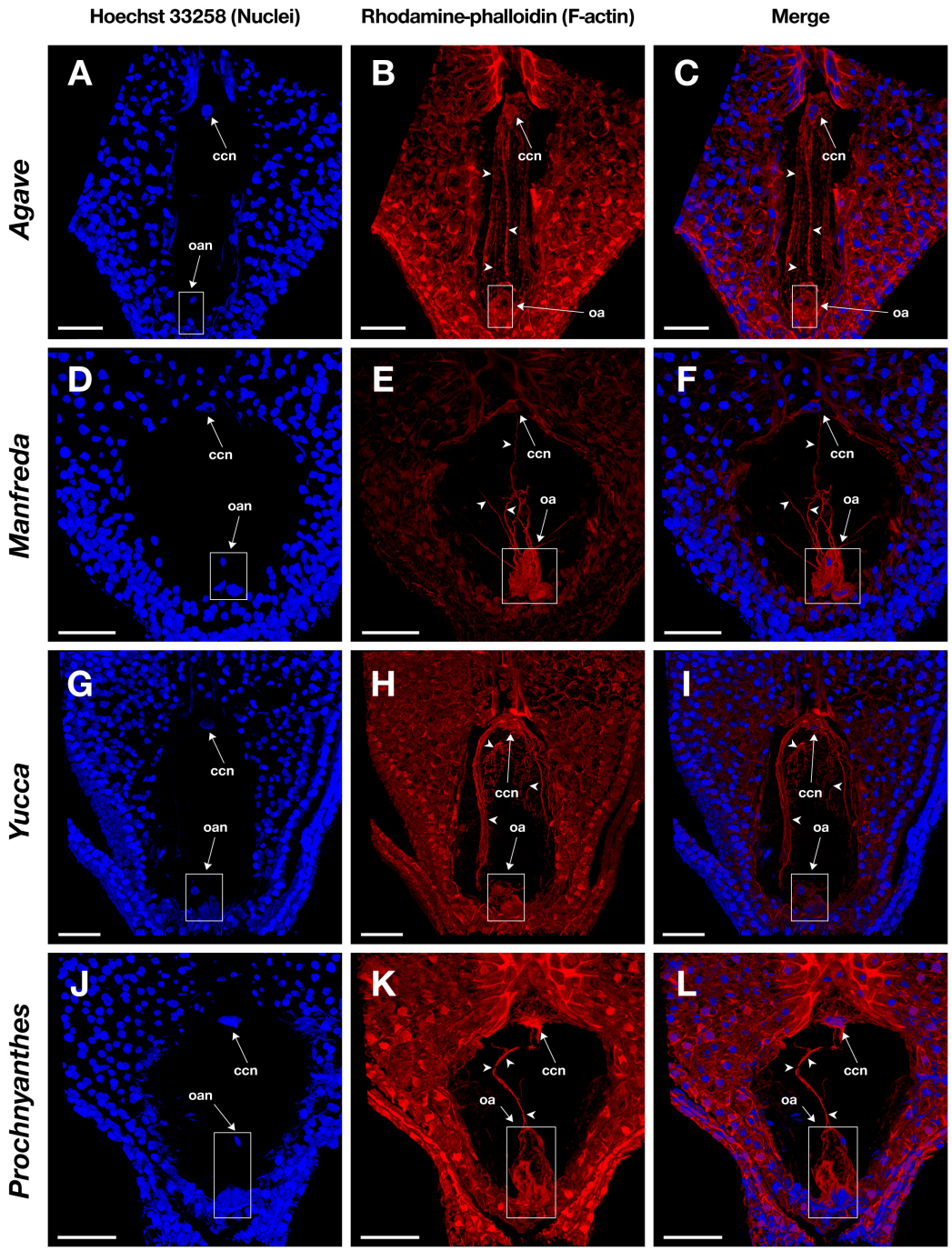


FIGURE 4 | Mature embryo sacs of four different species belonging to the Agavoideae sub-family stained with Hoechst 33258 (left column) and rhodamine-phalloidin (central column), and clarified with methyl salicylate. **(A–C)** *Agave tequilana*; **(D–F)** *Manfreda elongata*; **(G–I)** *Yucca* sp.; and, **(J–L)** *Prochnyanthes* sp. Arrowheads indicate F-actin filaments and cables that form part of the cytoskeleton of each cell type in the female gametophyte. ccn, central cell nucleus; oan, ovular apparatus nuclei; oa, ovular apparatus. Bars, 40 μm.

proteins may induce actin bundles artifacts (Le et al., 2003); nevertheless, others suggest those actin forms are biologically active and produced by specific actin associated proteins (Bartles, 2000; Thomas et al., 2009; Zhang et al., 2018). In this work, we did not observe actin bundles in surface cells stained with phalloidin (data not shown), which were subjected to exactly the same staining conditions as embryo sacs, where actin bundles are abundant (**Figures 3A**, **4B,E,H**). The latter suggests that bundles are produced in specific cellular contexts and they are not phalloidin induced artifacts.

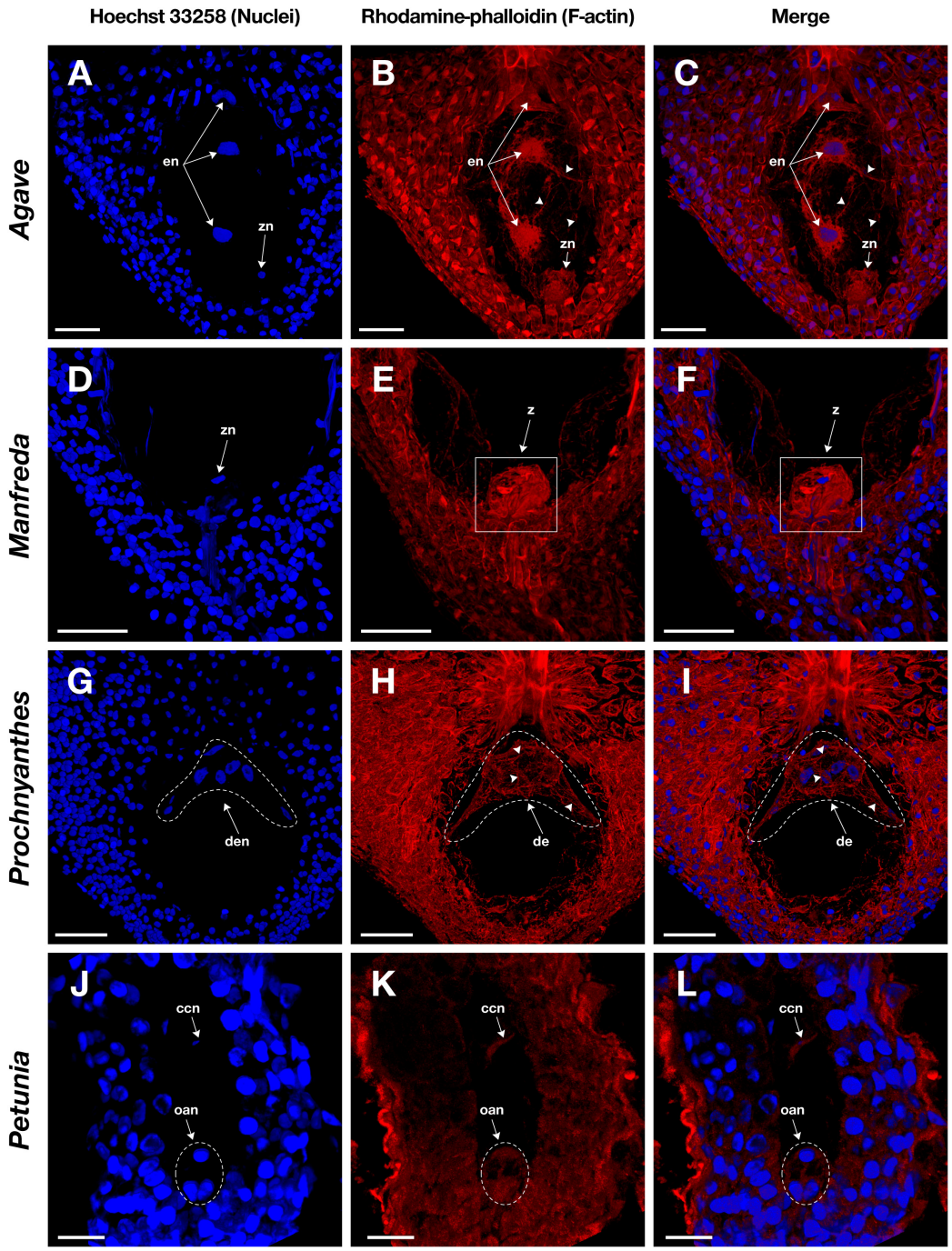


FIGURE 5 | Female gametophyte of four different species with crassinucellate ovules stained with Hoechst 33258 (left column) and rhodamine-phalloidin (central column), and clarified with methyl salicylate. **(A–C)** Endosperm nuclei of an immature seed of *Agave tequilana*; **(D–F)** zygote of *Manfreda elongata*; **(G–I)** chalazal endosperm chamber of *Prochnyanthes* sp.; and **(J–L)** mature embryo sac of *Petunia hybrida*. Arrowheads indicate F-actin filaments and cables that form part of the cytoskeleton inside the embryo sac. en, endosperm nuclei; z, zygote; zn, zygote nucleus; de, developing endosperm; den, developing endosperm nuclei; ccn, central cell nucleus; oan, ovular apparatus nuclei; oa, ovular apparatus. Bars, 40 μm.

Perhaps one of the main perks of immunolabeling is the possibility of applying two or more antibodies on the same sample to co-label several proteins (Shimamura, 2015) (e.g., actin and microtubules). Despite the later, antibodies are generally large; therefore, a proper fixation, membrane permeabilization, and cell wall digestion result critical for the successful diffusion of antibodies, especially into deeper cell layers (Pasternak et al., 2015), as is the case with the female gametophyte. On the other

hand, the relatively small size of phalloidin derivatives might be helpful in its permeation through the cell wall and membrane of plant cells. Thus, in the present protocol, permeabilization with detergents like DMSO or Triton X100, and degradation of the cell wall with enzymes were not necessary, which contribute to shorten the duration of the technique.

In this improved method, incubation, fixation, permeabilization, and clarification were performed in ASB to stabilize F-actin, after the successful experience of Płachno and Świątek (2012). They managed to stain the actin filaments in extra-ovular embryo sacs of *Utricularia nelumbifolia* (Płachno and Świątek, 2012). ASB contains EGTA, which binds Ca^{2+} ions that prevents actin filaments severing (Yin et al., 1981; Hepler, 2016).

Due to the intrinsic features of plant cells and tissues -like cell walls, vacuoles, and cuticle layers- most of the imaging work has been conducted on plant's surface cells. If inside cells need to be observed, microtome sectioning used to be the approach. This technique is time-consuming and provides images in only two dimensions (Haseloff, 2003). Here, by using methyl salicylate to clarify tissues, we accomplished the imaging of the complete F-actin cytoskeleton within the embryo sac (**Figures 3–5**), a highly vacuolated structure that is located inside the ovule and surrounded by one or more layers of nucellar tissue (**Figure 2**).

Methyl salicylate has been successfully used in the structural analysis of the female gametophyte development of *Solanum* (Stelly et al., 1984) and *Polianthes* (González-Gutiérrez and Rodríguez-Garay, 2016). Nevertheless, according to Richardson and Lichtman (2015), organic solvent-based clearing methods, which remove water from the cell, affect the capacity of fluorophores to maintain its emission. Despite those remarks, in our studies, the employment of methyl salicylate to clarify tissues did not interfere in the detection and quality of the fluorescent label.

Overall, this improved rhodamine-phalloidin staining followed by methyl salicylate clearing of whole ovules represents an option for the study of F-actin cytoskeleton in plant species where tagging with fluorescent proteins is not feasible. This approach is especially useful for imaging thick crassinucellate ovules, which till this report has not been successfully labeled and imaged (Escobar-Guzmán et al., 2015). Moreover, this technique could be useful as a first, easy, and rapid approach to visualize the actin cytoskeleton of the female gametophyte of different plant species.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

AG-G carried out the microscope analyses, the acquisition, and interpretation of images and drafted the manuscript. JV helped with interpretation of data and drafted the manuscript. BR-G conceived and coordinated the study, carried out the analysis and interpretation of the data, and drafted the manuscript. All authors read and approved the final manuscript.

FUNDING

This research was carried out with the support of the Frontiers of Science Program of the Mexican National Council of Science and Technology (CONACyT-Mexico), project 544, and Laboratorio Nacional PlanTECC, CONACyT project 293362.

ACKNOWLEDGMENTS

We thank H. Rodríguez-Julián and J. Aldana-Padilla for their assistance with the artwork. *Petunia hybrida* microphotographs were provided by L.A. Díaz-Godínez. AG-G is currently a graduate student at Doctorado en Ciencias en Innovación Biotecnológica, CIATEJ, Guadalajara, Jalisco, Mexico and receives a scholarship (462331) from CONACyT-Mexico.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00384/full#supplementary-material

MOVIE S1 | Animation of a 3D-projection of the micropylar region of an embryo sac of Agave spp stained with rhodamine-phalloidin. Note that F-actin bundles are conspicuously stained.

REFERENCES

Andersland, J. M., Fisher, D. D., Wymer, C. L., Cyr, R. J., and Parthasarathy, M. V. (1994). Characterization of a monoclonal antibody prepared against plant actin. *Cell Motil. Cytoskel.* 29, 339–344. doi: 10.1002/cm.970290406

Barranco-Guzmán, A. M., González-Gutiérrez, A. G., Rout, N. P., Verdín, J., and Rodríguez-Garay, B. (2019). Cytosolic calcium localization and dynamics during early endosperm development in the genus Agave (*Asparagales. Asparagaceae*). *Protoplasma* 256, 1079–1092. doi: 10.1007/s00709-019-01366-2

Barrell, P. J., and Grossniklaus, U. (2005). Confocal microscopy of whole ovules for analysis of reproductive development: the elongate1 mutant affects meiosis II. *Plant J.* 43, 309–320. doi: 10.1111/j.1365-313X.2005.02456.x

Bartles, J. R. (2000). Parallel actin bundles and their multiple actin-bundling proteins. *Curr. Opin. Cell Biol.* 12, 72–78. doi: 10.1016/s0955-0674(99)00059-9

Blancaflor, E. B., and Hasenstein, K. (2000). "Methods for detection and identification of F-actin in fixed and permeabilized plant tissues," in *Actin: A Dynamic Framework for Multiple Plant Cell Functions*, eds C. J. Staiger, F. Baluška, D. Volkmann, and P. W. Barlow (Dordrecht: Springer), 601–618. doi: 10.1007/978-94-015-9460-8_34

Cai, G., and Cresti, M. (2009). Organelle motility in the pollen tube: a tale of 20 years. *J. Exp. Bot.* 60, 495–508. doi: 10.1093/jxb/ern321

Chang, J., Xu, Z., Li, M., Yang, M., Qin, H., Yang, J., et al. (2019). Spatiotemporal cytoskeleton organizations determine morphogenesis of multicellular trichomes in tomato. *PLoS Genet.* 15:e1008438. doi: 10.1371/journal.pgen.1008438

Cheng, P. C. (2006). "Interaction of light with botanical specimens," in *Handbook of Biological Confocal Microscopy*, ed. J. Pawley (Boston, MA: Springer), 414–441. doi: 10.1007/978-0-387-45524-2_21

Cheung, A. Y., Duan, Q. H., Costa, S. S., de Graaf, B. H. J., Di Stilio, V. S., Feijo, J., et al. (2008). The dynamic pollen tube cytoskeleton: live-cell studies using actin-binding and microtubule-binding reporter proteins. *Mol Plant.* 1, 686–702. doi: 10.1093/mp/ssn026

Colling, D. A., and Wasteneys, G. O. (2005). Actin microfilament and microtubule distribution patterns in the expanding root of *Arabidopsis thaliana*. *Can. J. Bot.* 83, 579–590. doi: 10.1139/B05-032

Diaspro, A., and Robello, M. (2000). Two-photon excitation of fluorescence for three-dimensional optical imaging of biological structures. *J. Photochem. Photobio. B.* 55, 1–8. doi: 10.1016/S1011-1344(00)00028-2

Drøbak, B. K., Franklin-Tong, V. E., and Staiger, C. J. (2004). The role of the actin cytoskeleton in plant cell signaling. *New Phytol.* 163, 13–30. doi: 10.1111/j.1469-8137.2004.01076.x

Endress, P. K. (2011). Evolutionary diversification of the flowers in angiosperms. *Am. J. Bot.* 98, 370–396. doi: 10.3732/ajb.1000299

Era, A., Tominaga, M., Ebine, K., Awai, C., Saito, C., Ishizaki, K., et al. (2009). Application of lifeact reveals F-actin dynamics in *Arabidopsis thaliana* and the liverwort. *Marchantia polymorpha*. *Plant Cell Physiol.* 50, 1041–1048. doi: 10.1093/pcp/pcp055

Escobar-Guzmán, R. E., Rodríguez-Leal, D., Vielle-Calzada, J. P., and Ronceret, A. (2015). Whole-mount immunolocalization to study female meiosis in *Arabidopsis*. *Nat. Protoc.* 10, 1535–1542. doi: 10.1038/nprot.2015.098

Feijó, J. A., and Moreno, N. (2004). Imaging plant cells by two-photon excitation. *Protoplasma* 223, 1–32. doi: 10.1007/s00709-003-0026-2

Flores, L. R., Keeling, M. C., Zhang, X., Sliogeryte, K., and Gavara, N. (2019). Lifeact-TagGFP2 alters F-actin organization, cellular morphology and biophysical behaviour. *Sci. Rep.* 9, 1–13. doi: 10.1038/s41598-019-40092-w

González-Gutiérrez, A. G., and Rodríguez-Garay, B. (2016). Embryogenesis in *Polianthes* tuberosa L var. simple: from megasporogenesis to early embryo development. *SpringerPlus* 5:1804.

Haseloff, J. (2003).). Old botanical techniques for new microscopes. *Biotechniques* 34, 1174–1182. doi: 10.2144/03346bi01

Hepler, P. K. (2016). The cytoskeleton and its regulation by calcium and protons. *Plant Physiol.* 170, 3–22. doi: 10.1104/pp.15.01506

Herr, J. M. Jr. (1993). "Clearing techniques for the study of vascular plant tissues in whole structures and thick sections," in *Tested Studies for Laboratory Teaching, Proceedings of the 5th Workshop/Conference of the Association for Biology Laboratory Education (ABLE)*, eds C. I. A. /I. Goldman, P. L. Hauta, M. A. O'Donnell, S. E. Andrews, and R. van der Heiden (Columbia, South Carolina), 63–84.

Higaki, T., Sano, T., and Hasezawa, S. (2007). Actin microfilament dynamics and actin side-binding proteins in plants. *Curr. Opin. Plant Biol.* 10, 549–556. doi: 10.1016/j.pbi.2007.08.012

Huang, B. Q., Fu, Y., Zee, S. Y., and Hepler, P. K. (1999). Three-dimensional organization and dynamic changes of the actin cytoskeleton in embryo sacs of *Zea mays* and *Torenia fournieri*. *Protoplasma* 209, 105–119. doi: 10.1007/BF01415706

Huang, B. Q., and Sheridan, W. F. (1998). Actin coronas in normal and indeterminate gametophyte1 embryo sacs of maize. *Sex. Plant Reprod.* 11, 257–264. doi: 10.1007/s004970050151

Kawashima, T., and Berger, F. (2015). The central cell nuclear position at the micropylar end is maintained by the balance of F-actin dynamics, but dispensable for karyogamy in *Arabidopsis*. *Plant Reprod.* 28, 103–110. doi: 10.1007/s00497-015-0259-1

Kawashima, T., Maruyama, D., Shagirov, M., Li, J., Hamamura, Y., Yelagandula, R., et al. (2014). Dynamic F-actin movement is essential for fertilization in *Arabidopsis thaliana*. *eLife* 3:e04501.. doi: 10.7554/eLife.04501

Kimata, Y., Higaki, T., Kawashima, T., Kurihara, D., Sato, Y., Yamada, T., et al. (2016). Cytoskeleton dynamics control the first asymmetric cell division in *Arabidopsis* zygote. *Proc. Nat. Acad. Sci. U.S.A.* 113, 14157–14162. doi: 10.1073/pnas.1613979113

Kost, B., Spielhofer, P., and Chua, N. H. (1998). A GFP-mouse talin fusion protein labels plant actin filaments *in vivo* and visualizes the actin cytoskeleton in growing pollen tubes. *Plant J.* 16, 393–401. doi: 10.1046/j.1365-313X.1998.00304.x

Lazarides, E., and Weber, K. (1974). Filaments Non-Muscle. *Proc. Nat. Acad. Sci. U.S.A.* 71, 2268–2272.

Le, J., El-Assal, S. E. D., Basu, D., Saad, M. E., and Szymanski, D. B. (2003). Requirements for *Arabidopsis ATARP2* and *ATARP3* during epidermal development. *Curr. Biol.* 13, 1341–1347. doi: 10.1016/S0960-9822(03)00493-7

McCurdy, D. W., and Gunning, B. E. S. (1990). Reorganization of cortical actin microfilaments and microtubules at preprophase and mitosis in wheat root-tip cells: a double-label immunofluorescence study. *Cell Motil. Cytoskeleton* 15, 76–87. doi: 10.1002/cm.970150204

Melak, M., Plessnet, M., and Grosse, R. (2017). Actin visualization at a glance. *J. Cell Sci.* 130, 525–530. doi: 10.1242/jcs.189068

Pasternak, T., Tietz, O., Rapp, K., Begheldo, M., Nitschke, R., Ruperti, B., et al. (2015). Protocol: an improved and universal procedure for whole-mount immunolocalization in plants. *Plant Methods* 11, 1–10. doi: 10.1186/s13007-015-0094-2

Płachno, B. J., and Świątek, P. (2012). Actin cytoskeleton in the extra-ovular embryo sac of *Utricularia nelumbifolia* (*Lentibulariaceae*). *Protoplasma* 249, 663–670. doi: 10.1007/s00709-011-0306-1

Povarova, O. I., Sulatskaya, A. I., Kuznetsova, I. M., and Turoverov, K. K. (2012). Actin folding, structure and function: is it a globular or an intrinsically disordered protein? 57–80.

Reddy S. M. (2007). *University Botany-iii:(Plant Taxonomy, Plant Embryology, Plant Physiology)*, Vol. 3. New Delhi: New Age International.

Rezanejad, F. (2008). Development of ovule, female gametophyte, embryo and endosperm in *Petunia hybrida* grandiflora. *Res. J. Univ. Isfahan Sci.* 31, 57–66.

Richardson, D. S., and Lichtman, J. W. (2015). Clarifying tissue clearing. *Cell* 162, 246–257. doi: 10.1016/j.cell.2015.06.067

Roldán, J. A., Rojas, H. J., and Goldraij, A. (2012). Disorganization of F-actin cytoskeleton precedes vacuolar disruption in pollen tubes during the *in vivo* self-incompatibility response in *Nicotiana alata*. *Ann Bot.* 110, 787–795. doi: 10.1093/aob/mcS153

Rudall, P. (1997). The nucellus and chalaza in monocotyledons: structure and systematics. *Bot. Rev.* 63, 140–181. doi: 10.1007/bf02935930

Schneitz, K., Hülskamp, M., and Pruitt, R. E. (1995). Wild-type ovule development in *Arabidopsis thaliana*: a light microscope study of cleared whole-mount tissue. *Plant J.* 7, 731–749. doi: 10.1046/j.1365-313X.1995.07050731.x

Sheahan, M. B., Staiger, C. J., Rose, R. J., and McCurdy, D. W. (2004). A green fluorescent protein fusion to actin-binding domain 2 of *Arabidopsis* fimbrin highlights new features of a dynamic actin cytoskeleton in live plant cells. *Plant Physiol.* 136, 3968–3978. doi: 10.1104/pp.104.049411

Shimamura, M. (2015). "Whole-Mount immunofluorescence staining of plant cells and tissues," in *Plant Microtechniques and Protocols*, eds E. Yeung, C. Stasolla, M. Sumner, and B. Huang (Cham: Springer), 181–196. doi: 10.1007/978-3-319-19944-3_11

Spector, I., Shochet, N. R., Kashman, Y., and Groweiss, A. (1983). Latrunculins: Novel marine toxins that disrupt microfilament organization in cultured cells. *Science* 219, 493–495. doi: 10.1126/science.6681676

Stelly, D. M., Peloquin, S. J., Palmer, R. G., and Crane, C. F. (1984). Mayer's hemalum-methyl salicylate: a stain clearing technique for observations within whole ovules. *Stain Technol.* 59, 155–161. doi: 10.3109/10520298409113849

Świerczyńska, J., and Bohdanowicz, J. (2003). Microfilament cytoskeleton of endosperm chalazal haustorium of *Rhinanthus serotinus* (Scrophulariaceae). *Acta Biol. Cracov. Ser. Bot.* 45, 143–148. doi: 10.1007/s00709-013-0520-0

Tang, X., Lancelle, S. A., and Hepler, P. K. (1989). Fluorescence microscopic localization of actin in pollen tubes: comparison of actin antibody and phalloidin staining. *Cell Motil Cytoskel* 12, 216–224. doi: 10.1002/cm.970120404

Thomas, C., Tholl, S., Moes, D., Dieterle, M., Papuga, J., Moreau, F., et al. (2009). Actin bundling in plants. *Cell Motil. Cytoskel.* 66, 940–957. doi: 10.1002/cm.20389

Vandekerckhove, J., Deboben, A., Nassal, M., and Wieland, T. (1985). The phalloidin binding site of F-actin. *EMBO J.* 4, 2815–2818. doi: 10.1002/j.1460-2075.1985.tb04008.x

Vidali, L., McKenna, S. T., and Hepler, P. K. (2001). Actin polymerization is essential for pollen tube growth. *Mol Biol. Cell* 12, 2534–2545. doi: 10.1091/mbc.12.8.2534

Wulf, E., Deboben, A., Bautz, F. A., Faulstich, H., and Wieland, T. (1979). Fluorescent phallotoxin, a tool for the visualization of cellular actin. *Proc. Nat. Acad. Sci. U.S.A.* 76, 4498–4502. doi: 10.1073/pnas.76.9.4498

Yin, H. L., Hartwig, J. H., Maruyama, K., and Stossel, T. P. (1981). Ca2+ control of actin filament length. Effects of macrophage gelsolin on actin polymerization. *J. Biol. Chem.* 256, 9693–9697.

Yuan, M., Fu, Y., Wang, F., Huang, B., Sze-Yong, Z., and Hepler, P. K. (2002). Fertilization in Torenia Fournieri: actin organization and nuclear behavior in the central cell and primary endosperm. *Sci China Series C Life Sci.* 45, 211–214. doi: 10.1360/02yc9024

Zhang, S., Wang, C., Xie, M., Liu, J., Kong, Z., and Su, H. (2018). Actin bundles in the pollen tube. *Int. J. Mol. Sci.* 19:3710. doi: 10.3390/ijms19123710

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 26 May 2020
doi: 10.3389/fpls.2020.00640

Strong Selection Against Early Generation Hybrids in Joshua Tree Hybrid Zone Not Explained by Pollinators Alone

Anne M. Royer[1], Jackson Waite-Himmelwright[2] and Christopher Irwin Smith[2]*

[1] Biology Department, The University of Scranton, Scranton, PA, United States, [2] Biology Department, Willamette University, Salem, OR, United States

Coevolution frequently plays an important role in diversification, but the role of obligate pollination mutualisms in the maintenance of hybrid zones has rarely been investigated. Like most members of the genus *Yucca*, the two species of Joshua tree (*Yucca brevifolia* and *Yucca jaegeriana)* are involved in a tightly coevolved mutualism with yucca moths. There is strong evidence of a history of coevolution between Joshua trees and their moth pollinators. We use a geographic clines approach in the Joshua tree hybrid zone to ask if selection by the moths may currently contribute to maintaining separation between these species. We compare genomic, phenotypic, and pollinator frequency clines to test whether pollinators maintain the hybrid zone or follow it as passive participants. The results reveal dramatic overlapping genomic and pollinator clines, consistent with a narrow hybrid zone maintained by strong selection. Wider phenotypic clines and a chloroplast genomic cline displaced opposite the expected direction suggest that pollinators are not the main source of selection maintaining the hybrid zone. Rather, it seems that high levels of reproductive isolation, likely acting through multiple barriers and involving many parts of the genome, keep the hybrid zone narrow.

Keywords: *Yucca*, Joshua tree, *Tegeticula*, geographic clines, hybrid zone, coevolution

OPEN ACCESS

Edited by:
Jim Leebens-Mack, University of Georgia, United States

Reviewed by:
Antonio Gonzalez-Rodriguez, National Autonomous University of Mexico, Mexico
Daniel Blaine Marchant, Stanford University, United States

****Correspondence:***
Anne M. Royer
anne.royer@scranton.edu

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *04 January 2020*
Accepted: *24 April 2020*
Published: *26 May 2020*

Citation:
Royer AM, Waite-Himmelwright J and Smith CI (2020) Strong Selection Against Early Generation Hybrids in Joshua Tree Hybrid Zone Not Explained by Pollinators Alone. Front. Plant Sci. 11:640. doi: 10.3389/fpls.2020.00640

INTRODUCTION

Understanding the forces that drive speciation has been a question of interest since the field of evolutionary biology began, and continues to generate complex questions, including the role of intrinsic and ecological factors (Coyne and Orr, 2004; Sobel et al., 2010; Nosil, 2012). The importance of intrinsic isolation, including Bateson–Dobzhansky–Müller incompatibilities in the later stages of speciation, is well established (Coyne and Orr, 1989, 1997) and the role of ecology has attracted particular interest in recent years (Schluter, 2000; Nosil, 2012). In seeking to understand the final stages of speciation, hybrid zones provide a particularly valuable resource—laboratories of speciation where the full range of forces (including genetics, geography, and ecology) can act and interact to shape the trajectory of diversification (Hewitt, 1988; Barton and Hewitt, 1989). Pollinator mediated selection, particularly within obligate pollination mutualisms, has been suggested as an important mechanism by which ecological factors may promote speciation and reproductive isolation (Kiester et al., 1984; Armbruster and Muchhala, 2008), but hybrid zone analyses have rarely been used to test this idea. Examining hybrid zones and clinal variation within presents

a rich framework in which to understand both speciation and the potential role of pollinators in promoting plant diversification.

Ecological forces contributing to speciation can be divided into biotic and abiotic forces, with both playing important and varying roles depending on the system (Coyne and Orr, 2004; Lowry et al., 2008; Nosil, 2012; Baack et al., 2015). Interactions between species are particularly interesting because of the potential for the species to both evolve in response to each other (Janzen, 1980), offering opportunities for direct and diffuse coevolution. Such coevolution can contribute to explaining phenomena involving speciation including adaptive radiations (Ehrlich and Raven, 1964; Kiester et al., 1984; Smith and Benkman, 2007; Althoff et al., 2014; Hembry et al., 2014; Marquis et al., 2016) and the latitudinal biodiversity gradient (Mittelbach et al., 2007). Although there is abundant evidence of pairwise interspecific interactions spurring speciation, largely in the context of antagonisms (Mitter et al., 1988, 1991; Farrell, 1998; Yoder and Nuismer, 2010), the potential for mutualisms to foster diversification is a matter of debate and interest (Yoder and Nuismer, 2010; Hembry et al., 2014). In spite of a large body of literature on interspecific interactions and speciation, there are relatively few studies examining obligate pollination mutualisms in hybrid zones (but see Leebens-Mack et al., 1998; Moe and Weiblen, 2012).

The genus *Yucca* in general is a useful system for studying speciation and mutualism due to the tight relationship between many *Yucca* species and their moth pollinators (*Tegeticula* spp.). The vast majority of *Yucca* are pollinated exclusively by the obligate mutualist moths, often in a reciprocally obligate relationship (i.e., one species of *Tegeticula* for one species of *Yucca*). *Tegeticula* are nursery pollinators, laying their eggs in the developing fruits of the flowers they pollinate. The resulting larvae exact a cost on the plants, eating many of the developing seeds. In many of these plant-moth pairs, there is phylogenetic evidence of co-speciation, with the patterns of divergence in moths and plants frequently matching (Althoff et al., 2012). This raises the question of whether the mutualism contributes to speciation or simply causes one mutualist to follow its partner in diversification (Althoff et al., 2014). One group that shows promise for addressing these questions is Joshua trees and their pollinators.

Joshua trees (*Yucca brevifolia* and *Yucca jaegeriana*) offer the opportunity to fill this gap, taking advantage of natural variation in both mutualists across an existing hybrid zone (Rowlands, 1978; Smith et al., 2008) to understand how the interspecific interaction contributes to late-stage speciation. In Joshua trees, there is additional evidence that the sister species (*Y. brevifolia* and *Y. jaegeriana*) and their moth pollinators (*Tegeticula synthetica* and *Tegeticula antithethica*) have coevolved (Pellmyr and Segraves, 2003; Godsoe et al., 2008; Smith et al., 2008). Work on coevolution in these taxa has focused on two key characters that appear to be integral to the mutualism: style length in the trees and body size in the moths (Godsoe et al., 2008; Cole et al., 2017) (**Figure 1**). Of a suite of phenotypic traits measured in the trees, style length (the distance between the stigma and the ovules, and the site of moth oviposition; Trelease, 1893; Cole et al., 2017) is one of the most consistently and dramatically

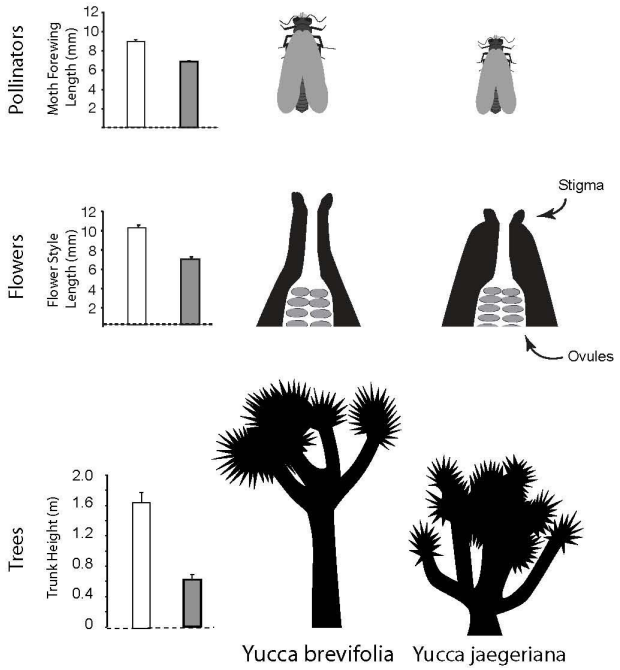


FIGURE 1 | Variation in body size in Joshua tree moth pollinators (graph illustrates a common proxy, wing length), and variation in two key characters differentiating the two species of Joshua tree: style length and trunk height.

differentiated between *Y. brevifolia* and *Y. jaegeriana*, with trunk height being just behind (Godsoe et al., 2008) or slightly more strongly differentiated (Royer et al., 2016). Most importantly, style length and moth ovipositor length are strongly correlated in comparisons across species (with *Y. brevifolia* having long styles and *T. synthetica* a larger body) (Godsoe et al., 2008, 2010; Yoder et al., 2013). This pattern of trait matching, consistent with coevolution, may also be present (although weaker) in intraspecific variation (compare Godsoe et al., 2010 vs Yoder et al., 2013).

The existence of a hybrid zone between *Y. brevifolia* and *Y. jaegeriana*, located in Tikaboo Valley in southern Nevada (Starr et al., 2013; Royer et al., 2016), has made more detailed studies of potential coevolution and the dynamics of speciation between this species pair possible. Data on pollinator fidelity and performance on the alternate host show that both moth species prefer their normal host and have lower fitness on the alternate, which may be related to phenotype matching (Smith et al., 2009). There is also asymmetry in crossing success, with *T. antithethica* found more frequently, and producing more offspring, on the "wrong" host than *T. synthetica* (Smith et al., 2009). This is expected to result in gene flow from *Y. jaegeriana* into *Y. brevifolia* via pollen. Such gene flow is supported by population genetics work using microsatellite markers in and around the hybrid zone indicating nuclear gene flow into *Y. brevifolia* (Starr et al., 2013), as well as range wide chloroplast data (Smith et al., 2008) showing *Y. brevifolia* chloroplasts moving into the *Y. jaegeriana* nuclear background (presumably as they are

swamped by *Y. jaegeriana* pollen). Additionally, analyses of single nucleotide polymorphisms associated with trait variation show strong divergent selection in allopatry and disruptive selection in the hybrid zone on style length, as expected if style length is a key trait in divergence (Royer et al., 2016). Thus, key pieces of ecological and population genetic data are consistent with the pollination mutualism playing a role in diversification.

However, other pieces of evidence cloud the picture, suggesting that the pollinators are close followers rather than drivers of speciation in the Joshua tree hybrid zone. The two *Yucca* species are highly genetically differentiated, with F_{ST} across the hybrid zone estimated at 0.29–0.30 (Royer et al., 2016). This level of differentiation in close proximity supports strong reproductive isolation, in spite of range-wide microsatellite and chloroplast data suggesting a long history of gene flow between species (Smith et al., 2008; Yoder et al., 2013). Range wide nuclear data show gene flow primarily from the hybrid zone into *Y. jaegeriana*, opposite the direction predicted by pollinator behavior (Yoder et al., 2013) and gene flow found in proximity of the hybrid zone (Starr et al., 2013). Finally, the same analysis that showed strong disruptive selection on style length also showed the same results for other strongly differentiated traits in and around the hybrid zone, particularly trunk height (Royer et al., 2016). In order to disentangle how the suite of forces acting in the Joshua tree hybrid zone are shaping the final stages of speciation, a new approach is clearly required.

Geographic clines analyses, which quantify how allele frequencies and phenotypes shift as one moves across a hybrid zone, can help discern what evolutionary and ecological processes are currently most important in shaping the hybrid zone (Barton and Hewitt, 1985; Mallet et al., 1990; Stankowski et al., 2017). This explicitly spatial approach can give us insight into how selection changes across the hybrid zone: which traits and parts of the genome are under stronger selection, how the distribution of the moth species changes across the hybrid zone, and how moths and trees interact. If taxa that diverge in allopatry come back into contact and hybridize, there are several possible evolutionary outcomes. If there is no reproductive isolation, no selection against hybrids, one expects abundant gene flow between the parental taxa with unimpeded symmetrical introgression. This creates a non-clinal pattern; a figure depicting change in phenotype or allele frequency over space would be a straight line with constant slope, and over many generations the hybrid zone would broaden and eventually disappear (Barton and Hewitt, 1985). On the other hand, if hybrids are less fit, selection will favor a quick transition over space from one parental taxon to another, producing a cline—a line depicting allele frequency or phenotype changes over space that will be flat at the edges of the hybrid zone, and then suddenly assume a steep slope at the center (**Figure 2**). The stronger the selection, the narrower the cline (the transition happens over a smaller space). The center of the cline—the inflection point where the slope is the steepest—then indicates the geographic location where selection switches from favoring the genotype/phenotype of one parent to the other. Stepped clines, with particularly steep centers and long introgression tails, indicate particularly strong selection against early-generation hybrids and/or hybrids in the geographic center of the hybrid zone, with only sections of the genome separated by recombination successfully introgressing. Consistent selection against hybrids will thus result in a narrow straight tension zone, with many coincident clines, which is expected if speciation is advanced and selection is limiting introgression across the genome (Key, 1968; Barton, 1979; Barton and Hewitt, 1985; Buggs, 2007).

Here, we use a geographic clines approach to analyze changes in moth abundance, nuclear and chloroplast genomic markers, and phenotypic traits across the hybrid zone to understand the process of speciation in Joshua trees. First, we hypothesize that selection for differentiation in style length, driven by coevolution with moth pollinators, is maintaining the separation of the species. We predict that strong selection from the pollinators will result in the geographic cline for style length being narrower than the clines for other traits, and closely coincident with the distribution of the moths. We expect geographic clines for SNPs associated with variation in style length to show the same pattern, coinciding closely with the distribution of moths shifting across the hybrid zone when compared to the genome-wide average and SNPs associated with other traits. Second, we hypothesize that trait mismatch between moths and floral features results in asymmetry in the direction of successful crosses, which should impact cytoplasmically inherited genomic material differently from the nuclear genome. Accordingly, we predict chloroplast haplotype clines should be shifted east relative to nuclear SNP clines, consistent with existing data on pollen flow and pollinator behavior.

MATERIALS AND METHODS

Genotype and/or trait data were collected from a total of 1734 trees in the hybrid zone and nearby allopatric areas in Tikaboo Valley, Nevada, United States (37.46°N, 115.5°W), with heaviest sampling in the hybrid zone (**Table 1** and **Figure 3**). Trees were sampled across the entire known hybrid zone, with sampling most intense near the center; a gap in sampling west of the hybrid zone represents the discontinuous distribution of trees rather than undersampling (**Figure 3**). Variation in floral and vegetative traits between species has been described previously (Godsoe et al., 2008; Royer et al., 2016). We used geographic cline analysis to look at the spatial distribution of hybrid zone variation in traits that differ significantly between the two species, including branch number, tree height, leaf length and width, trunk height, petal length and width, pistil length and width, style base width, and style length. Because Joshua trees do not flower each year, fewer trees have measurements for floral traits than trunk height (**Table 1** and **Supplementary Figure S1**). For details on how traits were measured, see Royer et al. (2016).

To estimate the distribution of moths, sticky traps were placed on flowering trees (including members of both parental species and hybrids) across the zone of sympatry during the springs of 2011–2013 and 2015. In each year, the sticky traps were left in place for the length of the flowering season. *Tegeticula* moths stuck to the traps were identified to species using diagnostic variations in the cytochrome oxidase gene sequenced in one

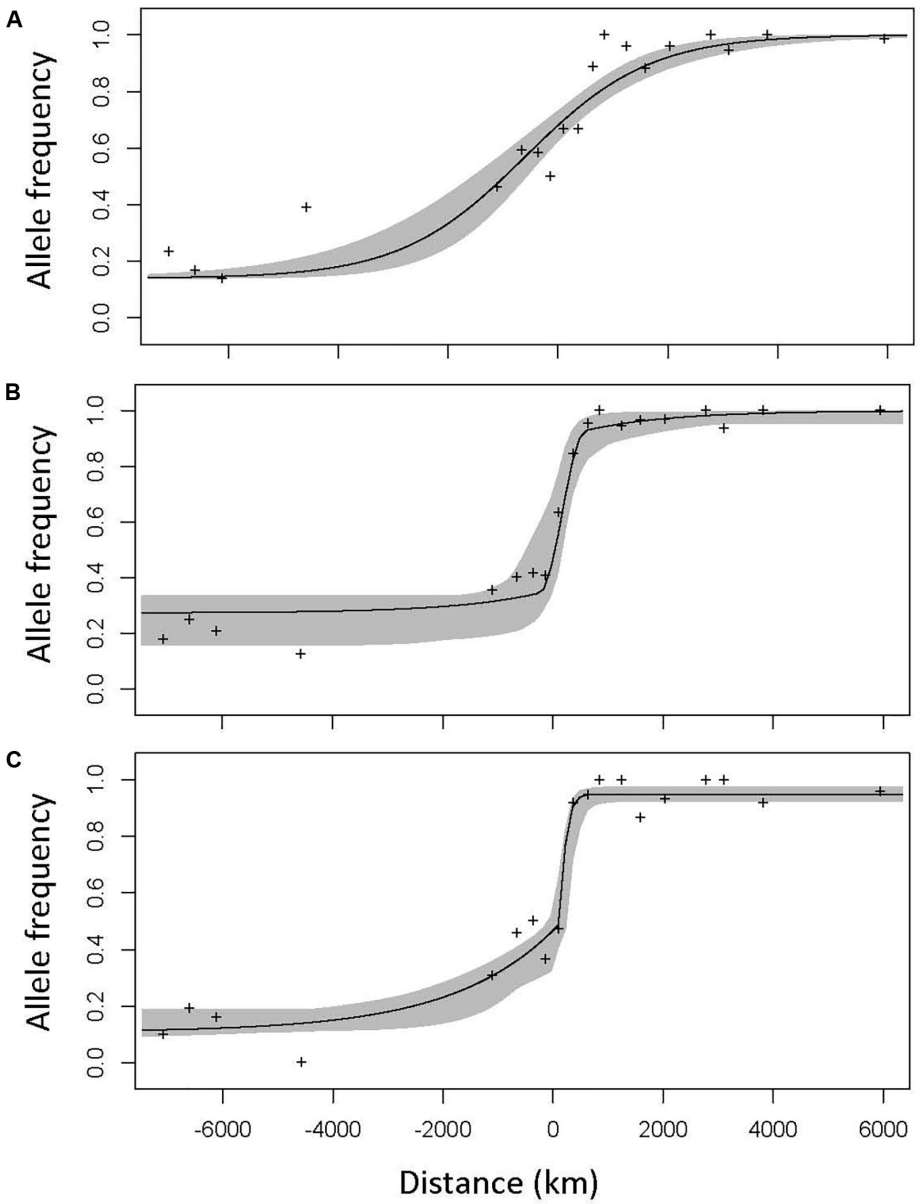


FIGURE 2 | The three cline model types, illustrated using the clines of high−F_{ST} SNPs in the Joshua tree hybrid zone. The center of the hybrid zone (0.5 hybrid contour, determined using microsatellite genotypes) is marked zero. Model I **(A)** is characterized by a gradual, symmetrical cline. Model II **(B)** is a narrower cline with long, symmetrical introgression tails, characteristic of strong selection against hybrids in the early generations and/or in the center of the hybrid zone, but with portions of the genome escaping and successfully introgressing into both parental backgrounds. Model III **(C)** shows similarly strong selection, but with substantially more introgression in one direction (in the case of the Joshua tree hybrid zone, that direction is always to the west—from *Y. jaegeriana* into *Y. brevifolia*). A null model would be a perfectly straight line from the allele frequency on the western edge to the allele frequency on the eastern edge, often (but not necessarily) with slope zero.

direction from the 5′ end using Sanger sequencing by Eurofins (Louisville, KY, United States) (Pellmyr and Segraves, 2003; Smith et al., 2008, 2009).

Leaf tissue for genotyping trees was collected in the field and dried with silica gel, or flash frozen in liquid nitrogen and later transferred to −80°C. DNA was extracted using the Qiagen (Germantown, MD, United States) DNEasy Plant mini kit. Ten microsatellite loci were amplified from 1469 trees following the protocol of Flatz et al. (2011). RAD-seq data were produced as described in Royer et al. (2016), resulting in 9516 SNPs at 4000 loci. Because of the greater time and expense involved in generating RADseq data, these SNPs were genotyped in a smaller, but still substantial number of trees (319) sampled in and immediately around the hybrid zone. Previous work in *Yucca* using sequence from five chloroplast genes found no variation in the hybrid zone (Smith et al., 2008). To identify variation in the chloroplast genome within the hybrid zone, we compared whole chloroplast genome resequencing

TABLE 1 | Number of trees sampled from locations outside the hybrid zone (indicated with species name) and within the hybrid zone.

Cline	*Y. brevifolia*	Hybrid zone	*Y. jaegeriana*	Total
Vegetative traits				
Branch number	288	2353	240	2881
Leaf length	278	2398	241	2917
Leaf width	288	2398	243	2929
Trunk height	254	1783	231	2268
Height	288	2398	235	2921
Floral traits				
Petal length	36	424	61	521
Petal width	36	424	61	521
Pistil length	36	426	63	525
Pistil width	36	426	63	525
Style length	36	426	63	525
Style base width	36	426	63	525
Genomic				
Chloroplast genotypes	59	127	29	215
SNP genotypes	62	183	63	308
Pollinator				
Moth sp. frequency	16	72	13	101

The SNP category covers SNP Qscores as well as high F_{ST}, random SNPs, and analyses of loci associated with trait variation.

data from populations near the hybrid zone (Smith et al., in review). We identified a number of polymorphic regions that showed promise of carrying species-specific variants based on 16 trees previously genotyped. These included two short regions that were readily amplified by PCR and that contained one SNP and two variable number tandem repeats (VNTRs). We then genotyped 215 trees at those loci. Primers were identified using the consensus chloroplast genome sequence (first set of primers: cpF4416, 5′-CCA-TTG-TCA-ATA-TGA-GTG-GG-3′; cpR4581, 5′-AAG-GAC-GAA-CCT-TGC-TTA-TT-3′. Second set of primers: cpF9932, 5′-TAT-ACG-TTC-TCG-CGA-TTT-GT-3′; cpR10134, 5′-ATT-TGG-CTT-CAA-TCT-TCC-CT-3′). The 200 bp regions were amplified using Qiagen multiplex PCR kits (Qiagen; Hilden, Germany) with 35 cycles of PCR with a 55°C annealing temperature (94°C to denature, 72°C for elongation). The PCR product was purified with ExoSAP-IT (Affymetrix, Inc.; Cleveland, OH, United States) and sequenced at Eurofins MWG Operon USA (Louisville, KY, United States). The sequence was aligned using CodonCode Aligner (CodonCode Corporation[1]), including segments from the consensus chloroplast genome of each species in the alignment. We combined the genotypes at both variant sites to identify chloroplast haplotypes.

Assigning Individual Trees to Species

Trees were assigned a hybrid index score (identifying parentals and, for hybrids, estimating the proportions of genomes of each tree that came from either species) using the Bayesian clustering software STRUCTURE 2.3.4 (Pritchard et al., 2000). We first confirmed that grouping trees into two species best describes the genetic variation in this region by comparing STRUCTURE runs at levels of K from 1 to 7 (one more than the number of populations sampled). We performed 10 runs per level of K using random seeds with 30,000 burnin followed by 150,000 iterations of MCMC, using the ancestry model with alpha inferred and default parameters (including the admixture model, which produces estimates of proportion genetic identity). Summary statistics for the runs were obtained using STRUCTURE HARVESTER (Earl and vonHoldt, 2012), and indicated convergence across runs at each level of K. We used the ΔK statistic (Evanno et al., 2005) to confirm that K = 2 (corresponding to the two species) is the best fit for our data. Mean probability of a tree belonging to each species was then calculated across the 10 runs at K = 2 using CLUMPAK (Kopelman et al., 2015). We used the proportion of each individual's genome estimated by STRUCTURE as originating in the eastern species *Y. jaegeriana* as the hybrid index score (*Q*). Qscores calculated using microsatellite genotypes at 10 loci previously shown to be informative in these species (Flatz et al., 2011; Starr et al., 2013) were used to construct the geographic cline transect (see below). This maximized the geographic coverage to get the best estimate of where the center of the hybrid zone is. To calculate the genomic cline itself, we used Qscore calculated using SNPs instead. This reduced the number of trees, but increased our confidence in the estimate of genomic composition for each individual tree.

Modeling Pollinator Preference

To test the roles of location in the hybrid zone and genetic identity of the tree in shaping pollinator host choice, we performed a Pearson product-moment correlation test to quantify the correlation between Qscore and location in the hybrid zone, and attempted to disentangle the effects of the two variables on pollinator host choice using a general linear model designating the binomial family, with a model: (Frequency of *T. synthetica/T. antithethica*) ∼ Qscore + distance from center of geographic cline transect.

Constructing the Geographic Cline Transect

To construct a two-dimensional transect for plotting geographic clines, we plotted the GPS coordinates of all sampled trees in ARCGIS ARCMAP 10.3 (ESRI). We used the hybrid index scores to estimate a 0.5 hybrid contour (the center of the hybrid zone) using Empirical Bayesian Kriging (an interpolation method) with prediction and eight sectors. We then estimated the shortest distance between each tree and the 0.5 hybrid contour line using the Near Table tool. The geographic cline of the hybrid zone results from plotting the location of each tree along a single line representing the distance of each from the center of the hybrid zone (the 0.5 hybrid contour line), and then grouping trees that are a similar distance from the center (**Figure 3**). To produce the final dataset for the geographic cline analysis, trees along the cline were divided into bins set every 250 m along the cline. For all clines (vegetative traits, floral traits, pollinator identity, Qscore, SNPs, and chloroplast haplotype), bins with fewer than three trees for at least one category were collapsed with the

[1]www.codoncode.com

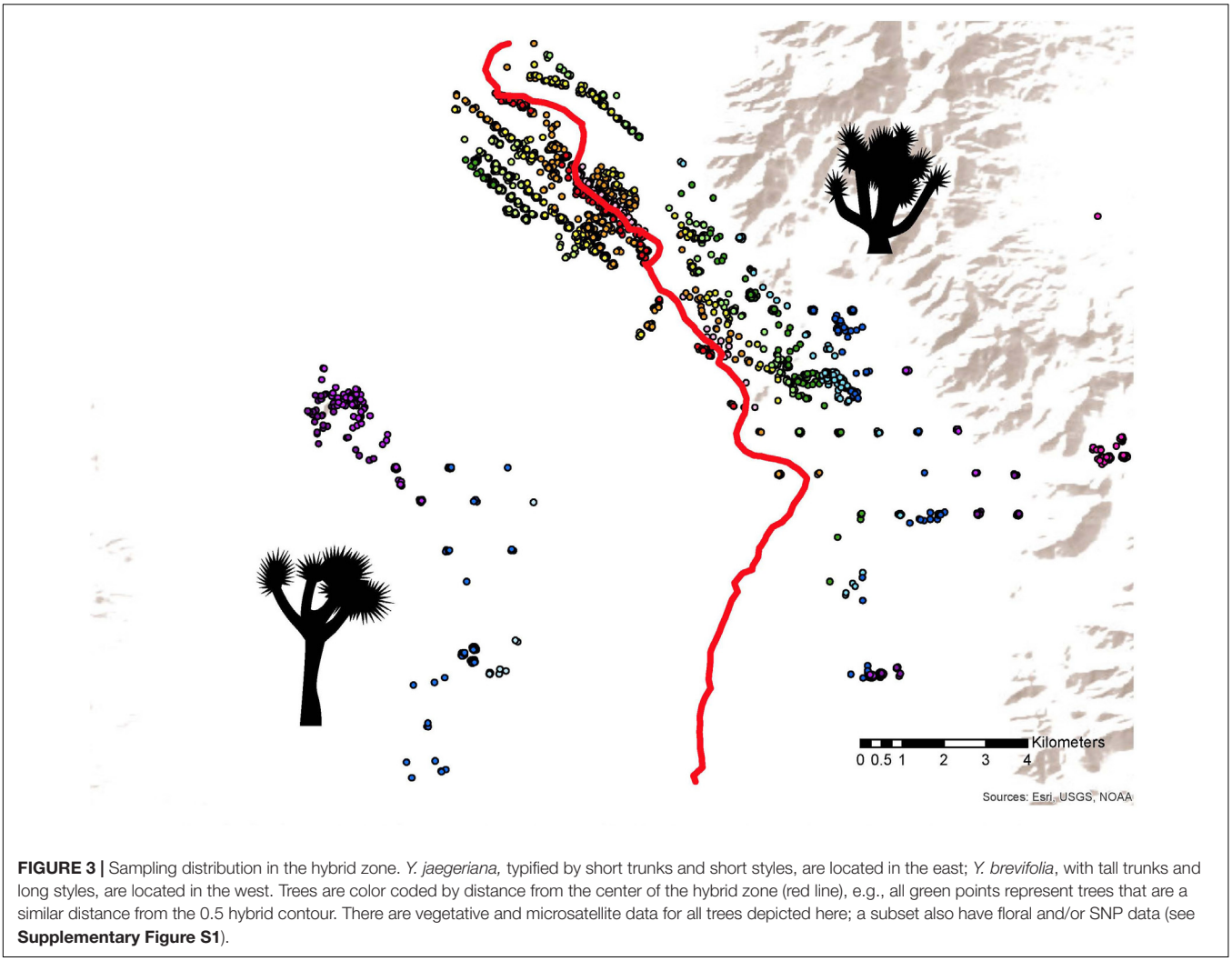


FIGURE 3 | Sampling distribution in the hybrid zone. *Y. jaegeriana,* typified by short trunks and short styles, are located in the east; *Y. brevifolia*, with tall trunks and long styles, are located in the west. Trees are color coded by distance from the center of the hybrid zone (red line), e.g., all green points represent trees that are a similar distance from the 0.5 hybrid contour. There are vegetative and microsatellite data for all trees depicted here; a subset also have floral and/or SNP data (see **Supplementary Figure S1**).

closest adjacent bin, producing 19 standard bins used for all cline analyses (**Supplementary Table S1**).

Fitting Geographic Clines

Genomic Clines: SNPs (Random, High-F_{ST}, and GWAS) and Chloroplast Haplotypes

For calculating nuclear genomic clines, we used SNP data previously collected by Royer et al. (2016), described above. When there were loci with multiple SNPs, we reduced the dataset to one SNP per locus because they are expected to be closely linked and result in identical clines. We retained the most statistically significant SNP (with the lowest *p*-value in the F_{ST} or GWAS analysis) when there was a difference, and eliminated one haphazardly when they were statistically indistinguishable. We fitted clines to the SNPs with the highest F_{ST} (top 1%, for a total of 84), using the F_{ST} analysis performed by Royer et al. (2016). We also fitted clines to SNPs significantly associated with the variation in style length (19 SNPs) and trunk height (29 SNPs) in a genome-wide association (GWAS) analysis of the hybrid zone (Royer et al., 2016). We note that there is some overlap in the SNP datasets: four style SNPs and seven trunk height SNPs exhibited high F_{ST}, and one SNP was significantly associated with variation in both style length and trunk height. The 100 random SNPs did not include any from the GWAS or high F_{ST} sets. For each cline, allele frequencies for each bin were output directly from the genomic analysis software STACKS (Catchen et al., 2013) in HZAR (Derryberry et al., 2014) format.

For the chloroplast data, we identified only two haplotypes within the hybrid zone. We calculated clinal variation in the frequency of these two haplotypes as described for SNP data above.

Trait Clines on Continuous Variables (Phenotypes, Qscores, Moth sp. Frequency)

For phenotypic traits, SNP-based Qscore, and moth abundance data, we standardized the data to a scale from 0 to 1, then ran a linear regression of bin mean trait value on distance along the transect of the hybrid zone. Traits with no significant linear regression were considered to not have clines (i.e., best fit a null model). For those that did change significantly across the hybrid zone, we fitted clines to the data using HZAR, a software

package implemented in R that uses the Metropolis–Hastings Markov Chain Monte Carlo algorithm to estimate a curve describing how allele frequencies of phenotypes change across the hybrid zone (Derryberry et al., 2014). Trait distributions were visually assessed for normality (an assumption of HZAR), and transformed when possible to attain normality (tree height was log-transformed, trunk height was square-root transformed). Only moth frequency (a strongly bimodal distribution) was unable to be successfully transformed; we fitted the clines on the untransformed distribution. Distance along the cline for each bin was calculated as the mean distance of all the trees included in the bin.

For all clines, we fitted three models: (I) a simple two-parameter model, setting only a center and width for the cline; (II) a four-parameter model, describing the shape of symmetric introgression tails ($\beta_0/2$ and theta0); and (III) six-parameter models with two separate parameters for each asymmetric introgression tail ($\beta_0/2$ and theta0, and $\beta_1/2$ and theta1) (**Figure 2**). All models also include a *y*-axis minimum and maximum (pmin and pmax), corresponding to a mean trait value or allele/haplotype frequency in parental populations. For each model type, the parameters and log likelihood of the best-fitting model were used for analysis. For SNP clines, all three models were compared to a null model run in HZAR, which would be a line with constant slope between pmin and pmax (no cline—analogous to the non-significant linear regression described for phenotypic clines above, as HZAR provides no null model for phenotypic clines). We determined the best-fitting model for each trait using AIC values output by HZAR; we selected the model with the lowest AIC, or the simplest model if dAIC < 2.

Analysis Comparing Clines

We compared clines by assessing overlap in centers and widths, using the 2 log likelihood (km) estimate (95% credible interval). SNP clines that best fit the null model were excluded from calculations of the mean SNP cline parameters for each category (high F_{ST}, random, trunk height, and style length). We then assessed whether there was overlap with the 95% CI of other clines (lack of overlap indicates a significant difference). To evaluate possible sources of selection, we compared cline centers.

Inferring Levels and Patterns of Hybridization Inside the Zone

Understanding what classes of hybrids are present in a hybrid zone (e.g., F1s, F2s, advanced hybrids, backcrosses) can be helpful in interpreting clines and understanding how advanced speciation is. While hybrid index scores give information about the proportion of the genome originating in each parent, they cannot distinguish between F1s and advanced generation hybrids that maintain equal proportions of the two genetic backgrounds. To do this, it is necessary to incorporate the degree of heterozygosity at each locus into the analysis. When populations are fixed for different alleles at a set of unlinked, biallelic loci, patterns of hybridization can be inferred from hybrid index scores combined with heterozygosities. For example, crossing pure individuals will produce an F_1 individual that has a hybrid index of 0.5 (50% of alleles inherited from populations and *i* and *j*) will be heterozygous at all loci ($H = 1.0$). Similar analytical predictions can be made for other early generation classes, including backcrossed individuals and hybrid intercrosses (i.e., F_2s). However, if loci are not completely diagnostic, allele sharing at loci muddies these predictions, making it more difficult to interpret empirical patterns.

Given that the 101 most highly differentiated loci are only semi-diagnostic in these taxa, we used individual-based simulations to generate distributions of multilocus genotypes that are expected to result from hybridization. To do this, we created two groups of "pure" individuals of *Y. jaegeriana* and *Y. brevifolia* based on their STRUCTURE (*Q*) scores and geographic distributions. We then used the program hybrid lab (Nielsen et al., 2006) to produce hybrid genotypes by "crossing" individuals from these groups and subsequent hybrid classes *in silico*. We produced an F_1, an F_2, an F_1 backcross to *jaegeriana*, and an F_1 backcross to *brevifolia*. 500 individuals were simulated for each cross type. For each simulated and empirical individual, we plotted the proportion of heterozygous loci (*H*) on its hybrid index score (HI) defined here as the proportion of alleles coming from pure population *i*. Note that the simulations assume that markers recombine freely ($r = 0.5$), so if some of these loci are tightly linked, we would be underestimating the number of generations of hybridization required to generate hybrid classes beyond the F_1.

RESULTS

Pollinator Occurrence

We found both pollinator species co-occurring in the center of the hybrid zone, over a 4.2 km wide band ranging from ∼1.1 km west of the center to ∼3.1 km east of the center, with 90% of trees with both species observed on them located between 0.4 km west and 2.4 km east. There are instances of moths choosing the "wrong" host, including trees with Q-scores beyond the threshold for identification as "pure" parentals. However, the strong geographic structure in the location of different tree identities leads to a strong correlation between tree location in the hybrid zone and tree Qscore (Pearson's product-moment correlation = 0.76, $t = 7.23$, $p = 1.22^{*}10^{-8}$, $df = 38$). The general linear model showed significant effects of both tree Qscore and location in predicting moth visitation, with Qscore being much more strongly significant (Qscore: $z = 4.384$, $p = 1.16^{*}10^{-5}$; location, $z = 2.16$, $p = 0.03$).

Model Fit and Cline Centers

We found that the hybrid zone is delineated by a clear and narrow nuclear genomic center, whether it is measured using SNP-Qscore, high F_{ST} SNPs, or a random set of nuclear SNPs (**Table 2** and **Figures 4**, **5**). The frequency of different model types (**Supplementary Table S2**) reinforces this. As expected, the random SNPs have a much higher frequency of null clines than the other sets of SNPs, indicating that they are more likely to freely introgress (only one of the high-F_{ST} clines is a null, and only 5% of style length and 17% of trunk height SNPs). However, the majority of even random SNPs (80%) exhibit clines

TABLE 2 | Cline centers, maximums, and minimums (for the two ends of the cline) for non-SNP geographic clines (means of all SNP clines from the high-F_{ST} and random-SNP sets are included).

Trait	Best model	Center	Width	Center low estimate	Center high estimate	Pmin (west)	Pmax (right)
Style base width	II	−2265.88	10,613.62	−4336.89	531.03	0.71	0.41
tree height	II	−2155.21	2702.79	−2284.25	−2065.65	0.81	0.50
Leaf width	II	−2038.1	11,696.69	−3754.44	−1289.55	0.60	0.14
Cp haplotype	I	−852	276.40	−1036.3	−781.56	0.02	1.00
Mean random SNP	–	−124.22	961.96	−1945.28	1006.96	0.83	0.97
SNP Qscore	II	0	2236.39	−82.32	93.26	0.02	0.99
Trunk height	II	56.02	2139.12	49.81	118.1	0.10	0.40
Mean high F_{ST}	–	105.85	2170.11	−193.77	404.56	0.12	0.98
Leaf length	III	157.37	1297.79	136.66	242.06	0.44	0.22
Style length	II	521.8	674.61	390.58	876.81	0.19	0.61
Moth	III	676.95	1229.31	585.21	676.95	0.00	0.99
Petal width	I	3251.09	6626.81	2280.26	5438.5	0.50	0.27

*Clines are listed in order of center location, from west to east. All measurements in meters. Low and high estimates for center are ± two log likelihoods of the mean, as produced by Hzar; for mean random SNP and mean high F_{ST} SNPs, the mean high and low estimates were used. For data on individual SNP clines, see **Appendix**. Traits for which linear regression across the cline was non-significant were excluded (branch number, petal length, pistil length, and pistil width). Pmin and Pmax for alleles are set with the minor allele frequency to the west; for phenotypes, pmin represents the west and pmax the eastern part of the distribution.*

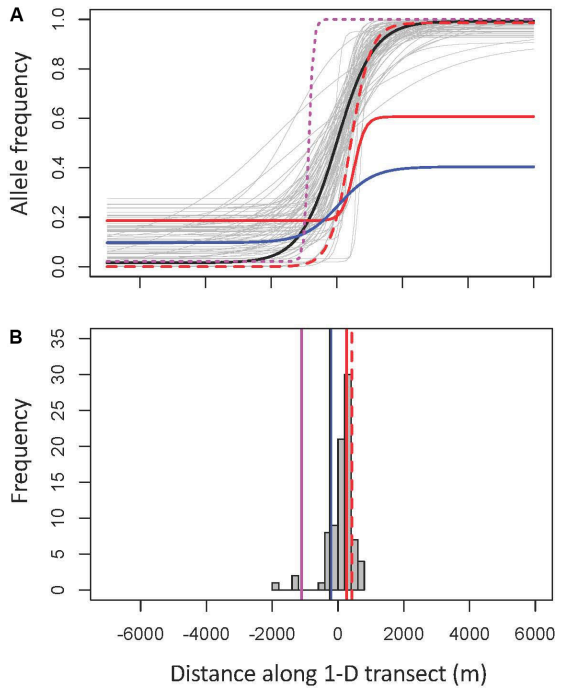


FIGURE 4 | Geographic clines of traits and genotypes in the Joshua tree hybrid zone. Gray represents high F_{ST} SNPs (portrayed as a histogram in **B**), included as a reference showing the cline distribution of parts of the genome likely under the strongest natural selection. Dashed red is moth frequency, black is Qscore, solid red is style length, solid blue is trunk height, and magenta is chloroplast haplotype frequency. **(A)** Full clines. All clines are oriented with the minimum on the left for clarity, which required flipping trait clines vertically. **(B)** Cline centers.

(all but one the simplest model) (**Supplementary Table S2**). The majority of high-F_{ST} clines (62%) and several trunk height and style length clines fit the more complex model II, with non-negligible proportion of high-F_{ST} clines fit best by model III (19%) (**Supplementary Table S2**). All of the asymmetric (model III) clines show longer introgression tails toward the west (see **Appendix**).

The two morphological traits that vary most clearly between the species, trunk height and style length, have phenotypic clines that are closely coincident with the nuclear genomic clines, as does leaf length (**Figure 4** and **Tables 2**, **3A**). While the average of individual clines for SNPs associated in GWAS with trunk height and style length was coincident with the nuclear genomic clines, there was a lot of variation between loci; for style length in particular, some SNPs had clines with centers far from the hybrid zone center (**Figure 5** and **Supplementary Table S2**). Comparing them to the mean high-F_{ST} cline shows the least consistency in the style SNPs, with half having centers significantly offset from the high-F_{ST} mean (one to the west, eight to the east). Trunk SNPs showed a more bimodal pattern, with one group fitting the high-F_{ST} pattern fairly well and the other diverging in similar ways to the style SNPs. Just over a third of the trunk height SNP clines had centers significantly offset from the high-F_{ST} mean (one to the west, five to the right). In addition to being more frequent, the aberrations in style SNPs are more extreme than those in trunk SNPs (**Figure 5** and **Appendix Table**).

We found two chloroplast haplotypes in the hybrid zone, one clearly associated with each Joshua tree species, allowing us to perform a successful geographic cline analysis on this genotype. The cline center for chloroplast haplotype frequency is significantly offset from the nuclear genomic clines, and opposite the direction we predicted—to the west rather than the east (**Table 2** and **Figure 4**). The chloroplast cline center is just east of the gap in the Joshua tree distribution, which is at most 2 km wide at the narrowest point with *Y. brevifolia* on the west and the hybrid zone to the east (**Figure 3**) (it may in fact be somewhat narrower, as we have not methodically sampled the edges). The cline is also startlingly steep, with only a single bin containing both genotypes (the bin directly adjacent to the gap, 1.07 km

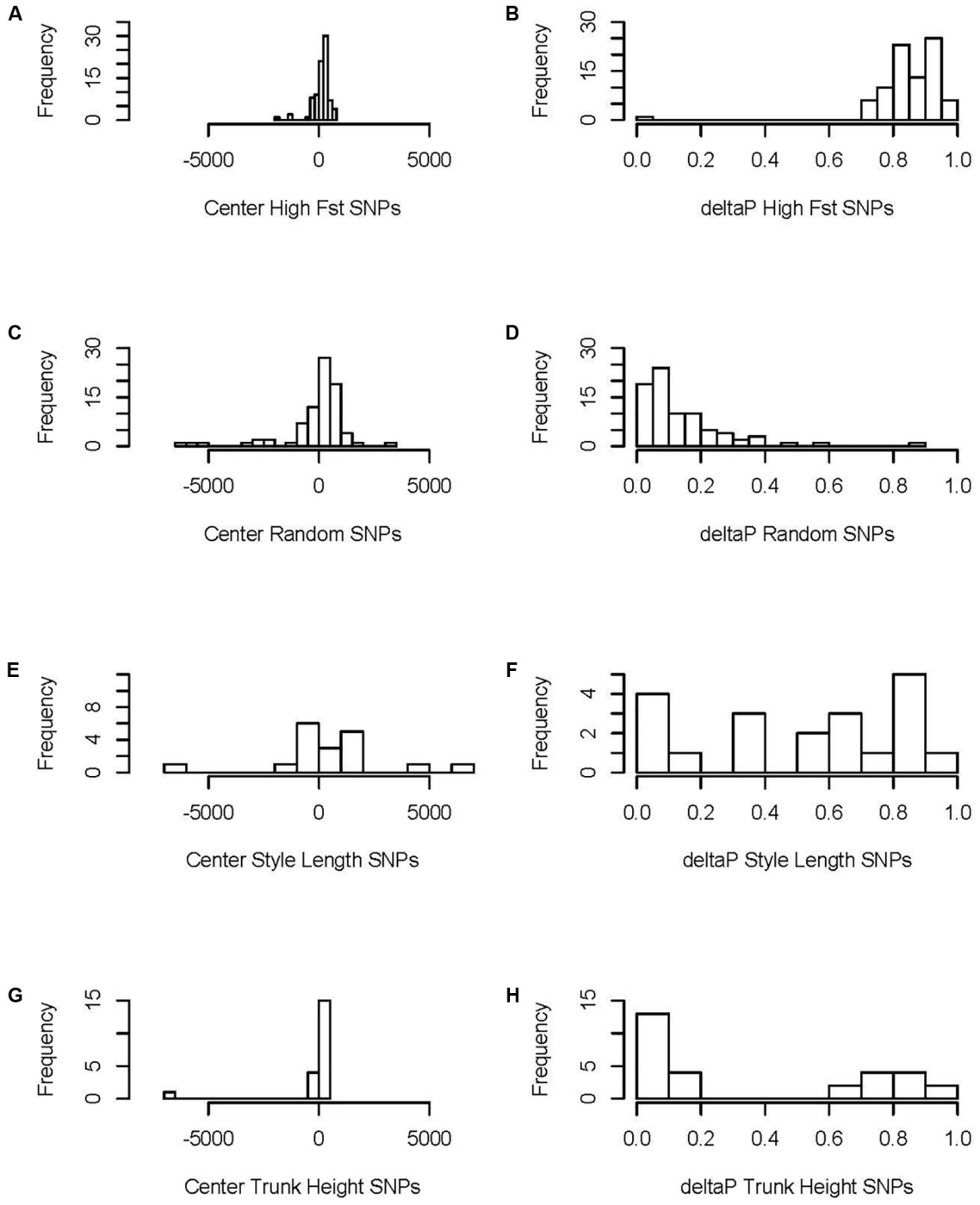


FIGURE 5 | Distribution of cline parameters: cline centers (left) and delta P (the variation in allele frequency between the two ends of the hybrid zone, right) for high F_{ST} SNPs, randomly selected (non-high F_{ST}) SNPs, and SNPs associated with style length and trunk height. The left column of the figure depicts the distribution of cline centers for SNPs, broken down by **(A)** high Fst, **(C)** random, **(E)** Style length, and **(G)** trunk height. The right column of the figure depicts the distribution of the difference between values on the extreme ends of the clines for SNPs, broken down by **(B)** high Fst, **(D)** random, **(F)** Style length, and **(H)** trunk height.

wide—with the exception of a single *Y. jaegeriana* haplotype appearing on the west side of the gap).

The phenotypic clines for other morphological traits deviated from the center of the hybrid zone in ways we did not predict. Tree height, style base width, and leaf width clines are offset in the same direction as the chloroplast haplotype cline, but even more dramatically (although the confidence intervals for estimating style base width cline center are so wide that they cross the center

of the hybrid zone). A single morphological trait cline, petal width, is significantly offset to the east (**Table 2**).

Pollinator species abundances shift sharply in the center of the hybrid zone. The cline for moth species is, as predicted, most closely coincident with the style length cline. It also clusters with the nuclear genomic, phenotypic trunk height, and phenotypic leaf length clines (**Figure 4** and **Table 2**).

The nuclear genomic clines together show a clear picture of higher genetic diversity (measured as more intermediate allele frequencies) in *Y. brevifolia* than in *Y. jaegeriana* (**Figure 4**).

Cytonuclear Genotype Data

When we compare nuclear genotypes (summarized using Qscore) to chloroplast genotype, we see some dissociation of nuclear and cytoplasmic genomes, with *Y. jaegeriana* chloroplasts dominating in the hybrid zone. Because the sample size of trees genotyped at both SNPs and the chloroplast was small, we used Qscores calculated with microsatellite genotypes for the nuclear genome estimate (the same used to set the 0.5 hybrid index line for the clines) instead. This produced a set of 183 trees with data on cytonuclear genotypes (a distinct dataset, as not all of the trees with SNP genotypes had chloroplast genotypes and vice versa). Of the 43 trees in this set that are genomically hybrids (SNP Qscores between 0.15 and 0.85, Royer et al., 2016), only 21% (nine trees) have the *Y. brevifolia* chloroplast haplotype. There are no trees with Qscores above 0.44 that have a brevifolia chloroplast haplotype.

Level and Pattern of Hybridization in the Zone

We estimated how many of the trees in our study were likely F1 hybrids, parental, backcrosses, or advanced generation hybrids (F2+) by plotting heterozygosity (*H*) on hybrid index score (HI), which revealed clear evidence for extensive genetic mixing between *Y. brevifolia* and *Y. jaegeriana* (**Figure 6**). Individuals from within the hybrid zone nearly spanned the full range of hybrid index scores, ranging from 0.14 to 0.97. Some individuals from the hybrid zone have genotypes consistent with being pure individuals of each species, as they overlap the clusters of *Y. brevifolia* and *Y. jaegeriana* from outside the hybrid zone. However, most individuals show some evidence of mixed ancestry. Even within the hybrid zone, the distribution of hybrid index scores was strongly clinal (Pearson correlation coefficient distance of hybrid index and geography = 0.66, $p < 0.0001$), so there may be little opportunity for pure individuals to mate with one another.

The distributions of simulated F_1, F_2 and backcross individuals provide deeper insight into the patterns and levels of hybridization between *Y. brevifolia* and *Y. jaegeriana.* For instance, none of the empirical hybrids had heterozygosities that were as high as is expected for an F_1 hybrid, suggesting that they are absent from the zone. Very few individuals also overlapped the distributions of the first-generation backcrosses, mainly because their heterozygosities were lower than expected for these hybrid classes. Rather, most individuals are likely to be later generation backcrosses, or the results of mating between backcrosses and later-generation hybrids. However, some individuals overlap the distribution of simulated F_2s, so may be early generation hybrids (**Figure 6**). Overall, these data indicate that individuals from this zone are highly mixed.

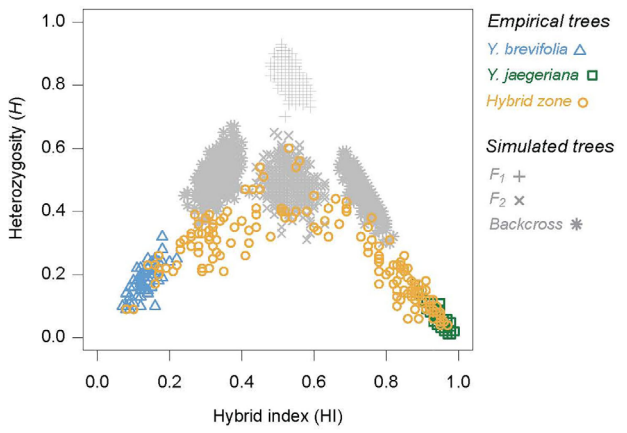


FIGURE 6 | The joint distribution of hybrid index scores and heterozygosities for empirical and simulated trees based on the 101 highly differentiation loci. Empirical trees include pure individuals of both taxa, and individuals from inside the hybrid zone. The distribution of four simulated hybrid classes are shown in gray, including F_1s, simulated by crossing the pure populations, F_2s simulated by intercrossing the simulated F_1s, and backcrosses, simulated by crossing the simulated F1s back to each parental type.

DISCUSSION

Our analysis of the Joshua tree hybrid zone found a narrow tension zone dominated by advanced generation hybrids (**Figure 6**), with largely coincident clines for tree genotype, phenotype, and pollinator occurrence (**Figure 4**). This is consistent with selection against hybrids resulting in moderate to strong reproductive isolation acting across the whole genome. The exception is the chloroplast cline, which was significantly offset to the west (opposite the direction predicted by previously documented gene flow) (**Figure 4** and **Table 2**).

The concordance of the clines for moths and tree hybrid index is striking, with both being quite narrow, showing sudden and dramatic transitions in the center of the hybrid zone. The cline centers are offset, but overlap substantially over the geographic area of transition (width) in spite of steep transitions (**Figure 4**). There are two possible, non-mutually exclusive causes for the close fit of moth and tree clines. First, it may be that moth species perform poorly on non-native hosts, and do not disperse far. This is analogous to the classic definition of a tension zone (Barton and Hewitt, 1985), although in a pair of "good" species: moths may choose the "wrong" host readily, but such moths have fitness that is reduced (*T. antithethica*) or negligible (*T. synthetica*), reproducing little if at all. Moths would therefore rarely be found more than one generation's dispersal distance from their host species. The other possibility is that they are excellent at detecting the correct host species, and their distribution mirrors that of the host because of behavior rather than selection-dispersal balance.

TABLE 3 | Comparisons of geographic clines.

	Style base	Tree height	Leaf width	Chloroplast	Random SNP	SNP Qscore	Trunk height	High F_{ST}	Leaf length	Style length	Moth	Petal width
Style base width	–	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	no	no
Tree height		–	**Yes**	No	**Yes**	No	No	No	No	No	no	no
Leaf width			–		No	No	No	No	No	No	no	no
Chloroplast				–	**Yes**	No	No	No	No	No	no	no
Mean random SNP					–	**Yes**	No	**Yes**	**Yes**	**Yes**	**yes**	no
SNP Qscore						–	**Yes**	**Yes**	No	No	no	no
Trunk height							–	**Yes**	No	No	no	no
Mean high F_{ST}								–	**Yes**	No	no	no
Leaf length									–	No	no	no
Style length										–	**yes**	no
Moth											–	no
Petal width												–

Cline coincidence, whether the cline centers are the same. Clines are listed in order of center location, from west to east. Cline comparisons with statistically indistinguishable parameters—that is to say, ones that are the same—are indicated with "yes" in bold.

Our statistical test suggests both mechanisms may be important, but the genetic identity of the tree is likely most important (this in spite of the strong correlation between genotype and location in the hybrid zone). Consistent with this, Smith et al. (2009) finds that pollinators show a strong preference for their host species, and that offspring survival on the wrong host is reduced at best. Preliminary data on *Tegeticula* movement in the hybrid zone show individual moths move an average of at least ∼30 meters (Autumn and Smith, unpublished data). Thus it seems likely that the matching clines of pollinator and host are due to a combination of choice, natural selection, and low dispersal.

Given the steep, narrow cline for pollinator species occurrence, we would expect strong selection from moths on style length to produce coincident and equally dramatic clines for the trait and correlated SNPs. Evidence for this is mixed. The phenotypic style length cline is coincident and concordant with the moth cline—in fact, the moth cline fits the style length cline better than it fits the hybrid index cline, and style length is the only phenotypic cline for which both the center and width are statistically indistinguishable from that of moths (**Table 3**). However, the clines for SNPs associated with variation in style length do not consistently show the same pattern, and include loci with cline centers significantly displaced from the moth cline in both directions (**Figure 5** and **Supplementary Table S2**). The relatively low number of SNPs identified through RADseq means it is likely that we missed selection on many parts of the genome, making our pool of SNPs correlated with style length smaller than it should be (Lowry et al., 2017a,b). However, there is no reason to suspect this would bias our sample in favor of weaker correlations or false positives. On the other hand, if our threshold for significance in the GWAS was too low, it would lead to the presence of false positives in the style SNP pool. This could explain why the SNP clines are so much more diffuse than the phenotypic cline for style length. Regardless, given the data we have, our genotypic clines are not consistent with strong selection on style length in the hybrid zone.

When all the clines are viewed as a whole, the main theme is overall coincidence of narrow genomic clines (including random SNPs not under selection as well as Qscore and high-F_{ST} SNPs) (**Figure 5**). This is characteristic of tension zones, where hybrids are less fit than parents and selection is acting against hybrids on many different parts of the genome (Key, 1968; Barton, 1979; Barton and Hewitt, 1985) (**Table 2**). The comparison of hybrid index to heterozygosity reinforces this; first-generation hybrids are completely absent in our sample, and F2s are rare. Most of the hybrids are backcrosses and advanced-generation hybrids (**Figure 6**). These trends are all consistent with previous high F_{ST} estimates, confirming the species are highly differentiated (Royer et al., 2016). The inability to identify particular targets or agents of selection is a predictable side effect of this phenomenon.

We found higher genetic diversity in *Y. brevifolia*, based on higher rates of allele fixation in *Y. jaegeriana* across all nuclear markers (i.e., higher minor allele frequencies in *Y. brevifolia*) (**Figures 4A**, **5**). This could be caused by introgression into *Y. brevifolia*, differences in effective population size, or differential maintenance of ancestral polymorphism. Previous population genetic work using microsatellites has suggested gene flow into *Y. brevifolia* (Starr et al., 2013), and we see evidence of introgression in the same direction in the clines—the few asymmetrical (model 3) clines (19% of the high-F_{ST} SNP pool) show introgression into *Y. brevifolia* (**Supplementary Table S3**), in the same direction that moth movement and population genetic data suggest gene flow should move (Smith et al., 2009; Starr et al., 2013).

The exception to the overall coincidence of clines, and perhaps our most surprising result, is the chloroplast cline center that is offset from nuclear clines opposite the direction we predicted (**Figure 4** and **Table 2**). Cytonuclear discordance is generally common in hybrid zones, but diagnosing the cause is challenging (Toews and Brelsford, 2012). The three basic mechanisms of all short-term evolution are the same ones that can explain a mismatch between cytoplasmic and nuclear

genomes in hybrid zones: natural selection, drift, or gene flow (Arnold, 1993).

If natural selection is causing the introgression of *Y. jaegeriana* chloroplasts into *Y. brevifolia*, introgression could be due to the *Y. jaegeriana* haplotype being inherently superior, the *Y. jaegeriana* haplotype being better adapted locally (e.g., Sambatti et al., 2008; Campbell et al., 2010), or cytonuclear interactions in hybrids with cytoplasm from one parent and nuclear genome from the other favoring the spread of the *Y. jaegeriana* haplotype. We have no data to test global superiority of one cytoplasmic genome in this system. Previous niche modeling has found little or no difference in the ecological niches of *Y. brevifolia* and *Y. jaegeriana* in and around the hybrid zone (Godsoe et al., 2009), so we have little support for local adaptation favoring one haplotype on the east side of the hybrid zone and another on the west. The last possible source of selection, cytonuclear interactions, commonly affect fitness (Burke and Arnold, 2001; Toews and Brelsford, 2012; Dobler et al., 2014), particularly in hybrids (Dobler et al., 2014), where they frequently appear in the form of cytoplasmic male sterility (CMS) (Frank, 1989). If interactions between the *Y. jaegeriana* cytoplasmic genome and the *Y. brevifolia* nuclear genome result in reduced male fitness, it could cause increased female fitness in hybrid plants with the *Y. jaegeriana* cytoplasmic genome, leading to the spread of that haplotype (Tsitrone et al., 2003). Such patterns of chloroplast capture have been shown to be widespread in plants (Rieseberg and Soltis, 1991; Bock et al., 2014) and would be sufficient to create the pattern we see in the Joshua tree hybrid zone.

Either drift or gene flow can also contribute to cytonuclear discordance, but neither seems likely to be responsible for the offset centers in the Joshua tree hybrid zone. Drift can shape clines of loci not under strong selection, as evidenced by the wider range of centers in our randomly selected set of low-F_{ST} SNP clines (centers range from −6573 to 2945, with 11% having centers farther from the center of the hybrid zone than the chloroplast cline; **Table 2** and **Supplementary Table S2**). Cytoplasmic genomes, with effective population sizes 1/4 that of the nuclear genome (Birky et al., 1983), are particularly prone to drift (Hudson and Turelli, 2003). However, because we have evidence from ecological and population genetic data that the chloroplast introgression in the Joshua tree hybrid zone is happening in spite of nuclear gene flow that should oppose it (compare our result with Smith et al., 2009 and Starr et al., 2013), we conclude selection is most likely responsible for the western displacement of the chloroplast cline center.

The cytonuclear genotypes present in the hybrid zone suggest cytonuclear incompatibilities may well play a role. With the large number of advanced hybrids and backcrosses present (**Figure 6**), the hybrid swarm should create opportunities for nuclear and cytoplasmic parental genomes to become dissociated, allowing chloroplast introgression to potentially happen in either direction. In the absence of selection, observed pollinator preferences would still push chloroplast introgression into *Y. jaegeriana* rather than what we observe. Instead, we find only partial disassociation of chloroplast and nuclear genomes. Opposite our prediction, most hybrids (∼80%) carry the *Y. jaegeriana* chloroplast haplotype. In addition, while all trees with genomic compositions consistent with backcrossing to *Y. jaegeriana* (Qscore > 0.5) have *Y. jaegeriana* chloroplast haplotypes, those with Qscores consistent with backcrossing to *Y. brevifolia* are narrowly dominated by the *Y. jaegeriana* chloroplast haplotype as well (58%). While far from decisive, these data suggest a role for selection and reinforce that gene flow cannot explain the chloroplast cline.

The explicitly geographic approach of our analysis highlights a significant feature of the hybrid zone relevant to understanding gene flow that has been ignored in previous studies: the ancient, dry lakebed that causes a roughly 1km gap in the distribution of Joshua trees just west of, and roughly parallel to, the center of the hybrid zone (**Figure 3**). The very low density of trees in this zone (there are a few, unsampled, individuals scattered throughout it) could be caused by sandier or saltier soil resulting in moisture levels too low for Joshua trees to tolerate. Across their range, Joshua trees are typically absent from pluvial lakebeds (C. Smith, personal observation), which are often dominated by *Atriplex canescens*, a highly salt tolerant species of salt bush that is widespread throughout western North America (Glenn et al., 1996). In fact, the entire hybrid zone is located east of the lakebed, with the nearest individuals west of the gap considered part of a population categorized as *Y. brevifolia* rather than hybrid (Starr et al., 2013), and all RADseq-genotyped individuals having Qscores below 0.06. This suggests that gene flow across the gap is relatively rare. That would be consistent with what we know about gene flow in Joshua trees; estimates of seed dispersal are around 30.0 ± 16.8m from the maternal tree (Vander Wall et al., 2006), and measurements of movement in other species of *Tegeticula* estimate moths move less than 50 m (Marr et al., 2000); as mentioned above, preliminary work in Joshua tree *Tegeticula* shows individual moths move an average of ∼30 m (31.79 ± 34.17 m; Autumn and Smith, unpublished data) (although this is likely an underestimate, Smith pers. comm.). This is roughly similar to estimates of *Yucca filamentosa* pollen dispersal by its moth pollinator, *Tegeticula yuccasella*: mean 4.66 ± 10.23 m (Marr et al., 2000). Occasional long-distance dispersal events certainly occur, but at a scale of >30x the mean, they may be quite rare.

It is likely that this area of low population density is the feature anchoring the hybrid zone in its current location. Movement of hybrid zones is generally a common feature; although many tension zones are currently stable, others have been shown to move (Buggs, 2007; Wielstra et al., 2017; van Riemsdijk et al., 2019). Directional hybrid zone movement can be in response to asymmetric gene flow, because one species is favored by natural selection, or in response to differences in population density (Buggs, 2007). Movement toward areas of low population density has long been thought to be a predominant driver of location and a consistent feature of tension zones (Barton, 1979; Barton and Hewitt, 1985). After being pulled toward an area of low population density, a hybrid zone will stop altogether and be held stationary if it reaches an uninhabitable or extreme low density region (Barton and Hewitt, 1985). The distribution gap formed by the dry lakebed west of the Joshua tree hybrid zone meets these criteria.

If the Joshua tree hybrid zone originally formed substantially east of its current location and subsequently moved in response to low density to the west, we would expect to see asymmetric clines with longer tails to the east across the genome, signaling greater introgression where the center of the hybrid zone used to be (Barton and Hewitt, 1985; Currat et al., 2008; Wielstra et al., 2017; van Riemsdijk et al., 2019). For clines with neutral variation, the centers should also be offset to the east, including the cytoplasmic genome (assuming selection is not acting on it, which is obviously a big assumption—see above) (Wielstra et al., 2017). This signal of introgression is expected to linger indefinitely (Currat et al., 2008). In fact, we see none of these features in the Joshua tree hybrid zone; all nuclear clines, including those for loci not under selection, share roughly the same center. The most likely history for the Joshua tree hybrid zone is what Barton (1979) believed to be the predominant model—the hybrid zone barely moved from the point where it first arose before being anchored long-term as a tension zone next to an area of low population density, the ancient lakebed on the west side of Tikaboo Valley.

CONCLUSION

We find that the patterns we expected to see based on previous ecological and population genetic work in and around the hybrid zone were not fully supported: the Joshua tree hybrid zone appears to be dominated by advanced-generation hybrids, is most strongly shaped by selection against early-generation hybrids, and displays a chloroplast cline displaced to the west (opposite our prediction). Evidence for direct phenotypic selection from pollinators is weak by comparison. These results suggest several new clear avenues for further work. First, a focus on how pollinators interact with advanced-generation hybrids (pollinator choice and performance). Previous work focused on reproductive isolation, comparing performance of pollinators on the opposite host. In fact, it seems that the formation of F1 hybrids is quite rare, so the movement of pollen between advanced generation hybrids (particularly backcrossed ones) and parentals is much more relevant to understanding gene flow in the current hybrid zone. In addition, the clear and unexpected pattern of chloroplast haplotype displacement to the west sets up several testable alternate hypotheses. The work described above on pollinator behavior may reveal pollen movement patterns that differ from those described in crosses between pure parental plants, consistent with chloroplast introgression. Additional work quantifying investment in male and female reproduction by hybrids with different chloroplast haplotypes (ovule and pollen counts) should reveal whether cytonuclear incompatibility is lowering overall fitness in hybrids with the *brevifolia* chloroplast haplotype, or increasing female fitness of those with the *jaegeriana* chloroplast haplotype. Finally, fully understanding the barriers contributing to reproductive isolation in Joshua trees will require going beyond population genomic methods to experimental approaches quantifying each barrier, prezygotic and postzygotic, to reproduction across these species (Sobel and Chen, 2014).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

AR contributed to design, performed the analyses, and wrote the manuscript. JW-H collected and genotyped moths. CS provided phenotypic and genomic data, funding, and input into writing and interpretation.

FUNDING

This work was supported by a grant from the National Science Foundation to CIS (DEB 1253849).

ACKNOWLEDGMENTS

We gratefully acknowledge Sean Stankowski's assistance with analysis and interpretation. Furey Stirrat developed the cpDNA primers. Ramona Flatz performed the microsatellite and RADseq lab work. Hunter Smith genotyped the cpDNA markers. The Bureau of Land Management Caliente Field office provided research and collecting permits for the work. Todd Esque of the US Geological Survey provided logistic support. We gratefully acknowledge these contributions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00640/full#supplementary-material

REFERENCES

Althoff, D. M., Segraves, K. A., and Johnson, M. T. J. (2014). Testing for coevolutionary diversification: linking pattern with process. *Trends Ecol. Evol.* 29, 82–89. doi: 10.1016/j.tree.2013.11.003

Althoff, D. M., Segraves, K. A., Smith, C. I., Leebens-Mack, J., and Pellmyr, O. (2012). Geographic isolation trumps coevolution as a driver of yucca and yucca moth diversification. *Mol. Phylogenet. Evol.* 62, 898–906. doi: 10.1016/j.ympev.2011.11.024

Armbruster, W. S., and Muchhala, N. (2008). Associations between floral specialization and species diversity: cause, effect, or correlation? *Evol. Ecol.* 23, 159–179. doi: 10.1007/s10682-008-9259-z

Arnold, J. (1993). Cytonuclear disequilibria in hybrid zones. *Annu. Rev. Ecol. Syst.* 24, 521–553. doi: 10.1146/annurev.es.24.110193.002513

Baack, E., Melo, M. C., Rieseberg, L. H., and Ortiz-Barrientos, D. (2015). The origins of reproductive isolation in plants. *New Phytol.* 207, 968–984. doi: 10.1111/nph.13424

Barton, N., and Hewitt, G. (1985). Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* 16, 113–148. doi: 10.1146/annurev.es.16.110185.000553

Barton, N. H. (1979). The dynamics of hybrid zones. *Heredity* 43, 341–359. doi: 10.1038/hdy.1979.87

Barton, N. H., and Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature* 341, 497–503. doi: 10.1038/341497a0

Birky, C. W., Maruyama, T., and Fuerst, P. (1983). An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* 103, 513–527.

Bock, D. G., Andrew, R. L., and Rieseberg, L. H. (2014). On the adaptive value of cytoplasmic genomes in plants. *Mol. Ecol.* 23, 4899–4911. doi: 10.1111/mec.12920

Buggs, R. J. A. (2007). Empirical study of hybrid zone movement. *Heredity* 99, 301–312. doi: 10.1038/sj.hdy.6800997

Burke, J. M., and Arnold, M. L. (2001). Genetics and the fitness of hybrids. *Annu. Rev. Genet.* 35, 31–52. doi: 10.1146/annurev.genet.35.102401.085719

Campbell, D. R., Wu, C. A., and Travers, S. E. (2010). Photosynthetic and growth responses of reciprocal hybrids to variation in water and nitrogen availability. *Am. J. Bot.* 97, 925–933. doi: 10.3732/ajb.0900387

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1111/mec.12354

Cole, W. S., James, A. S., and Smith, C. I. (2017). First recorded observations of pollination and oviposition behavior in *Tegeticula antithetica* (Lepidoptera: Prodoxidae) suggest a functional basis for coevolution with joshua tree (yucca) hosts. *Ann. Entomol. Soc. Am.* 110, 390–397. doi: 10.1093/aesa/sax037

Coyne, J., and Orr, H. (1989). Patterns of speciation in drosophila. *Evolution* 43, 362–381. doi: 10.1111/j.1558-5646.1989.tb04233.x

Coyne, J. A., and Orr, H. A. (1997). "'Patterns of speciation in *Drosophila*'" revisited. *Evolution* 51, 295–303. doi: 10.1111/j.1558-5646.1997.tb02412.x

Coyne, J. A., and Orr, H. A. (2004). *Speciation*. New York, NY: W.H. Freeman.

Currat, M., Ruedi, M., Petit, R. J., and Excoffier, L. (2008). The hidden side of invasions: massive introgression by local genes. *Evolution* 62, 1908–1920. doi: 10.1111/j.1558-5646.2008.00413.x

Derryberry, E. P., Derryberry, G. E., Maley, J. M., and Brumfield, R. T. (2014). hzar: hybrid zone analysis using an R software package. *Mol. Ecol. Resour.* 14, 652–663. doi: 10.1111/1755-0998.12209

Dobler, R., Rogell, B., Budar, F., and Dowling, D. K. (2014). A meta-analysis of the strength and nature of cytoplasmic genetic effects. *J. Evol. Biol.* 27, 2021–2034. doi: 10.1111/jeb.12468

Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Ehrlich, P. R., and Raven, P. H. (1964). Butterflies and plants – A study in coevolution. *Evolution* 18, 586–608. doi: 10.1111/j.1558-5646.1964.tb01674.x

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Farrell, B. D. (1998). "Inordinate fondness". Explained: why are there so many beetles? *Science* 281, 555–559. doi: 10.1126/science.281.5376.555

Flatz, R., Yoder, J. B., Lee-Barnes, E., and Smith, C. I. (2011). Characterization of microsatellite loci in Yucca brevifolia (Agavaceae) and cross-amplification in related species. *Am. J. Bot.* 98, e67–e69. doi: 10.3732/ajb.1000468

Frank, S. A. (1989). The evolutionary dynamics of cytoplasmic male sterility. *Am. Nat.* 133, 345–376. doi: 10.1086/284923

Glenn, E., Pfister, R., Brown, J. J., Thompson, T. L., and O'Leary, J. (1996). Na and k accumulation and salt tolerance of Atriplex canescens (Chenopodiaceae) genotypes. *Am. J. Bot.* 83, 997–1005. doi: 10.1002/j.1537-2197.1996.tb12796.x

Godsoe, W., Strand, E., Smith, C. I., Yoder, J. B., Esque, T. C., and Pellmyr, O. (2009). Divergence in an obligate mutualism is not explained by divergent climatic factors. *New Phytol.* 183, 589–599. doi: 10.1111/j.1469-8137.2009.02942.x

Godsoe, W., Yoder, J. B., Smith, C. I., Drummond, C. S., and Pellmyr, O. (2010). Absence of population-level phenotype matching in an obligate pollination mutualism. *J. Evol. Biol.* 23, 2739–2746. doi: 10.1111/j.1420-9101.2010.02120.x

Godsoe, W., Yoder, J. B., Smith, C. I., and Pellmyr, O. (2008). Coevolution and divergence in the Joshua tree/yucca moth mutualism. *Am. Nat.* 171, 816–823. doi: 10.1086/587757

Hembry, D. H., Yoder, J. B., and Goodman, K. R. (2014). Coevolution and the diversification of Life. *Am. Nat.* 184, 425–438. doi: 10.1086/677928

Hewitt, G. M. (1988). Hybrid zones-natural laboratories for evolutionary studies. *Trends Ecol. Evol.* 3, 158–167. doi: 10.1016/0169-5347(88)90033-X

Hudson, R. R., and Turelli, M. (2003). Stochasticity overrules the "three-times rule": genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution* 57, 182–190. doi: 10.1111/j.0014-3820.2003.tb00229.x

Janzen, D. H. (1980). When is it coevolution. *Evolution* 34, 611–612. doi: 10.1111/j.1558-5646.1980.tb04849.x

Key, K. H. L. (1968). The concept of stasipatric speciation. *Syst. Biol.* 17, 14–22. doi: 10.1093/sysbio/17.1.14

Kiester, A. R., Lande, R., and Schemske, D. W. (1984). Models of coevolution and speciation in plants and their pollinators. *Am. Nat.* 124, 220–243. doi: 10.1086/284265

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* 15, 1179–1191. doi: 10.1111/1755-0998.12387

Leebens-Mack, J., Pellmyr, O., and Brock, M. (1998). Host specificity and the genetic structure of two yucca moth species in a yucca hybrid zone. *Evolution* 52, 1376–1382. doi: 10.1111/j.1558-5646.1998.tb02019.x

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., et al. (2017a). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol. Ecol. Resour.* 17, 142–152. doi: 10.1111/1755-0998.12635

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., et al. (2017b). Responsible RAD: striving for best practices in population genomic studies of adaptation. *Mol. Ecol. Resour.* 17, 366–369. doi: 10.1111/1755-0998.12677

Lowry, D. B., Rockwood, R. C., and Willis, J. H. (2008). Ecological reproductive isolation of coast and inland races of mimulus guttatus. *Evolution* 62, 2196–2214. doi: 10.1111/j.1558-5646.2008.00457.x

Mallet, J., Barton, N., Lamas, G., Santisteban, J., Muedas, M., and Eeley, H. (1990). Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in heliconius hybrid zones. *Genetics* 124, 921–936.

Marquis, R. J., Salazar, D., Baer, C., Reinhardt, J., Priest, G., and Barnett, K. (2016). Ode to ehrlich and raven or how herbivorous insects might drive plant speciation. *Ecology* 97, 2939–2951. doi: 10.1002/ecy.1534

Marr, D. L., Leebens-Mack, J., Elms, L., and Pellmyr, O. (2000). Pollen dispersal in Yucca filamentosa (Agavaceae): the paradox of self-pollination behavior by *Tegeticula yuccasella* (Prodoxidae). *Am. J. Bot.* 87, 670–677. doi: 10.2307/2656853

Mittelbach, G. G., Schemske, D. W., Cornell, H. V., Allen, A. P., Brown, J. M., Bush, M. B., et al. (2007). Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecol. Lett.* 10, 315–331. doi: 10.1111/j.1461-0248.2007.01020.x

Mitter, C., Farrell, B., and Futuyma, D. J. (1991). Phylogenetic studies of insect-plant interactions: insights into the genesis of diversity. *Trends Ecol. Evol.* 6, 290–293. doi: 10.1016/0169-5347(91)90007-K

Mitter, C., Farrell, B., and Wiegmann, B. (1988). The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *Am. Nat.* 132, 107–128. doi: 10.1086/284840

Moe, A. M., and Weiblen, G. D. (2012). Pollinator-mediated reproductive isolation among dioecious fig species (Ficus, Moraceae). *Evol. Int. J. Org. Evol.* 66, 3710–3721. doi: 10.1111/j.1558-5646.2012.01727.x

Nielsen, E. E. G., Bach, L. A., and Kotlicki, P. (2006). HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. *Mol. Ecol. Notes* 6, 971–973. doi: 10.1111/j.1471-8286.2006.01433.x

Nosil, P. (2012). *Ecological Speciation*. New York, NY: Oxford University Press, Inc. doi: 10.1093/acprof:osobl/9780199587100.001.0001

Pellmyr, O., and Segraves, K. A. (2003). Pollinator divergence within an obligate mutualism: two yucca moth species (Lepidoptera; Prodoxidae: Tegeticula) on the joshua tree (*Yucca brevifolia*; Agavaceae). *Ann. Entomol. Soc. Am.* 96, 716–722. doi: 10.1603/0013-8746(2003)096[0716:PDWAOM]2.0.CO;2

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

Rieseberg, L. H., and Soltis, D. E. (1991). Phylogenetic consequences of cytoplasmic gene flow in plants. *Evol. Trends Plants* 5, 65–84.

Rowlands, P. G. (1978). *The Vegetation Dynamics of the Joshua tree (Yucca brevifolia Engelm.) in the Southwestern United States of America*. Riverside: University of California.

Royer, A. M., Streisfeld, M. A., and Smith, C. I. (2016). Population genomics of divergence within an obligate pollination mutualism: selection maintains differences between Joshua tree species. *Am. J. Bot.* 103, 1730–1741. doi: 10.3732/ajb.1600069

Sambatti, J. B. M., Ortiz-Barrientos, D., Baack, E. J., and Rieseberg, L. H. (2008). Ecological selection maintains cytonuclear incompatibilities in hybridizing sunflowers. *Ecol. Lett.* 11, 1082–1091. doi: 10.1111/j.1461-0248.2008.01224.x

Schluter, D. (2000). *The Ecology of Adaptive Radiation*. New York, NY: Oxford University Press, Inc.

Smith, C. I., Drummond, C. S., Godsoe, W., Yoder, J. B., and Pellmyr, O. (2009). Host specificity and reproductive success of yucca moths (Tegeticula spp. Lepidoptera: Prodoxidae) mirror patterns of gene flow between host plant varieties of the Joshua tree (*Yucca brevifolia*: Agavaceae). *Mol. Ecol.* 18, 5218–5229. doi: 10.1111/j.1365-294X.2009.04428.x

Smith, C. I., Godsoe, W. K. W., Tank, S., Yoder, J. B., and Pellmyr, O. (2008). Distinguishing coevolution from covicariance in an obligate pollination mutualism: asynchronous divergence in joshua tree and its pollinators. *Evolution* 62, 2676–2687. doi: 10.1111/j.1558-5646.2008.00500.x

Smith, J. W., and Benkman, C. W. (2007). A coevolutionary arms race causes ecological speciation in crossbills. *Am. Nat.* 169, 455–465. doi: 10.1086/511961

Sobel, J. M., and Chen, G. F. (2014). Unification of methods for estimating the strength of reproductive isolation. *Evolution* 68, 1511–1522. doi: 10.1111/evo.12362

Sobel, J. M., Chen, G. F., Watt, L. R., and Schemske, D. W. (2010). The biology of speciation. *Evolution* 64, 295–315. doi: 10.1111/j.1558-5646.2009.00877.x

Stankowski, S., Sobel, J. M., and Streisfeld, M. A. (2017). Geographic cline analysis as a tool for studying genome-wide variation: a case study of pollinator-mediated divergence in a monkeyflower. *Mol. Ecol.* 26, 107–122. doi: 10.1111/mec.13645

Starr, T. N., Gadek, K. E., Yoder, J. B., Flatz, R., and Smith, C. I. (2013). Asymmetric hybridization and gene flow between Joshua trees (Agavaceae: Yucca) reflect differences in pollinator host specificity. *Mol. Ecol.* 22, 437–449. doi: 10.1111/mec.12124

Toews, D. P. L., and Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Mol. Ecol.* 21, 3907–3930. doi: 10.1111/j.1365-294X.2012.05664.x

Trelease, W. (1893). Further studies of yuccas and their pollination. *Mo. Bot. Gard. Annu. Rep.* 1893, 181–226. doi: 10.2307/2992178

Tsitrone, A., Kirkpatrick, M., and Levin, D. A. (2003). A Model for chloroplast capture. *Evolution* 57, 1776–1782. doi: 10.1111/j.0014-3820.2003.tb00585.x

van Riemsdijk, I., Butlin, R. K., Wielstra, B., and Arntzen, J. W. (2019). Testing an hypothesis of hybrid zone movement for toads in France. *Mol. Ecol.* 28, 1070–1083. doi: 10.1111/mec.15005

Vander Wall, S. B., Esque, T., Haines, D., Garnett, M., and Waitman, B. A. (2006). Joshua tree (*Yucca brevifolia*) seeds are dispersed by seed-caching rodents. *Écoscience* 13, 539–543. doi: 10.2980/1195-6860(2006)13[539:JTYBSA]2.0.CO;2

Wielstra, B., Burke, T., Butlin, R. K., Avcı, A., Üzüm, N., Bozkurt, E., et al. (2017). A genomic footprint of hybrid zone movement in crested newts. *Evol. Lett.* 1, 93–101. doi: 10.1002/evl3.9

Yoder, J. B., and Nuismer, S. L. (2010). When does coevolution promote diversification? *Am. Nat.* 176, 802–817. doi: 10.1086/657048

Yoder, J. B., Smith, C. I., Rowley, D. J., Flatz, R., Godsoe, W., Drummond, C., et al. (2013). Effects of gene flow on phenotype matching between two varieties of Joshua tree (*Yucca brevifolia*; Agavaceae) and their pollinators. *J. Evol. Biol.* 26, 1220–1233. doi: 10.1111/jeb.12134

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 27 May 2020
doi: 10.3389/fpls.2020.00573

Mayahuelin, a Type I Ribosome Inactivating Protein: Characterization, Evolution, and Utilization in Phylogenetic Analyses of *Agave*

*Fernando Lledías[1†], Jesús Gutiérrez[2†], Aída Martínez-Hernández[3], Abisaí García-Mendoza[2], Eric Sosa[2], Felipe Hernández-Bermúdez[2], Tzvetanka D. Dinkova[4], Sandi Reyes[2], Gladys I. Cassab[1] and Jorge Nieto-Sotelo[2]**

[1] Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Mexico, [2] Jardín Botánico, Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City, Mexico, [3] Colegio de Postgraduados, Campus Campeche, Campeche, Mexico, [4] Departamento de Bioquímica, Facultad de Química, Universidad Nacional Autónoma de México, Mexico City, Mexico

Agaves resist extreme heat and drought. In *A. tequilana* var. *azul*, the central spike of the rosette -containing the shoot apical meristem and folded leaves in early stages of development- is remarkably heat tolerant. We found that the most abundant protein in this organ is a 27 kDa protein. This protein was named mayahuelin to honor Mayáhuel, the agave goddess in the Aztec pantheon. LC-MS/MS analyses identified mayahuelin as a type I RIP (**R**ibosome **I**nactivating **P**rotein). In addition to the spike, mayahuelin was expressed in the peduncle and in seeds, whereas in mature leaves, anthers, filaments, pistils, and tepals was absent. Anti-mayahuelin antibody raised against the *A. tequilana* var. *azul* protein revealed strong signals in spike leaves of *A. angustifolia*, *A. bracteosa*, *A. rhodacantha*, and *A. vilmoriniana*, and moderate signals in *A. isthmensis*, *A. kerchovei*, *A. striata* ssp. *falcata*, and *A. titanota*, indicating conservation at the protein level throughout the *Agave* genus. As in charybdin, a type I RIP characterized in *Drimia maritima*, mayahuelin from *A. tequilana* var. *azul* contains a natural aa substitution (Y76D) in one out of four aa comprising the active site. The RIP gene family in *A. tequilana* var. *azul* consists of at least 12 genes and *Mayahuelin* is the only member encoding active site substitutions. Unlike canonical plant RIPs, expression of *Mayahuelin* gene in *S. cerevisiae* did not compromise growth. The inhibitory activity of the purified protein on a wheat germ *in vitro* translation system was moderate. *Mayahuelin* orthologs from other *Agave* species displayed one of six alleles at Y76: (Y/Y, D/D, S/S, Y/D, Y/S, D/S) and proved to be useful markers for phylogenetic analysis. Homozygous alleles were more frequent in wild accessions whereas heterozygous alleles were more frequent in cultivars. *Mayahuelin* sequences from different wild populations of *A. angustifolia* and *A. rhodacantha* allowed the

OPEN ACCESS

Edited by:
Luis Enrique Eguiarte, National Autonomous University of Mexico, Mexico

Reviewed by:
Laura Trejo, National Autonomous University of Mexico Tlaxcala, Mexico
Benjamín Rodríguez-Garay, CONACYT Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco (CIATEJ), Mexico
June Simpson, Centro de Investigación y Estudios Avanzados, Instituto Politécnico Nacional de México, Mexico

***Correspondence:**
Jorge Nieto-Sotelo
jorge.nieto@ib.unam.mx

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *14 December 2019*
Accepted: *17 April 2020*
Published: *27 May 2020*

Citation:
Lledías F, Gutiérrez J, Martínez-Hernández A, García-Mendoza A, Sosa E, Hernández-Bermúdez F, Dinkova TD, Reyes S, Cassab GI and Nieto-Sotelo J (2020) Mayahuelin, a Type I Ribosome Inactivating Protein: Characterization, Evolution, and Utilization in Phylogenetic Analyses of Agave. Front. Plant Sci. 11:573. doi: 10.3389/fpls.2020.00573

identification of accessions closely related to *azul, manso, sigüín, mano larga, and bermejo* varieties of *A. tequilana* and var. *espadín* of *A. angustifolia*. Four *A. rhodacantha* accessions and *A. angustifolia* var. *espadín* were closer relatives of *A. tequilana* var. *azul* than *A. angustifolia* wild accessions or other *A. tequilana* varieties.

Keywords: RIP (ribosome inactivating protein), active site substitution, plant domestication, protein translation, disjoint distributions, agave evolution

INTRODUCTION

The *Agave* genus is a member of the Agavoideae subfamily within the Asparagaceae family of plants (The Angiosperm Phylogeny Group, 2009; Chase et al., 2009). The natural distribution of *Agave* encompasses the United States, Mexico, Central America, the Caribbean islands, and South America as far south as Paraguay (García-Mendoza, 1998). The *Agave* genus contains approximately 206 species; Mexico has the highest diversity of species (159, of which 119 are endemic) and it is considered its center of origin (Gentry, 1982; García-Mendoza, 1998; García-Mendoza and Chávez-Rendón, 2013). Most species in the genus are adapted to and play important ecological roles as part of dry ecosystems or arid microenvironments within mesic habitats. *Agave* species also are a food source for bats of the *Leptonycteris* genus that migrate long distances in Mexico and the Sonoran desert (Howell and Roth, 1981; Rojas-Martínez et al., 1999). The cultural importance of agaves in Mexico and the United States Southwest is enormous since pre-historical times to the present. More than 70 known traditional uses are documented for species in the genus (Castetter et al., 1938; Nobel, 1988; García-Mendoza, 1998). In addition, agaves show a great potential as bioenergy crops and as sources of bioactive compounds with anticancer, antioxidant, antimicrobial, antifungal, pre-biotic, and anti-inflammatory properties (Barreto et al., 2010; Escamilla-Treviño, 2011; Simpson et al., 2011; Santos-Zea et al., 2012; Hernández-Valdepeña et al., 2016).

The morphological and physiological adaptations of agaves to high temperature and aridity include succulency of leaves and stems, long and narrow leaves, rosettes sitting near the soil that facilitate nocturnal water collection from dew that is funneled to the base of the plant, shallow roots, thick cuticles, low stomatal densities, and CAM metabolism (Nobel, 1988; Martorell and Ezcurra, 2007; Luján et al., 2009).

In *A. tequilana* var. *azul* the structure with the highest heat resistance is the spike (Luján et al., 2009) which is composed by several folded leaves, located at the center of the rosette, that surround and protect the shoot apical meristem. Heat resistance in the spike is mostly due to its higher levels of heat shock proteins (HSP), higher stomatal density, and greater capacity for leaf cooling relative to more mature sectors of the rosette (Luján et al., 2009). During the progress of the previous study, we detected a 27 kDa protein as the most abundant protein in the spike leaves; we further studied it suspecting to be an HSP. We named this protein mayahuelin after *Mayáhuel*, the agave goddess, a prominent member of the Mesoamerican pantheon, and venerated by ancient náhuatl-speaking cultures (i.e. aztecs, mexicas, etc.) (Tena Martínez, 2002). Amino acid and nucleotide sequence analyses revealed that mayahuelin is a Type I ribosome inactivating protein (RIP) and represents the first RIP described within the subfamily Agavoideae.

Ribosome inactivating proteins are a family of cytotoxic polypeptides with the capacity to bind ribosome large subunits. This interaction causes the irreversible blockage of protein synthesis (De Virgilio et al., 2010). RIPs are found in bacteria, fungi and especially in plants, where they have been described in 22 families representing 14 angiosperm orders (Puri et al., 2012; Di Maro et al., 2014). Most RIPs have a N-glicosidase activity [EC. 3.2.2.22] that removes an adenine residue from the highly conserved rRNA structure known as "Sarcin-Ricin Loop" (SRL). This structure is a central element for the interaction between the ribosome and the Elongation Factor II (EFII). SRL takes its name from ricin, a RIP from *Ricinus communis* seeds that depurinates (A4324) rat 28S rRNA, and from sarcin, a RIP from *Aspergillus giganteus* that breaks the phosphodiester bond between the G4325-A4326 residues of the 28S rRNA (Szewcsak and Moore, 1995; Spackova and Sponer, 2006). Despite SRL structural conservation, RIP specificity for ribosomes shows clear differences (May et al., 2013) while ricin severely damages mammalian and yeast ribosomes, its effects on plants are minimum and null for *E. coli*. In contrast, the "Pokeweed Antiviral Protein" or PAP (a type I RIP from *Phytolacca americana*) equally depurinates plant, bacterial, and animal ribosomes (Peumans et al., 2001). Apart from rRNA, many RIPs depurinate DNA, adenine polynucleotides, and different viral RNAs. Because of this multi-substrate activity, RIPs are also known as "Polynucleotide-Adenosine Glicosidases" (Barbieri et al., 1997).

Ribosome inactivating proteins are classified as type I, formed by one chain named A (MW 30 kDa), and type II, heterodimers between an A chain (type I-like) and a B chain with lectin properties. Both chains have a MW from 56 to 65 kDa (Stirpe and Batelli, 2006). A 60 kDa type III RIPs (or pro-RIPs) with a type I N-terminus, and a C-end with unknown function has been described as an inactive precursor that requires processing to obtain a functional RIP (Puri et al., 2012). Type II RIPs are highly toxic on account of the B chain, that promotes entry of the A chain into the cell (Stirpe, 2013). In some plant species RIPs are widely expressed in different tissues (e.g. saporin from *Saponaria oficinalis* is found in leaves, roots and seeds), while in others show tissue-specific location (e.g. ricin from *R. communis* found in seeds only).

Ribosome inactivating protein first enzymatic mechanisms were elucidated in ricin A chain, where the catalytic site residues responsible for SRL depurination were identified as Y80, Y123, E177, and R180 (Kim and Robertus, 1992). Catalytic site amino

acids and their tertiary structure are highly conserved in at least 10 published RIP crystal structures (Peumans et al., 2001). Individual catalytic site amino acid substitutions have different impact on enzymatic activity of RIPs. The R180H substitution in ricin rearranges the active site and decreases activity 500-fold (Day et al., 1996) whereas substitutions in active site tyrosine residues, Y80S or Y123S, in charge of adenine-substrate stabilization, decrease depurination activity by 160- and 70-fold, respectively, (Ready et al., 1991; Kim and Robertus, 1992).

Ribosome inactivating protein expression increases under different stress conditions, during senescence, and in response to microbial and viral infections (Stirpe and Batelli, 2006). Under osmotic and heat stress, the translation inhibitory activity and DNA deadenylation activities of RIPs increase in *Hura crepitans* and *Phytolacca americana* (Stirpe et al., 1996). Moreover, RIP overexpression increases drought and salt tolerance in rice, this gain attributed to the up-regulation of stress genes through unknown mechanisms (Jiang et al., 2012). Accordingly, exogenous administration of purified RIP from *P. americana* on tobacco leaves protects them from tobacco mosaic virus infection and increases the levels of antioxidant enzymes (Zhu et al., 2016).

Several *Agave* molecular phylogenies have been proposed based on nucleotide sequences derived from *rbc*L, *trn*L+*trn*L-*trn*F, *ndh*F, and *ITS* markers (Bogler and Simpson, 1996; Bogler et al., 2006; Good-Avila et al., 2006; Heyduk et al., 2016; Huang et al., 2018) and have been very useful to resolve different genera within Agavoideae (*Agave*, *Beschorneria*, *Furcraea*, *Hesperaloe*, *Hesperoyucca*, *Manfreda*, *Polianthes, Prochnyanthes,* and *Yucca*). However, these markers offer very poor intrageneric resolution. In contrast, AFLP, SSR, and SSAP markers offer very high genetic resolutions, allowing population genetic analyses in different *Agave* species (Gil-Vega et al., 2006; Bousios et al., 2007; Trejo et al., 2018; Rivera-Lugo et al., 2018). Because false positive and false negative data can be obtained using fragment length as an estimate for genetic identity between individuals, especially when dealing with genetically distant species, methods alternative to AFLP, SSR, and SSAP have been proposed to derive molecular phylogenies based on nucleotide sequence, such as transcriptomics, RNA/DNA hybrid enrichment, and phylogenomics, all of which are based on next-generation sequencing (Lemmon and Lemmon, 2013; Huang et al., 2018; Zhang et al., 2018; Fernández et al., 2018). However, they are very costly, time-consuming, and not very practical when large numbers of samples are studied. Moreover, the large number of genetic markers available with these methods does not warrant phylogenetic resolution, as shown by a recent study of CAM (Crassulacean acid metabolism) evolution in Agavoideae that used 272 CAM-related genes to derive a phylogeny of 60 Agavoideae species (Heyduk et al., 2016) giving good resolution at the intergeneric level, but poor resolution at the intrageneric level, displaying hard polytomies in *Agave sensu lato*. Thus, nuclear markers that allow sufficient genetic resolution at the species level are still needed for plant phylogenetic analysis. *A. tequilana* var. *azul*, the exclusive cultivar approved for tequila production according to Mexican law, is cultivated asexually to maintain its varietal qualities; as a result its populations have extremely low genetic diversity (Gil-Vega et al., 2001). Hence, it is critical to identify the wild populations that originated this variety and other clonally propagated cultivars used for. tequila and mescal production. This knowledge is fundamental for both their conservation and improvement as they represent potential sources of genetic variation. Phylogenetic methods are very powerful tools to achieve these goals.

Here, we present results on the isolation, characterization, expression, and evolution of mayahuelin and its use as a phylogenetic marker within the *Agave* genus. We report the expression analyses of mayahuelin in different plant organs. Interestingly, a highly conserved tyrosine (Y76) in the active site of all RIPs (corresponding to Y80 in ricin) was naturally mutated to aspartate or serine in mayahuelin from *A. tequilana* var. *azul* and in other species within the *Agave* genus. Mayahuelin from *A. tequilana* var. *azul* containing the Y76D substitution was not toxic *in vivo* when expressed in yeast and only moderately toxic *in vitro*. Mayahuelin was immunodetected in the spike of several species of both the Littaea and Agave subgenera, indicating conservation at the protein level throughout the genus. Phylogenetic analyses using *Mayahuelin* ortholog sequences identified accessions of *A. rhodacantha* and *A. angustifolia* as close relatives of five *A. tequilana* and one *A. angustifolia* cultivars. Several accessions of *A. rhodacantha* and *A. angustifolia* intermixed in different *Agave* clades and were more genetically distant from these cultivars. We discuss the implications of the Y76 substitutions in terms of *Agave* as a natural resource and in domesticated plants.

MATERIALS AND METHODS

Plant Materials

Plant materials utilized in this work came from different sources: the National Collection of Agavaceae and Nolinaceae from Jardín Botánico, Instituto de Biología, Universidad Nacional Autónoma de México, in Mexico City; the agave collection of Jardín Botánico, Casa Sauza, in Tequila, Jalisco, Mexico, and from recent field trips made for this work (for a detailed list of plants studied and their sources see **Supplementary Table S1** and **Supplementary Figure S1**).

Mayahuelin Native Protein Purification

Native mayahuelin was obtained from fresh *A. tequilana* var. *azul* spike leaves cut with scissors and pulverized with a mortar and pestle in liquid nitrogen. 1 mL of frozen tissue powder was transferred to 2 mL tubes and 1 mL extraction buffer (200 mM Tris pH 7.2, 20 mM NaCl, 0.5% (v/v) β-mercapthoethanol, 2 mM EDTA pH 8.0, and 10X Complete protease inhibitor cocktail [Roche]). Tissue was resuspended with a stainless steel spatula, while thawing for 30 s. Each tube was perforated at the bottom with a syringe needle and placed atop 15 mL conic tubes and centrifuged (6 000 × *g* for 10 min, 6°C) to recover the liquid phase, taking the advantage that *A. tequilana* natural fibers worked as a filter. Supernatants were recovered and two volumes of cold acetone were added. After 30 min on ice, tubes were centrifuged (14,000 × *g* for 10 min, 25°C) and supernatants were discarded. Pellets were air-dried (30 min) and resuspended in

200 mM Tris pH 8.8, 1% glycerol. Tubes were centrifuged again. Supernatants were recovered for their separation on native gels (see **Supplementary Information**).

Mass Spectrometry

A published protocol for preparation of total protein extracts from *Agave*, separation by electrophoresis, and mass spectrometric analysis was followed (Lledías et al., 2017a,b). Samples recovered from "Little blue tank" (see **Supplemental Information**) traps were precipitated with methanol/chloroform, resuspended in 1X Laemmli sample buffer and directly loaded onto a stacking gel of a 12% polyacrylamide/SDS. The resulting band was excised for analysis. In-gel samples were chemically modified prior to mass spectrometry analysis. After reduction (dithiothreitol) and alkylation (iodoacetamide), samples were digested in gel with sequencing grade modified trypsin (Promega; Madison, WI, United States) in a solution of 50 mM ammonium bicarbonate pH 8.2 for 18 h at 37°C. Resultant peptides were desalted with Zip Tip C18 (Millipore; Billerica, MA, United States) and applied to a LC-MS system (Liquid Chromatography-Mass Spectrometry) composed by an EASY-nLC II nanoflow pump (Thermo Fisher Scientific; San Jose, CA, United States) coupled to a LTQ-Orbitrap Velos (Thermo Fisher Scientific; San Jose, CA, United States) mass spectrometer with a nano-electrospray ionization (ESI) source. The mass spectrometer was calibrated with a Calmix solution containing N-butylamine, caffeine, Met-Arg-Phe-Ala (MRFA) peptide, and Ultramark 1621. This solution was used to calibrate the LTQ Velos module with ion trap (IT) and Orbitrap FT (Fourier transform) mass detector on positive ionization ESI mode. N-butylamine (73.14 Da) was included to extend mass calibration to values less than *m*/*z*. Once calibrated, molecular mass accuracy at less than 5 ppm can be obtained. For LC, a 5%–85% gradient of solution B (water/acetonitrile, 0.1% formic acid) and solvent A (0.1% formic acid in water) was used for 160 min through a home-made capillary column (10 cm in length, ID 0.75 μm) made of TSP standard FS tubing with OD 363 μm (part no. TSP-075375BGB, Analytik, United States) packed with a C18-reversed phase silica gel (Jupiter 4 μm Proteo 90 Å, Phenomenex; Torrance, CA, United States) with a flux of 10 μL/min. For peptide fragmentation, Collision-Induced Dissociation (CID) and High-energy Collision Dissociation (HCD) methods were used with a resolution power (RP = m/FWHM) of 15,000 and selecting only 2+, 3+ and 4+ charged ions. A full scan of ions was performed with the Orbitrap analyser with a resolution power (RP = m/FWHM) of 60,000. For data acquisition, the positive ion mode was set. Capture and performance of fragmentation data were done according to the total ion scanning and predetermined charge with the following parameters: 2.0 (m/z) isolation width; collision energy, 35 arbitrary units; activation Q, 0.250; activation time, 10 ms; maximum injection time, 10 ms per micro-scanning. The automatic capture of data was done by ion dynamic exclusion: (i) exclusion list of 300 ions; (ii) pre-exclusion time of 30 s; and (iii) exclusion time of 70 s. Sequences obtained by electrospray LC-MS/MS were searched in.raw format with the Proteome Discoverer 1.4.1.14 (Thermo Fisher Scientific; San Jose; CA, United States) and the search engine Sequest HT. Since proteomic data in Agavoideae is lacking, we searched an EST library database from *A. tequilana* var. *azul* (Martínez-Hernández et al., 2010; Simpson et al., 2011). A minimal FDR (false discovery rate) of 0.01 and maximal FDR of 0.05, in addition to a decoy database, were used in the *Percolator* program. The maximum tolerance for molecular mass differences between theoretical and experimental values (precursor mass tolerance) was 20 ppm; tolerance for fragments obtained afer dissociation of precursor ion (fragment mass tolerance) was 0.6 Da. For automatic search mode, modification constants such as carbamido-methylation of cysteines (C) and variables such as methionine oxidation (M), asparagine (N) and glutamine (Q) deaminations were established.

N-terminal sequencing of mayahuelin, isolated by the native protein purification protocol described above, was accomplished by Edman degradation using an LF 3000 (Beckman Instruments, Irvine, CA, United States) automated protein sequencer coupled to a Beckman GoldHPLC system.

Other Methods for Protein Biochemical Analysis

Detailed protocols for native gel electrophoresis, native protein electroelution, polyclonal antibody production, SDS-PAGE analysis, and immunoblot analysis of mayahuelin are described in **Supplementary Information**.

Evaluation of the Effects of Mayahuelin on Protein Translation *in vitro*

The inhibitory effect of mayahuelin on protein translation was tested using luciferase as a reporter on a cell-free wheat germ protein synthesis system (cat. L4380, Promega). Mayahuelin was tested at different nM concentrations and, as a positive control, saporin [13.3 nM] was tested (cat. S9896, Sigma-Aldrich). Samples were preincubated at 25°C for 30 min with RNAsin, an RNAase inhibitor (cat. N2111, Promega). After preincubation, 50 ng of luciferase coding RNA were added and kept for an extra 1.5 h at 25°C. The reaction was stopped with 10 μL of 1X passive lysis buffer (cat. E1941, Promega). Luciferase reactant (50 μL) (cat. E1483, Promega) was added to 10 μL aliquots from each reaction, in triplicate. As a negative control, one reaction without protein and mRNA was included. Samples were analyzed on a Varioskan Lux 3020-176 luminometer (Thermo Fisher Scientific). Data was analyzed by one-way ANOVA using *GraphPad Prism* v6 software. To determine mayahuelin IC_{50}, data was adjusted to a non-linear dose-response curve using the logistic fitting with four parameters function (Origin v9.6 software).

Mayahuelin Homology Modeling

Mayahuelin amino acid sequence from *A. tequilana* var. *azul* was used to fetch protein sequences with the *BLASTP* program[1] restricting the search only to proteins with known X-ray crystalographic structures resolved at 2.5°A or lower and

[1]https://blast.ncbi.nlm.nih.gov

accepting only outputs with an identity larger than 30 and >90% coverage. Charybdin from *Drimia maritima* [sea onion, Asparagaceae, subfamily Scilloideae] (Touloupakis et al., 2006) obtained the best hit with 37% identity and 92% coverage, relative to mayahuelin. Charybdin structure was retrieved as a.pdb file from the Protein Data Bank[2] and additional lateral chains, ions, and ligands were removed to obtain the basic frame of the protein. Protein alignment between the amino acid sequences of mayahuelin and charybdin was performed with *T-COFFEE* v11.0 and hand edited before conversion to.pir format. Both.pir and.pdb files were uploaded to the *Modeller* v9.19 program generating 10,000 mayahuelin models. Only the top three models – those having the lowest DOPE (Discrete Optimized Protein Energy) scores – were further considered for analysis (Shen and Sali, 2006). The selected models were evaluated with *ERRAT*[3], a tool designed to identify protein regions on need of refinement. The model with the highest score was subject to refinement in regions with >99% rejection. Using *Modeller*, 1,000 new refined models were created by repeating a new cycle of evaluations with DOPE and ERRAT to identify the model with the best score. The final model was evaluated on a Ramachandran graph using *Molprobity*[4].

Other Molecular Biology Methods Used

Protocols for the cloning and expression of *Mayahuelin* gene in *S. cerevisiae*, for the estimation of *Mayahuelin* transcript levels, and for the amplification of *Mayahuelin* genes by RT-PCR and direct PCR are described in **Supplementary Information**.

Mayahuelin Orthology Tests

Protein-coding nucleotide sequences of *Mayahuelin* candidates were aligned together with the 12 RIP family members of *Agave tequilana* var. *azul* (**Supplementary Figures S2, S3, S11, S12**, and **Supplementary Table S4**). We used *TranslatorX*, a multiple-alignment method based on the corresponding aa alignments encoded by such sequences. *TranslatorX* first performs the aa aligment and from the output optimizes the aligment of the nucleotide sequences (Abascal et al., 2010). As an additional parameter, the algorithm *MUSCLE* was used (Edgar, 2004). The resulting multiple alignments were analyzed by both Maximun-likelihood (ML) and Bayesian inference (BI) methods.

The program *PhyML* v3.0 (Guindon et al., 2010)[5] was used to perform ML. To obtain the best nucleotide substitution reconstruction the *SMS* algorithm was selected (Lefort et al., 2017) and implemented in *PhyML*. Default values were used for construction of the starting tree (BIONJ option) and for tree improvement (NNI, Nearest-neighbor interchange option). Support values were obtained by bootstrapping with 1,000 pseudoreplicates.

BI analyses were performed with MrBayes v3.2 (Ronquist et al., 2012). To analyze the sequences based on codons, instead of single nucleotides, the *GTR* (General Time Reversible) nucleotide substitution refinement was selected with the *codon* option. Other parameters were used under default values and considering aC3095_122 RIP sequence as outgroup. Two independent and simultaneous Markov chain Monte Carlo simulations (MCMC) were run using six hot chains and two cold chains; random starting trees, sampling of refinement parameters, posterior probabilites every 500 generations, and discarding 25% of initial generations (burn-in) were implemented. A total of 1e+06 generations were run. Convergence of the two chains was determined by examining the average standard desviation of splites frequencies in MrBayes and by calculating the efective sample size (ESS) with the program *Tracer* v1.7.1 (Rambaut et al., 2018). An standard deviation <0.01 and a total ESS >200 were used as criteria to establish convergence in the stationary phase.

Using either ML or BI methods, only candidate sequences that clustered with clone aC630_3 (*Mayahuelin*) were accepted as orthologs for phylogenetic analysis.

Phylogenetic Analyses

A panel of 34 taxa within the Littaea and Agave subgenera was assembled (see **Supplementary Tables S1, S6**) with a focus on *A. tequilana* and other members of the Rigidae group (*A.angustifolia*, *A. rhodacantha*, and *A. aktites*) in addition to members of the Hiemiflorae (*A. isthmensis*), Americanae (*A. americana*), Parryanae (*A. guadalajarana* and *A. parryi*), and Marmoratae (*A. zebra*) groups within the Agave subgenus as well as the Littaea subgenus represented by Choripetalae (*A. guiengola*), Amolae (*A. vilmoriniana*) and Marginatae (*A. horrida*) groups. *Beschorneria calcicola* was chosen as an outgroup. *B. calcicola* is a member of Agavoideae consistently found in separate clades that stem from more basal nodes relative to the *Agave* genus in all published molecular phylogenies (Bogler and Simpson, 1996; Bogler et al., 2006; Good-Avila et al., 2006; Heyduk et al., 2016). The selection considered both cultivated and wild individuals: five cultivars of *A. tequilana* (*azul*, *manso*, *sigüín*, *mano larga*, and *bermejo*), three cultivars of *A. angustifolia* (*espadín*, *Huajuapan*, and *Ahuacuotzingo*), four cultivars of *A. rhodacantha* (*ixtlero amarillo*, *Ejutla*, *UNAM*, and *Nayarit*), seven wild accessions of *A. angustifolia* (from Sonora, Sinaloa, Jalisco, Guerrero, and Oaxaca), five wild accessions of *A. rhodacantha* (from Sonora, Sinaloa, Jalisco, and Oaxaca), nine additional *Agave* species with or without current/pre-historical evidence of utilization by humans (*A. americana*, *A. aktites*, *A. guadalajarana*, *A. guiengola*, *A. horrida*, *A. isthmensis*, *A. parryi*, *A. vilmoriniana*, and *A. zebra*), and *B. calcicola*.

Mayahuelin F2-R2 or F6-R6 specific primer pairs (**Supplementary Table S3** and **Supplementary Figure S8**), complementary to gene sequences encoding the N- and C-terminal ends of mature mayahuelin, were used to amplify *Mayahuelin* orthologs by either RT-PCR or direct genomic PCR. None of the amplified genomic sequences contained introns. *Mayahuelin* paralog sequences were excluded from the analyses (see **Supplementary Figures S4, S12** section for orthology tests). Validated sequences were aligned with the *TranslatorX* and *MUSCLE* programs (Edgar, 2004; Abascal et al., 2010). After trimming, the length of the alignment spanned 630 bp yielding 104 variable sites with a proportion of 0.165 (**Table 1**). The consensus tree for Maximum-likelihood was inferred with *PhyML* 3.0 and by bootstrapping with 1,000 pseudoreplicates,

[2]http://www.rcsb.org

[3]http://services.mbi.ucla.edu/ERRAT

[4]http://molprobity.biochem.duke.edu

[5]http://www.atgc-montpellier.fr/phyml/

TABLE 1 | A summary of the sequence alignment of 34 Agavoideae taxa.

No. of taxa	34
Alignment length (bp)	741
Alignment_length after trimming (bp)	630
Undetermined characters	216
Missing percent	1.008
Number of variable sites	104
Proportion of variable sites	0.165
Singleton sites (bp)	62
Parsimony informative sites	42
Proportion of parsimony informative sites	0.067
AT content	0.47
GC content	0.53

Data were obtained with the AMAS software (Borowiec, 2016).

as described above. For BI analyses, a consensus tree was derived with MrBayes program. The *GTR* reconstruction was selected under the *codon* option using *B. calcicola* sequence as an outgroup. A total of 1.2e+06 generations of MCMC simulations were run. Other parameters and convergence measurements were used as described above.

RESULTS

Mayahuelin Is Abundant in Spike Leaves and Seeds

Agaves form a spirally shaped rosette; at its center, a spike is visible (a group of folded leaves that surround and protect the shoot apical meristem) and several unfolded and more mature leaves in the periphery. During the progress of development of an *ad hoc* method for total protein extraction from *A. tequilana* var. *azul* leaves (Lledías et al., 2017a,b), we noticed a highly accumulated 27 kDa molecular mass protein present exclusively in the spike and we named this protein mayahuelin (**Figure 1A**). This structure is formed by a variable number of leaves (approximately 8 to 15, **Figures 1B,C**). Protein from each spike leaf was individually analyzed to estimate mayahuelin content. A strong signal was detected in leaves 10 to 4 (in a proximal to distal order, relative to the shoot apical meristem [SAM]) and lower levels in leaves 3 to 1 (**Figures 1C–E**). In samples from internal (I), middle (M), and outer (O) rosette sectors, as well as in stem (St) and roots (R), mayahuelin was absent (**Figures 1D,E**, right panels). Qualitative estimation of *Mayahuelin* RNA accumulation by quantitative RT-PCR indicated that *Mayahuelin* transcripts were expressed in spike leaves only and that their levels were near absent in leaf 1, reaching a peak in leaves 3 to 5, and levels decreasing in leaves 7 and 8 (**Figure 2**). Analyzed tissues from a mature *A. tequilana* var. *azul* plant (**Figure 3A**) undergoing sexual reproduction showed a high mayahuelin protein content (equivalent to spike leaves) in seeds (**Figure 3D–F**, lane 7). A faint signal was also observed in the floral peduncle (**Figures 3E,F**, lane 5). Mayahuelin was absent in floral structures (**Figures 3B,C,E,F**, lanes 2, 3, 4, and 6).

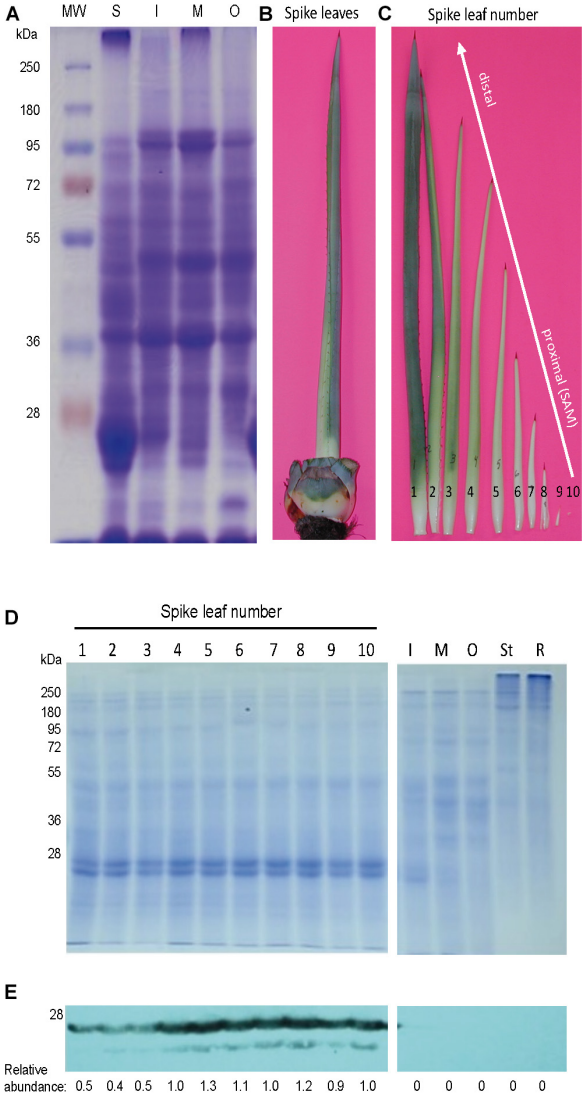


FIGURE 1 | In the rosette of *A. tequilana* var. *azul*, mayahuelin is exclusively expressed in spike leaves. **(A)** SDS/PAGE profile of total proteins (20 μg) in leaves from the spike (S) and from the internal (I), medium (M) and outer (O) regions of an *A. tequilana* var. *azul* rosette. The spike **(B)** consists of several folded leaves in different stages of development **(C)**. To estimate the levels of mayahuelin, total protein extracts (5 μg) were individually analyzed by SDS/PAGE **(D)** and by immunobloting using an anti-mayahuelin antibody **(E)**. Mayahuelin levels are high in spike leaves (**E,** left image). In leaves from the internal (I) medium (M), and outer (O) regions, or in tissues from the stem (St) and roots (R), mayahuelin was absent (**E**, right image). Protein levels shown in **(E)** panel are relative to spike leaf number 10.

Mayahuelin Purification

Native electrophoresis of solubilized proteins obtained from spike leaves (after acetone precipitation) revealed a pattern of two major bands (**Figure 4A**). Band 2, recovered by native electroelution, methanol/chloroform precipitation, and analyzed by SDS/PAGE, appeared as a unique 27 kDa molecular mass protein (**Figure 4B**, lane 2) that comigrated with the major band

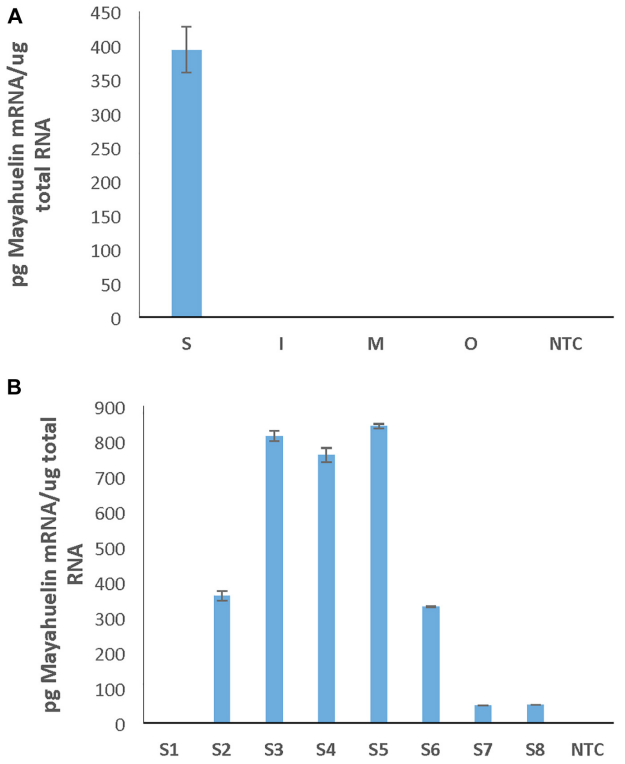


FIGURE 2 | Mayahuelin transcripts accumulate only in spike leaves of *A. tequilana* var. *azul*. Quantitative PCR assays were performed using total RNA isolated from leaves from different rosette sectors (**S**pike, **I**nternal, **M**edium, and **O**uter) **(A)** or from individual leaves of the spike (leaves S1–S8 represent a developmental leaf gradient, where leaf S1 is the most mature one and S8 the youngest one and the leaf closest to the SAM) **(B)**. NTC, no template control reaction. The spike of the *A. tequilana* specimen used in this study contained only 8 leaves and came from a different specimen to the one used in experiment on **Figure 1**.

present in a total protein preparation from *A. tequilana* var. *azul* from spike leaves (**Figure 4B**, lanes 1 and 2). The native protein, cross-linked to the polyacrylamide matrix, was used to obtain polyclonal antibodies for immunoblot analysis. A strong signal detected by the anti-mayahuelin sera was observed from protein band 2 (mayahuelin) obtained by electroelution (**Figures 4C,D**, lane 2) and from the 27 kDa protein present in total protein extracts made from spike leaves (**Figures 4C,D**, lane 1).

Accumulation of Mayahuelin in Spike Leaves Is Conserved in the *Agave* Genus

We searched for mayahuelin presence in spike leaves total protein extracts from a panel of species of the subgenera Littaea [*A. bracteosa*, *A. desertii*, *A. guiengola*, *A. isthmensis*, *A. kerchovei*, *A. striata sp. falcata*, *A. titanota*, *A. victoriae-reginae*, and *A. vilmoriniana*] and Agave [*A. angustifolia*, *A. petrophila*, *A. rhodacantha*, and *A. zebra*] (**Figure 5**). Mayahuelin was detected both in the Littaea and Agave subgenera. In *A. angustifolia*, *A. bracteosa*, *A. rhodacantha*, and *A. vilmoriniana* spike leaves, mayahuelin content was similar or above that in *A. tequilana* var. *azul* (**Figure 5B**, lanes 1, 2, 3, 9, 13). Low levels of mayahuelin were observed in spike leaves of *A. isthmensis*, *A. kerchovei*, *A. striata sp. falcata*, and *A. titanota* (**Figure 5B**, lanes 6, 7, 10, and 11). In *A. deserti*, *A. guiengola*, *A. petrophila*, *A. victoriae-reginae*, and *A. zebra* spike leaves (**Figure 5B**, lanes 4, 5, 8, 12, and 14) mayahuelin levels were below detection even at longer film exposure times.

Mayahuelin Protein and Gene Sequencing

To determine the primary aa sequence of mayahuelin, the purified protein was analyzed in two independent biological replicas by mass spectrometry (LC-MS/MS). *De novo* peptide sequencing of mayahuelin fragments (for experiment 2 see **Supplementary Table S2**, **Supplementary Information**) from both analyses showed complete identity with cDNA consensus sequence aC630_3 reconstructed from an EST *A. tequilana* var. *azul* library (Martínez-Hernández et al., 2010; Simpson et al., 2011). N-terminal sequencing of mayahuelin was accomplished by Edman degradation. After 15 cycles, the amino acid sequence obtained was VKFEVNLDVRTLXAA. This sequence matched perfectly well with two peptides (# 1 and # 2) sequenced by LC-MS/MS (**Supplementary Table S2**). One peptide from the first experiment (data not shown) sequenced by LC-MS/MS contained Q at its C-terminus; we assumed that it represented the C-terminus of the protein since it lacked R or K. In the second experiment, this peptide contained K at the C-terminus (peptide # 14 in the second experiment shown in **Supplementary Table S2**). The analyses of aC630_3 ESTs confirmed the presence of transcripts containing either AAA or CAA codons at this position, explaining the two versions obtained by LC-MS/MS. Therefore, the calculated molecular mass of mayahuelin is 27,251.93, matching very well mayahuelin's apparent molecular weight calculated by SDS-PAGE. Thus, mayahuelin aa sequence obtained by MS, after the assembly of 14 peptide sequences, covered 73% of the protein (**Figure 6** and **Supplementary Table S2**). The alignment of mayahuelin aa sequence obtained by MS and the sequence predicted by cDNA sequence aC630_3 suggested that mayahuelin was synthetized as a precursor of 310 amino acids in length to which the N- and C- termini were removed (**Figure 6**). Therefore, mature mayahuelin primary sequence consists of 245 amino acids; the N-terminal end of mayahuelin precursor protein contains a putative signal peptide for extracellular secretion (**Supplementary Figures S5A,B**). Since cDNA sequence aC630_3 is missing the initiation codon at the 5′ end, mayahuelin precursor must be longer in length (**Figure 6**). A protein BLAST search at the NCBI database showed that mayahuelin sequence matches to a group of proteins known as Type I RIPs (Ribosome Inactivating Proteins) (**Supplementary Figure S6**). Protein sequence alignment to known RIPs revealed an amino acid substitution in the active site of mayahuelin: an aspartate replacing a highly conserved tyrosine (Y76D) (**Figure 6** and **Supplementary Figure S6**) as a consequence of a single base change in the tyrosine codon (TAC→GAC). This amino acid change resembles a substitution in charybdin, a 29 kDa Type I RIP from the sea squill plant, (*Drimia maritima*) that

shows a valine substitution at the same tyrosine in the active site (Touloupakis et al., 2006; **Figure 6** and **Supplementary Figure S6**). *D. maritima* is the accepted name for its synonym *Charybdis maritima* and it is a member of the Asparagaceae family as *A. tequilana*. The toxicity of charybdin is low in a mouse reticulocyte *in vitro* translation system compared with a canonic RIP like saporin (Touloupakis et al., 2006). A search for *Mayahuelin* related sequences in the *A. tequilana* var. *azul* EST library (Martínez-Hernández et al., 2010; Simpson et al., 2011) uncovered eleven additional cDNA sequences encoding putative RIPs. Analysis of their open reading frames showed that *Mayahuelin* is the only member of the *A. tequilana* var. *azul* RIP family encoding substitutions in the active site of the protein (**Supplementary Figure S2**, **Supplementary Information**).

Mayahuelin Expression Is Harmless for the Growth of *S. cerevisiae* Cells

To evaluate the cytotoxicity of mayahuelin, two different plasmid constructs were engineered to express the *Mayahuelin* gene in the W303-1a yeast strain: pYES-DEST52::*Mayahuelin* (R1), pYES-DEST52::*Mayahuelin*::*V5*::*6his* (R2) (**Supplementary Figure S7**). Total protein extracts from transformed yeast cells (**Figure 7A**) were evaluated after galactose induction for the presence of mayahuelin protein by immunoblot analysis. In R1-transformed yeast extracts, a 27 kDa well-defined band was detected after 16h of galactose induction reaching maximum levels at 24 h (**Figure 7B**). This band corresponds to mature mayahuelin. Yeast cells transformed with *Mayahuelin*::*V5*::*6his* plasmid (R2) expressed a 32 kDa band (expected molecular mass for mayahuelin fused to *V5* and *6xHis* epitopes) at 20 and 24 h post induction with lower intensities than those observed for R1-transformed yeast (**Figure 7B**). 24 h growth curves profiles for R1-, R2- and mock- transformed (pYES) yeast were obtained in SD-galactose, plus requirements, induction medium (**Figure 7C**). Similar growth behavior was observed for all three strains with an exponential phase starting at 12 h that was maintained until 20 h. At 24 h, only pYES cultures upheld an exponential growth pattern, while R1 and R2 showed a slight decrease indicating a probable entrance to post-diauxic shift phase (**Figure 7C**).

Mayahuelin Inhibits Luciferase *in vitro* Translation on a Wheat Germ System

Mayahuelin was purified to homogeneity using a protocol (Lledías, Gutiérrez, and Nieto-Sotelo, in preparation) based on standard chromatographic methods to directly evaluate its effects on protein synthesis. Mayahuelin inhibited luciferase *in vitro* translation in a dose-dependent manner, when tested on a wheat germ cell-free system. At an initial concentration of 15.4 nM, luciferase translation was 0.65 relative to control with no RIP. At 30.8 nM the registered inhibition was 0.83, a similar value to the 0.85 obtained when saporin was added at 13.3 nM used as positive control. Full inhibition of luciferase translation was obtained when mayahuelin reached 123.2 nM (**Figure 8A**). The translation inhibitory effects at 30.8- and 61.6 nM mayahuelin were not statistically different to saporin at 13.3 nM (**Figure 8A**). Luciferase relative expression values were transformed to inhibition percentage and adjusted at a dose-response curve. Mayahuelin showed an IC_{50} = 10.43 nM (R^2 = 0.999) for protein *in vitro* translation in the wheat germ system employed.

A Phylogenetic Reconstruction of *Agave* Based on *Mayahuelin* DNA Sequences

We used *Mayahuelin* gene as a phylogenetic marker with the following three main goals: to understand whether the Y76D substitution was found in other Agavoideae species, to discern its possible relation to domestication/improvement in the genus, as there is an ancient human history of cultivation and exploitation of wild Agavoideae populations, and to identify close relatives of the cultivated varieties within wild populatons of Rigidae to contribute to their conservation.

Phylogenetic reconstructions were obtained by Maximum-likelihood (ML) and by Bayesian inference (BI) methods. Estimates of the ML and BI phylogenetic analyses are shown in **Table 2**. In ML, GTR + Γ + I was the best nucleotide substitution refinement. As expected for a protein coding marker such as *Mayahuelin*, in BI analyses codon + GTR proved to be the best reconstruction relative to nucleotide-based refinements. As shown in **Figure 9** and **Supplementary Figure S13**, the genetic diversity of *Mayahuelin* within Agavoideae has been sufficiently rapid to grant resolution at the intraspecific, intrageneric, and intergeneric levels. In both ML and BI analyses *B. calcicola* was resolved as a separate lineage (clade 1) from all *Agave* taxa. In BI (**Figure 9**) two *Agave* lineages were clearly resolved: one represented by *A. vilmoriniana*, *A. tequilana* var. *bermejo*, and *A. angustifolia* ssp. *rubescens* (clade 2), and a very strongly supported second lineage (posteriror probability = 99) including all other species analyzed (clade 3). *A. rhodacantha*, *A. angustifolia*, and *A. tequilana* accessions were polyphyletic, as they were dispersed among the two *Agave* clades in the tree (**Figure 9**). Clade 2 was the earliest diverging group suggesting a more distant relationship of *A. vilmoriniana*, *A. tequilana* var. *bermejo*, and *A. angustifolia* ssp. *rubescens* relative to the other species in the tree.

The major lineage (clade 3), was subdivided in seven subclades: four of them well resolved (A, D, F, and G) and three of them showing hard polytomy (B, C, and E). Subclade A had a single species (*A. isthmensis*) and its separation from subclades B, C, D, E, F, and G was well supported (posterior probability = 87). Subclade D was very strongly supported (posteriror probability = 99) and revealed *A. tequilana* var. *sigüín* as a close relative of a cultivated form (*ixtlero amarillo*) of *A. rhodacantha* from southern Jalisco, a wild specimen of *A. angustifolia* also from southern Jalisco, and *A. guiengola* from Oaxaca. *A. zebra*, *A. parryi*, *A. guadalajarana*, a wild accession of *A. angustifolia* from Sinaloa, and *A. aktites* conformed the well resolved subclade F, whereas *A. tequilana* var. *azul,* a wild (from Sonora) and one cultivated (from Oaxaca) forms of *A. rhodacantha*, *A. angustifolia* var. *espadín*, a wild *A. rhodacantha* accession from southern Jalisco, and a cultivar of *A. rhodacantha* from unknown origin (UNAM) clustered in clade G. The sub-branch of G subclade that included *A. tequilana* var. *azul, A. rhodacantha* from Alamos, Sonora,

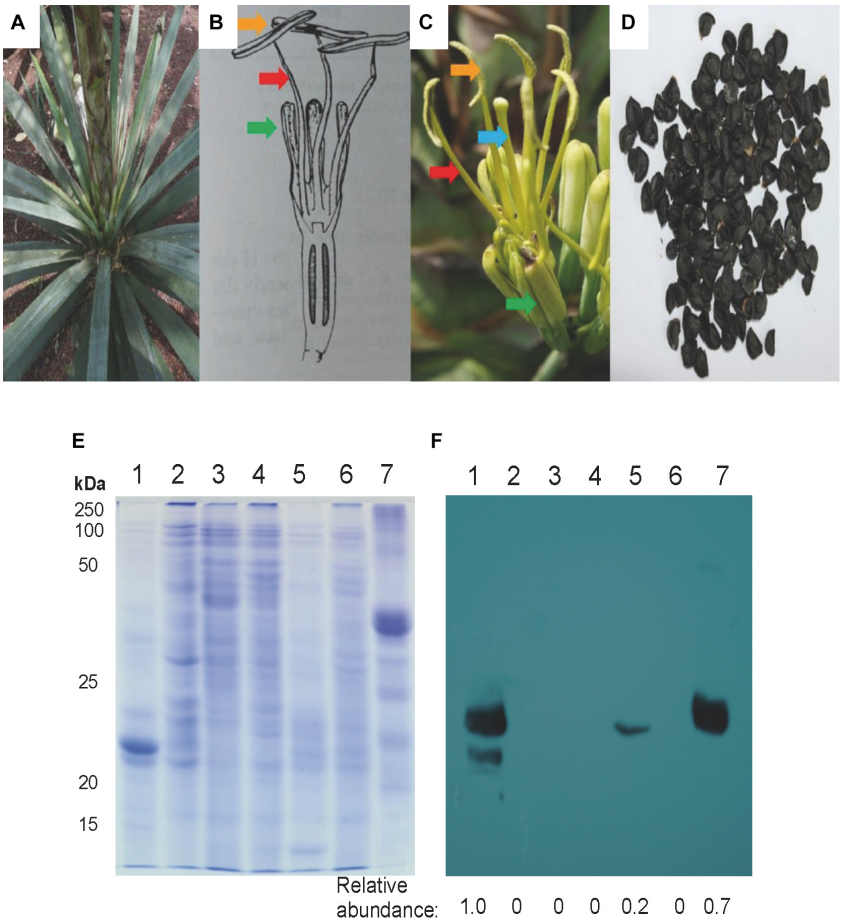


FIGURE 3 | Mayahuelin levels in *A. tequilana* var. *azul* mature/reproductive plant organs. Upper portion: Images of a whole plant **(A)**, a schematic representation and image of a flower **(B,C)**, and seeds **(D)** from *A. tequilana* var. *azul* plants used for analyses. In **(B,C)** green, red, blue, and orange arrows point tepals, filaments, pistil, and anthers, respectively. Lower portion: SDS/PAGE **(E)** and immuno-blot **(F)** analyses of total protein preparations from spike leaves (lane 1), anthers (lane 2), filaments (lane 3), pistils (lane 4), flowering spike (lane 5), tepals (lane 6) and seeds (lane 7). Protein levels shown in **(F)** panel are relative to spike leaves sample.

and *A. rhodacantha* from Ejutla, Oaxaca was well supported (posterior probability = 84).

Polytomic subclades B and C were composed by *A. rhodacantha*, *A. angustifolia*, and *A. tequilana* accessions in addition to *A. americana* present only in subclade C. Likewise, polytomic subclade E included *A. rhodacantha*, *A. angustifolia*, and *A. tequilana* accessions plus *A. horrida.*

A BI phylogenetic reconstruction that included only *Mayahuelin* sequences from *A. tequilana* and *A. angustifolia* var. *espadín*, showed that the *azul*, *espadín*, and *manso* varieties were the closest relatives, followed by *mano larga* and *sigüín*, whereas *bermejo* is a distant relative of the first five (**Figure 9** and **Supplementary Figure S9**, **Supplementary Information**).

The ML reconstruction (**Supplementary Figure S13**) confirmed subclades D, G, and, partially, clade 2 of the BI reconstruction (**Figure 9**). All other species of clade 3 were unresolved by ML methods. In general, the ML topology had low support values that ranged between 39 and 68 (**Supplementary Figure S13**); an exception was clade 2, showing a moderate support value (81).

Sequences of the 34 taxa showed high variability at amino acid position 76 of mature mayahuelin, an important residue in the active site of the protein, with six allelic states present: Y/Y, D/D, S/S, Y/D, Y/S, and D/S (**Supplementary Figure S10**, **Supplementary Table S5**). Both *A. tequilana* var. *azul* and *A. angustifolia* var. *espadín* were homozygous for the Y76D substitution, confirming their closer relationship relative to other *A. tequilana* varieties (**Figure 9**; **Supplementary Table S5**, and **Supplementary Figure S9**, **Supplementary Information**).

DISCUSSION

Mayahuelin Protein Is Highly Conserved and Active Site Substitution Alleles Are Common in Agavoideae

Mayahuelin from *A. tequilana* var. *azul* is an atypical RIP, as one of the canonical amino acids that compose its active site (tyrosine 76) is substituted by aspartate (**Figure 6** and

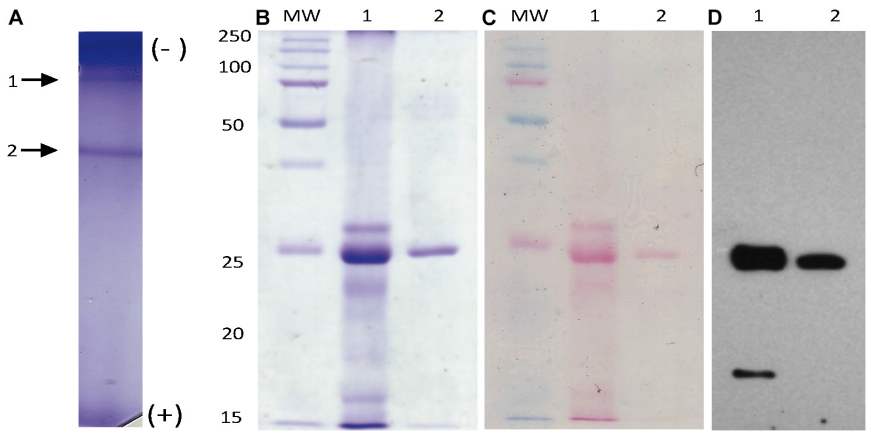


FIGURE 4 | Mayahuelin native-protein purification. Proteins from spike supernatants were separated by native PAGE **(A)**. After Coomassie blue-staining, two main bands were observed (arrows 1 and 2 in **A**). Band 2 (mayahuelin) was electroeluted, denatured, and analyzed by SDS/PAGE (**B**, lane 2). Mayahuelin was the major band (27 kDa) detected in a total protein preparation from spike leaves of *A. tequil*ana var. *azul* (**B**, lane 1). A total protein extract from spike leaves of *A. tequil*ana var. *azul* and purified mayahuelin were transferred to nitrocellulose filters, stained with Ponceau red (**C**, lanes 1 and 2) and probed with polyclonal immunoadsorbed rabbit-anti-mayahuelin antibodies (**D**, lanes 1 and 2).

Supplementary Figures S2, S6). A similar natural mutation was reported previously for charybdin, a RIP from *D. maritima*, also a member of the Asparagaceae family (Touloupakis et al., 2006). Here, we found that the frequency of *Mayahuelin* ortholog genes in other Agavoideae species, encoding amino acid substitutions at Y76, is unexpectedly high (**Supplementary Figure S10**, **Supplementary Table S5**, **Supplementary Information**), becoming even more intriguing the study of their physiological or ecological roles in the plant.

Mayahuelin expression in spike leaves of *A. tequilana* var. *azul* is quite high, representing at least 20% of the total protein (**Figure 1A**). This high level of RIP abundance is comparable to reports in other monocotyledonous plants, as for charybdin and for a type I RIP from *Iris hollandica* (Van Damme et al., 1997; Touloupakis et al., 2006). Both *D. maritima* and *I. hollandica* accumulate RIPs in the bulb, a storage organ where perhaps they could serve as storage proteins, given their large quantities. It is also tempting to speculate that the high levels of accumulation of mayahuelin in immature leaves and mature seeds serve as a storage protein. Expression of vegetative storage proteins is dynamic: they are regulated in response to nitrogen nutrition, wounding, and hervibory (Staswick, 1990; Berger et al., 2002). Thus, the physiological status of the specimens analyzed could explain the disparate levels of mayahuelin in spike leaves (**Figure 5**) and the lack of correlation with their genetic distance relative to *A. tequilana* var. *azul*. For example, *A. zebra* clusters in clade F, which is a sister group of clade G where *A. tequilana* var. *azul* belongs (**Figure 9**). However, mayahuelin was not detected in *A. zebra* (**Figure 5**, lane 14). In contrast, *A. vilmoriniana*, on clade 2 is very distant from *A. tequilana* var. *azul* (**Figure 9**), but contained very high levels of mayahuelin (**Figure 5**, lane 13). Alternatively, and not mutually exclusive, levels of mayahuelin could also be the result of the affinity of the mayahuelin primary polyclonal-antibody toward mayahuelins from other species. We found that, as soon as the most developed leaf in the spike unfolds, mayahuelin levels decrease drastically (**Figures 1D,E**). This evidence supports the role of mayahuelin as a vegetative storage protein. However, we do not exclude additional roles of mayahuelin in defense against hervibory or other regulatory or enzymatic roles (Stirpe and Batelli, 2006; Stirpe, 2013). As discussed forward, the role of humankind during plant domestication or exploitation of plant natural resources for its benefit could also be the driving force for selection of novel structures in the active site of type I RIPs, which by themselves, are not naturally cytotoxic, as they lack the protein (B chain) containing the cell-binding domain for cell internalization, typical of type II RIPs (Stirpe and Batelli, 2006; Stirpe, 2013). Interestingly, type I RIPs can be internalized and strongly enhance its cytotoxic effects when combined with triterpenoid- or steroidal- saponins, that increase membrane permeability (Korchowiec et al., 2015). As steroidal saponins are commonly found in *Agave* (Santos-Zea et al., 2012), this synergic interaction is a possibility worth studying further.

Activity of Mayahuelin Could Be Affected by Structural Changes in Its Active Site

When mature *Mayahuelin* gene was introduced in *S. cerevisiae* the cells expressed mayahuelin after galactose induction as a 27 kDa protein, in the R1 transformed strain, and as a 32 kDa protein, in R2 cells (**Figure 7B**). RIP heterologous expression is lethal in yeast when the canonic RIP catalytic site is intact, for example when PAP [type I RIP from *Phytolacca americana*] (Hur et al., 1995) or ricin A chain (Li et al., 2007) are expressed. We found that mayahuelin expression was not cytotoxic in *S. cerevisiae* (**Figure 7C**). The natural substitution of one tyrosine residue by aspartate in the mayahuelin active site (Y76D) could explain the null cytotoxic effect in yeast cells, although it is also conceivable that other catalytic or non-catalytic domains in mayahuelin are responsible for its null yeast growth inhibition.

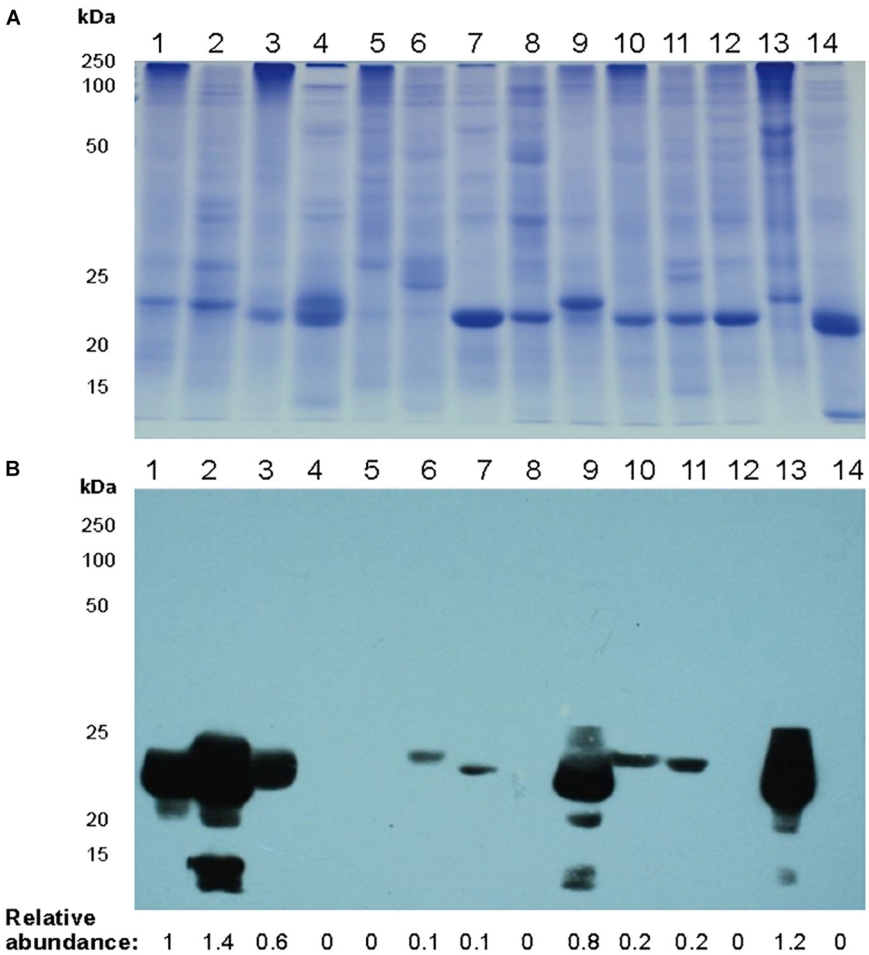


FIGURE 5 | Relative levels of mayahuelin in different *Agave* plant species. **(A)** SDS/PAGE of total protein extracts obtained from spike leaves of *Agave* species. **(B)** Immunoblot analysis of protein extracts [10 μg] obtained from spike leaves of *Agave* species using *A. tequilana* var. *azul* anti-mayahuelin antibody. *A. tequilana* var. *azul* (lane 1); *A. angustifolia* (lane 2); *A. bracteosa* (lane 3); *A. desertii* (lane 4); *A. guiengola* (lane 5); *A. isthmensis* (lane 6); *A. kerchovei* (lane 7); *A. petrophila* (lane 8); *A. rhodacantha* (lane 9); *A. striata* ssp. *falcata* (lane 10); *A. titanota* (lane 11); *A. victoriae-reginae* (lane 12); *A. vilmoriniana* (lane 13); and *A. zebra* (lane 14). Protein levels shown in b panel are relative to A. *tequilana* var. *azul* sample.

Although not as innocuous as in yeast cells, the IC_{50} obtained for mayahuelin on a cell-free wheat germ extract was 10.43 nM (**Figure 8**). The IC_{50} for inhibition of protein synthesis by RIPs varies according to the cell-free system used (Fuchs, 2019) and seems to be mediated by domains outside the catalytic site, which are highly variable among RIPs. For example, the IC_{50}s for a RIP from *Clerodendrum aculeatum* (CA-SRI) are 0.008- and 0.8 nM in rabbit reticulcyte lysate and wheat germ systems, respectively, Kumar et al. (1997). RIP IC_{50}s in the wheat germ system cover more than three orders of magnitude, ranging from 0.2 nM, in the case of dodecandrin, to 800 nM for ricin A chain (Harley and Beevers, 1982; Reisbig and Bruland, 1983; Ferreras et al., 1993; Bonnes et al., 1994; Fuchs, 2019). In comparison, mayahuelin classifies as a RIP with medium capacity for protein inhibition in the wheat germ system.

The low toxicity of mayahuelin *in vivo* and *in vitro* could be related to its active site Y76D substitution. The structural analysis of available RIPs indicate that two tyrosine residues in the active site are aligned in parallel and confine the substrate adenine (**Figure 10**; Monzingo and Robertus, 1992; Savino et al., 2000). A tyrosine in a homologous region in the ricin catalytic site diminished ricin catalytic activity 15 times, when substituted for phenylalanine [Y80F] (Ready et al., 1991) and the same substitution completely abolished saporin depurination activity (Bagga et al., 2003). Ricin substitution Y80S, of a tyrosine that stabilizes the adenine-substrate, decreases depurination activity by 160-fold (Ready et al., 1991; Kim and Robertus, 1992). A decrease in translation inhibitory activity -about two orders of magnitude compared with ricin- was found in charibdyn, the first reported naturally occurring RIP with an active site mutation (Touloupakis et al., 2006). In the crystal structure of charybdin valine 79 -which substitutes Y79 in the active site- is not aligned with tyrosine 117. As a result, an open conformation of the active site is adopted (**Figure 10**). In addition, the valine aliphatic chain is not capable to keep the substrate adenine residue in place (Touloupakis et al., 2006). According to our

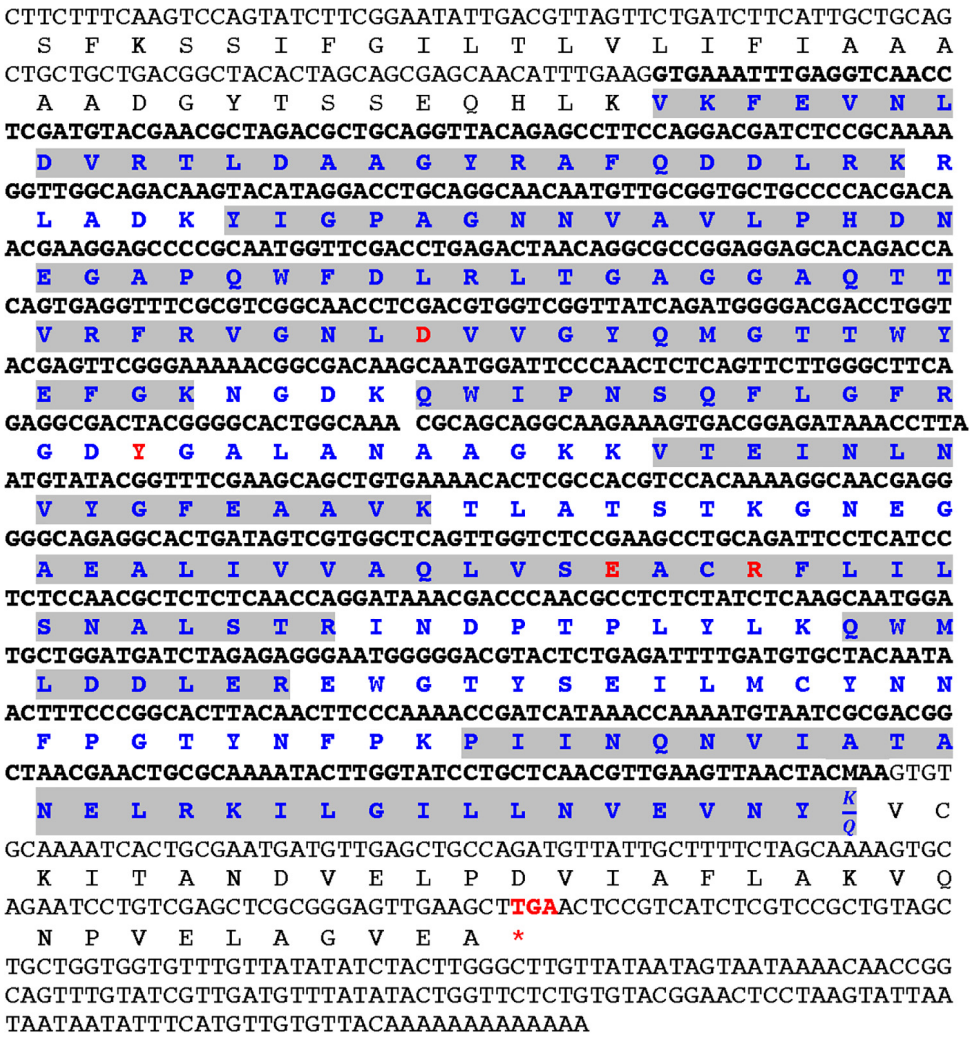


FIGURE 6 | Alignment of mayahuelin nucleotide and amino acid sequences from *A. tequilana* var. *azul*. The nucleotide sequence of clone aC630_3 from an *A. tequilana* var. *azul* cDNA library is shown in gray or black fonts. Predicted and experimentally determined (LC-MS/MS) amino acid sequences are shown below the nucleotide sequence. Highlighted in gray are the peptide sequences obtained by LC-MS/MS of pure mayahuelin isolated from spike leaves. Red codon (stop) indicates the end of the open reading frame in *Mayahuelin* cDNA. Bold letters in nucleotide and amino acid sequence indicate sequences of mature mayahuelin protein isolated from spike leaves. Red labeled amino acids show the position of the four conserved amino acids in RIP proteins. *Mayahuelin* Genbank accession number is MN913554. Mass spectrometry data was deposited at the Peptide Atlas repository (http://www.peptideatlas.org/repository/) under accession number PASS01536.

mayahuelin homology-based model, D76 does not align properly with Y110 causing a more open conformation of the active site, as in charybdin (**Figure 10**). This open conformation may cause an incorrect alignment of the substrate adenine in the SRL of the rRNA, decreasing depurination efficiency. A more complete understanding of the Y76D substitution in mayahuelin should emerge once a mayahuelin with a canonical Y76 version is characterized. This experiment remains pending.

Agave Phylogeny and Evidence of Disjunt Distribution Within Rigidae Group

The three Littaea species studied in this work were distributed in three different clades/subclades. *A. vilmoriniana* formed part of clade 2, with affinities to one *A. tequilana* and one *A. angustifolia* accessions, whereas *A. guiengola* and *A. horrida* were distributed in subclades D and E, respectively (**Figure 9**). This is in agreement with previous phylogenetic studies performed in Agavoideae (Bogler and Simpson, 1996; Bogler et al., 2006; Good-Avila et al., 2006; Eguiarte et al., 2013) showing the intermixing of species of the Littaeae and Agave subgenera, indicating that they are not monophyletic. However, two members of the Parrynae group (*A. guadalajarana* and *A. parryi*) clustered together and were consistent with their taxonomic classification (Gentry, 1982).

In Mexico, 53 *Agave* species are used for the production of mescal, tequila, and other alcoholic beverages. Out of these, eleven species are intensively cultivated by clonal propagation (Torres et al., 2015) causing an enormous decrease in genetic diversity as more areas are dedicated every year for their

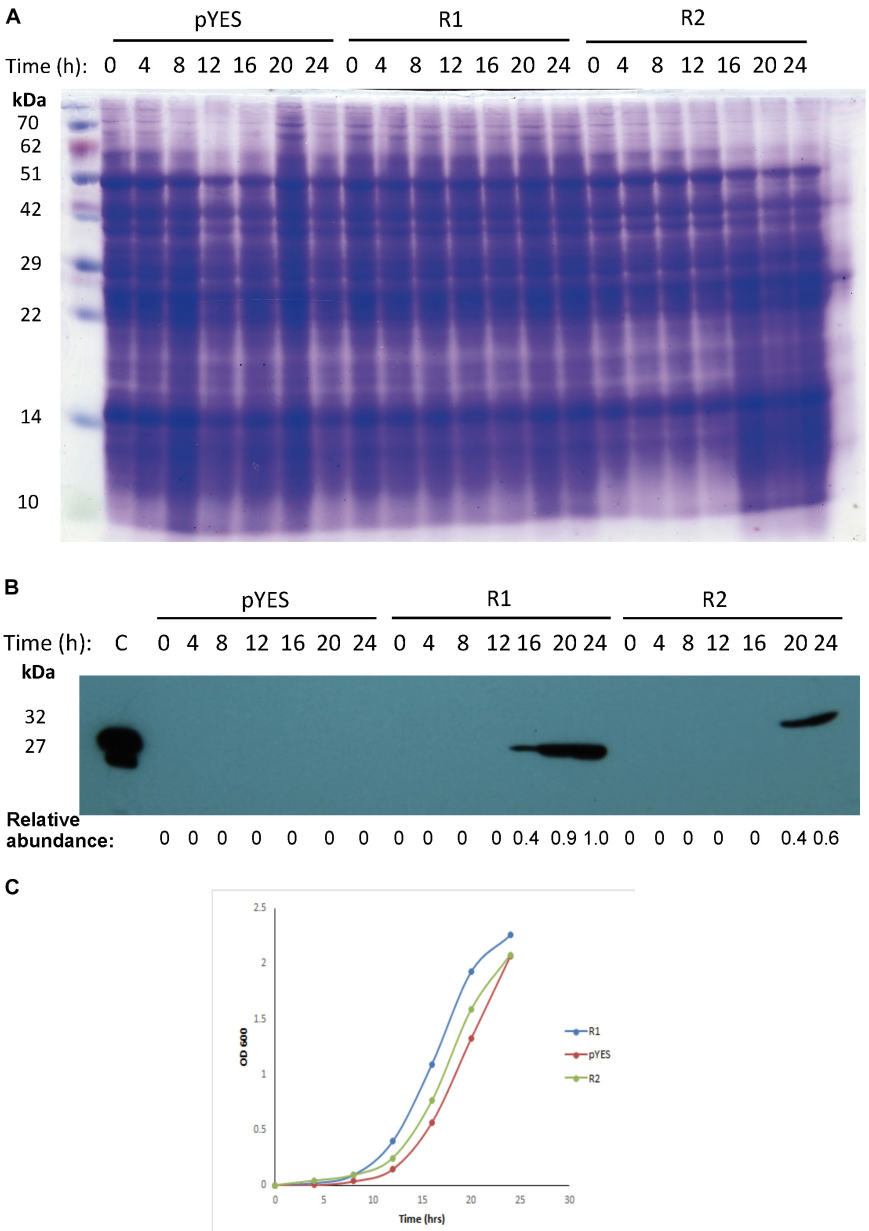


FIGURE 7 | Expression and effect of mayahuelin from *A. tequilana* var. *azul* on growth of *S. cerevisiae* cell cultures. **(A)** SDS-PAGE of yeast total protein extracts from strains transformed with an empty- or a pYES-DEST52 vector expressing *Mayahuelin* gene. Gel shows the protein profile after 0, 4, 8, 12, 16, 20, and 24 h of galactose induction in pYES-DEST52::*Mayahuelin* (R1), pYES-DEST52::*Mayahuelin::V5::6his* (R2) or mock-transformed (pYES) *S. cerevisiae* cells of the W303-a strain. **(B)** Analysis of mayahuelin content in yeast extracts by immunobloting using anti-mayahuelin antibodies. Lane marked **(C)** shows a positive control extract made from spike leaves of *A. tequilana* var. *azul*). **(C)** Growth curves of yeast cells transformed with R1(circles), R2 (triangles) or pYES (diamond). Protein levels shown in b panel are relative to 24 h time point in R1 sample.

cultivation and their wild relatives are continuously extracted from their natural habitat. Identification of the closest wild relatives of the cultivated forms is necessary for the deployment of conservation strategies and sustainable management practices for agave cultivation. More than 15 varieties are known for *A. tequilana* and their origins have been subject to speculation prior to the use of molecular phylogeny approaches. Using morphological characters, Gentry discussed the origin of four cultivars (*azul*, *listado al margen*, *manso*, and *pata de mula*) as well as *A. angustifolia* var. *espadín* from specific wild populations in Mexico (Gentry, 1982) while Valenzuela proposed the identity of four cultivars: *sigüín*, *moraleño*, *bermejo*, and *chato* (Valenzuela Zapata, 1995). More recently, the phylogenetic relationships between nine *A. tequilana* cultivars were derived using AFLP markers (Gil-Vega et al., 2006) recognizing three closely related groups: 1) *azul*, *azul listado*, *sigüín*, *manso*, and *moraleño*, 2)

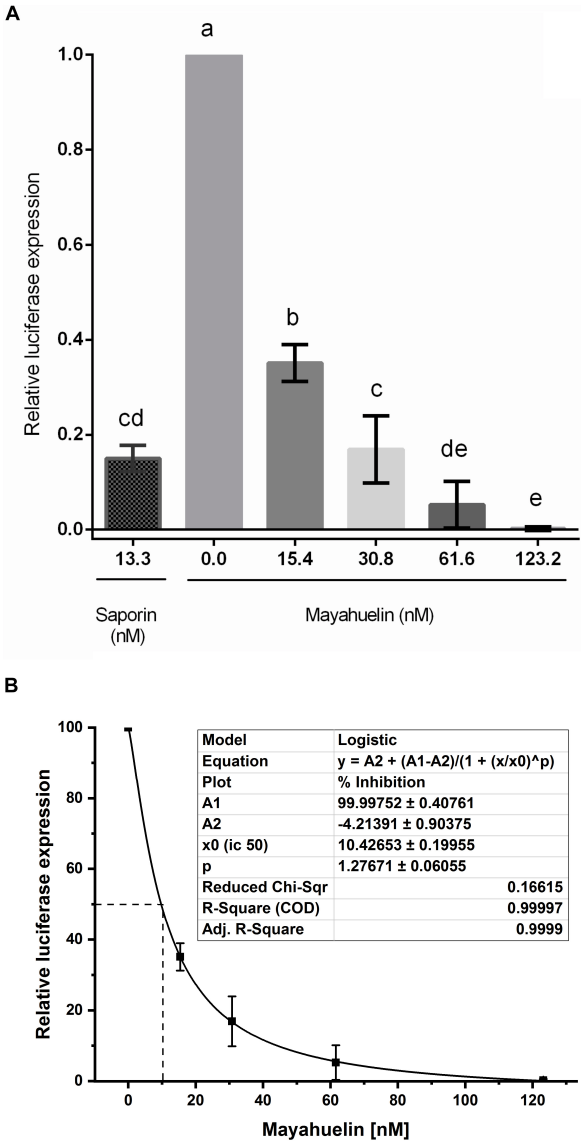


FIGURE 8 | Effect of mayahuelin from *A. tequilana* var. *azul* on protein translation *in vitro* on a wheat germ cell-free system. **(A)** A pure mayahuelin preparation obtained by chromatographic methods was assayed on a cell-free wheat germ translation system at the concentrations indicated (0–123.2 nM). A commercial preparation of saporin (from *Saponaria officinalis*) was used at 13.3 nM as a positive control. Luciferase translation was used as a reporter and its activity was measured on a luminometer. Data shown are the mean and standard deviation of three independent experiments. Statistical significant differences between treatments were calculated by one-way ANOVA using *GraphPad Prism v6* software. Bars with different letters are statistically different ($P < 0.05$). **(B)** Mayahuelin dose-response curve on luciferase traslation inhibition. Data from experiments in **(A)** was used to estimate the half-maximal inhibitory concentration (IC_{50}) of mayahuelin using Software *Origin v9.6.* Continuous line represents the graph adjustment. Dotted line corresponds to IC_{50}.

TABLE 2 | Estimates of substitution reconstructions used in Maximum-likelihood and Bayesian inference analyses.

	Maximum likelihood parameter estimates	95% credible interval of Bayesian parameter samples	
			Total ESS
Refinement	GTR + Γ + I	codon + GTR	NI
Llk	−179,358,746	NI	NI
AIC	373,717,492	NI	NI
TL	NI	0.20176–0.27790	4511
r(A<->C)	1.68181	0.15769–0.28965	2379
r(A<->G)	1.94162	0.20164–0.33773	1810
r(A<->T)	0.77428	0.04246–0.13191	2588
r(C<->G)	0.81229	0.07187–0.17619	2573
r(C<->T)	1.21419	0.09647–0.23001	1760
r(G<->T)	1.00000	0.07639–0.19743	2630
pi(A)	0.28822	NI	NI
pi(C)	0.26051	NI	NI
pi(G)	0.25992	NI	NI
pi(T)	0.19135	NI	NI
Alpha	0.72500	NI	NI
Pinvar	0.53100	NI	NI

Maximum-likelihood estimates for the selected reconstruction were obtained using the ML tree. Credible intervals were obtained from the stationary distribution. Abreviations: Llk, likelihood scores; AIC, Akaike Information Criterion; TL, tree length; r, reversible substitution rate between indicated bases; pi, stationary frequency of indicated base; alpha, shape parameter of the gamma distribution of rate variation; pinvar, proportion of invariable sites; ESS, estimated sample size; NI, parameter not included in selected refinement. For a complete description of the total stationary frequency of codons and the Effective Sample Size (ESS) of Bayesian analyses see ***Supplementary Table S7****.*

bermejo, and 3) *chato*, *hoja delgada*, and *pata de mula*. The relationships between *A. tequilana* cultivars and other *Agave* species, using SSAP of Ty1-*copia* retrotransposons, confirmed the close genetic distance between *azul*, *azul listado*, and *sigüín*, and revealed a close proximity between these varieties with a cultivated form of *A. rhodacantha* (var. *zopilote*), and a more distant relationship to *A. angustifolia*, *A. sisalana*, *A. americana*, and *A. filifera* (Bousios et al., 2007). In a study of three *Agave* species of the State of Jalisco (*A. angustifolia*, *A. tequilana*, and *A. rhodacantha*), where tequila is made, the use of SSR markers revealed that *A. angustifolia* populations from southern Jalisco are close to *A. tequilana* var. *azul*, whereas *A. tequilana* var. *sigüín* and *chato* are close to the cultivar *A. rhodacantha* var. *ixtlero amarillo* (Trejo et al., 2018). In a different study, and using AFLP markers, *A. tequilana* var. *azul* was found in close genetic proximity to *A. angustifolia* var. *espadín* from Oaxaca and to a cultivated form of *A. rhodacantha* also from Oaxaca (Rivera-Lugo et al., 2018). Based on SSR markers, *A. angustifolia* populations also show a large genetic diversity (Trejo et al., 2018).

Our study encompassed the most ample geographical distibution and the largest number of localities for both wild and cultivated accessions from *A. angustifolia* and *A. rhodacantha* used to date to derive *Agave* phylogenies (Rivera-Lugo et al., 2018; Trejo et al., 2018). This ensamble of sequences enabled a more comprehensive analysis of the genetic relationships between the *A. tequila*na varieties to members of the Rigidae group, represented in this work by *A. tequila*na, *A. angustifolia*,

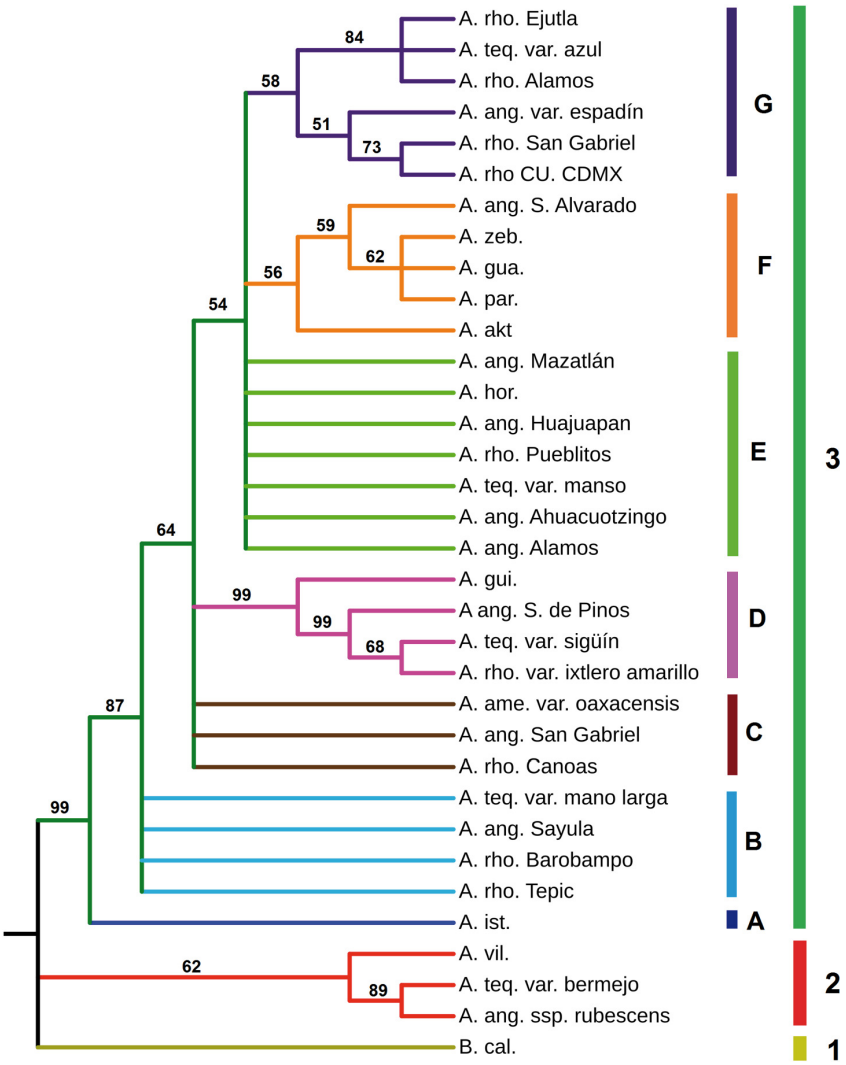


FIGURE 9 | Phylogenetic reconstruction derived from analyses of *Mayahuelin* ortholog gene sequences from Agavoideae. Ortholog mayahuelin sequences (see **Supplementary Table S6** and **Supplementary Figure S4**) were aligned and a phylogenetic reconstruction was obtained based on Bayesian inference algorithms, as described in "Materials and Methods" section. Posterior probability values are given for each branch of the tree. Abbreviations of species names are: A. akt. (*A. aktites*), A.ame. (*A. americana*), A.ang. (*A. angustifolia*), A.gua. (*A. guadalajarana*), A.hor. (*A. horrida*), A.gui. (*A. guiengola*), A.ist. (*A. isthmensis*), A.par. (*A. parryi*), A.rho. (*A. rhodacantha*), A.teq. (*A. tequilana*), A.vil. (*A.* vilmoriniana), A.zeb. (*A. zebra*), and B.cal. (*Beschorneria calcicola*). *Mayahuelin* Genbank accession numbers from all taxa analyzed in figure are indicated in **Supplementary Table S6**. Words after the species abbreviation refer to either the cultivar (i.e. azul) or the locatity of origin of the specimen (i.e. Alamos). Nucleotide sequence alignment and phylogenetic tree can be found at the TreeBASE website (http://purl.org/phylo/treebase/phylows/study/TB2:S25921).

A. rhodacantha, and *A. aktites*. BI algorithms very strongly support the idea that the Rigidae group is not monophyletic. Clearly, *A. tequilana* var. *azul* is close to *manso* and *sigüín* varieties (**Figure 9**) and agree with previous studies (Gil-Vega et al., 2006; Bousios et al., 2007). In addition, we found that *mano larga* is next of kin of *azul*, *manso*, and *sigüín*, whereas *bermejo* is very distant and closely related to *A. angustifolia* ssp. *rubescens* (**Figure 9** and **Supplementary Figure S9**, **Supplementary Information**). The large genetic distance between *bermejo* and other *A. tequilana* varieties was reported earlier (Gil-Vega et al., 2006). Our phylogeny also captured the close proximity between *A. tequilana* var. *azul* and *A. angustifolia* var. *espadín* from Oaxaca found by other authors using morphological or AFLP molecular markers (Gentry, 1982; Rivera-Lugo et al., 2018).

A. rhodacantha accessions from Sonora and Oaxaca showed the closest kinship to *A. tequilana* var. *azul* (from the Tequila region in Jalisco) (**Figure 9** and **Supplementary Figure S4**). Relative to *A. tequilana* var. *azul*, the percent identity of *Mayahuelin* gene from the former is 100% whereas that of the latter is 99.72%, although next in proximity to *A. tequilana* var. *azul*, according to BI and ML methods, are a wild *A. rhodacantha* accession from southern Jalisco, an ornamental *A. rhodacantha* of unknown origin (UNAM campus in Mexico

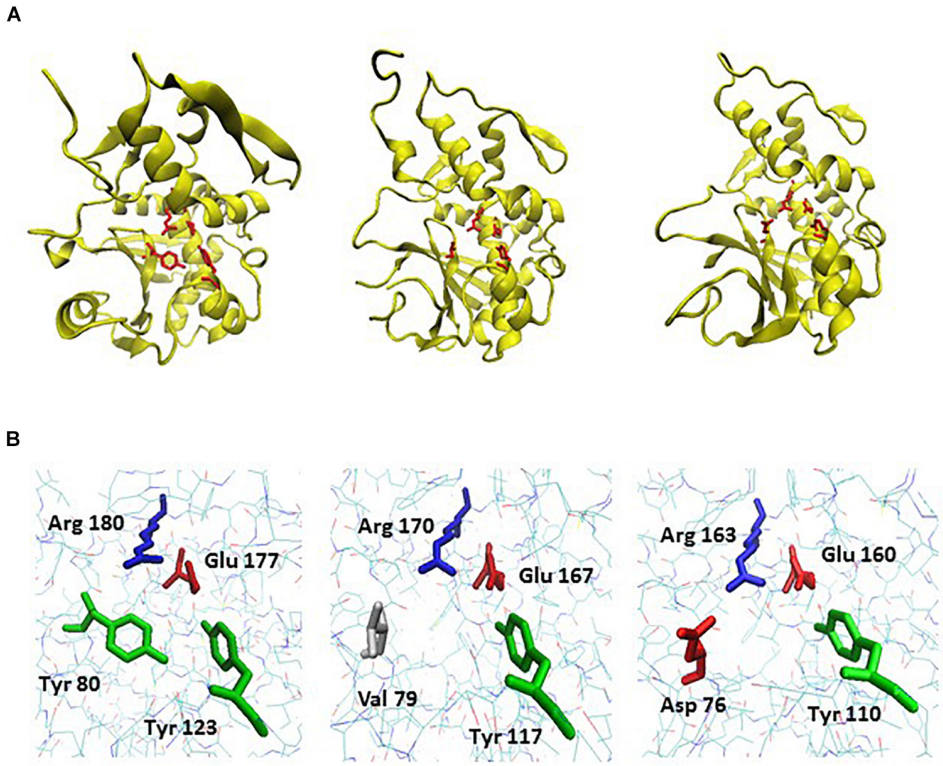


FIGURE 10 | A structural homology model for mayahuelin and its comparison to the crystal structural models of charybdin and ricin. **(A)** Backbone and ribbon diagrams of the proteins showing the side chains of invariable amino acids in their active sites. From left to right: crystal structure of ricin A chain at 1.8 °A [PDB: 1IFT], crystal structure of charybdin at 1.6 °A [PDB: 2B7U], and homology model of mayahuelin from *A. tequilana* var. *azul* (see "Materials and Methods" section). Catalytic amino acids are shown in red. **(B)** Wireframe diagrams of the same proteins focusing on active site region only. From left to right: ricin A-chain, charybdin, and mayahuelin. The position number and name of the four invariable residues of the active site are shown. The.pdb files were obtained from Protein Data Bank [https://www.rcsb.org]. Homology model for mayahuelin was obtained as described in "Materials and Methods" section using charybdin structure as a template. Images were edited with VMD v1.9.3 visualization program.

City), and *A. angustifolia* var. *espadin* from Oaxaca with percent identities relative to *A. tequilana* var. *azul* of 99.71, 99.71, and 96.65%, respectively. Nonetheless, it is striking the larger genetic distance between *A. tequilana* var. *azul* and its neighboring populations of *A. rhodacantha* from Nayarit and southern Jalisco. Similarly, *A. angustifolia* var. *espadín* from Oaxaca is closer to *A. rhodacantha* accessions from southern Jalisco than those from Oaxaca whereas *A. tequilana* var. *bermejo* is closer to *A. angustifolia* ssp. *rubescens* from Guerrero (**Figure 9**). This was not the case for *A. tequilana* var. *manso*, *sigüín*, and *mano larg*a that showed close kinship to agaves from southern Jalisco, according to BI (**Figure 9**). These results contribute to the conservation genetics of agaves used for tequila and mescal production, as they represent the first step toward the genetic identification of natural populations of their wild ancestors. One of the weaknesses of this work resides in the use of only one accession (living specimen) from each population analyzed, which does not allow to assess their genetic diversity. Therefore, future studies on the origin of domesticated agaves should include a larger number of individuals from each of these populations to have a more statistically significant estimation of their identity relative to their closest wild relatives. The domesticated alleles could be rare alleles in other wild populations that, on the other hand, may display similarities at different levels of complexity. Furthermore, it is believed that the success of the genus *Agave* is due in part to the high frequency of polyploid species within the group. Agaves have a basic chromosome number = 30 and euploid series of 2n, 3n, 4n, 5n, and 6n have been reported (Granick, 1944). Thus, the reconstruction of *Agave* phylogenies based on few or single genetic markers, such as *Mayahuelin*, could underestimate their full ancestry, especially in cases where allopolyploid speciation events occurred. Therefore, our results are indicative, but no fully conclusive, of the origin of *A. angustifolia*, *A. tequilana*, and *A. rhodacantha* cultivars.

A full understanding of the disjoint distributions in the phylogeographic pattern between these cultivars and their closest wild relatives requires both ecological and cultural elements for discussion. Although bats are long-distance flyers and agave seed set is dependent on bat pollination to achieve full potential (Howell and Roth, 1981) agave seed dispersal depends mostly on wind and water (Sánchez-Salas et al., 2017; Lindsay et al., 2018). During the warmest half of the year dominant winds in Mexico move with a NE to SW direction. During the coldest

half of the year dominant winds in northern Mexico blow from the W, while maintaining a NE to SW direction in southern Mexico and occasionally receiving northern winds from the Gulf of Mexico (García, 2003). Thus, prevailing wind patterns in Mexico would not favor connectivity between Sonora, Jalisco, and Oaxaca. However, from a theoretical perspective, wind dispersal of agave seed from Oaxaca to Sonora, but not the opposite, is possible since, during the summer and autumm, hurricanes run parallel to the coasts of Mexico moving on a SE to NW direction, ocassionaly entering inland (García, 2003; Rosengaus Moshinsky et al., 2002). Thus, it is theoreticaly unlikely the dispersion of agave seed from Sonora to Oaxaca, especially if we consider the large geographical distance between these states (**Supplementary Figure S1**). Assuming long distance wind dispersal of agave seeds, germination and establishment of agave seedlings in natural habitats is extremely low (Nobel, 1992; Arizaga and Ezcurra, 2002) making even more difficult to explain the observed disjoint distribution by wind seed dispersal. Alternatively, disjoint distributions and presence of plants far outside their natural range can be explained if we consider the long history of plant cultivation by human cultures. Because of their small size and weight, *Agave* seeds and bulbils could have been dispersed by humans along ancient trade routes. Such long distance trade was intense in Mesoamerica (present day Mexico and Guatemala) and North America (i.e. Mississippian culture of southeastern and midwestern United States) from around 1600 BCE and between 1000 and 1550 CE, respectively, Smith (2010). Moreover, cultural connections between Mesoamerica and Sonora exist since prehispanic times (Watson and García, 2016) and pinpoint Western Mexico (Jalisco, Nayarit, Colima, and Michoacán) as a macro-regional economy connecting Mesoamerica and the United States Southwest by both coastal and highland routes (Wilcox et al., 2008). Evidence for diffusion of plants along these routes is better understood in maize, whose introduction to the United States Southwest from Mesoamerica began around 2000 BCE, via the highland route, and 2000 years ago, via the coastal route (da Fonseca et al., 2015). In the case of some *Agave* species, ancient traces of their cultivation by Native Americans have been recognized based on genetic structure and the patchy distribution of colonies typically found in close proximity to prehistoric settlements (Minnis and Plog, 1976; Gentry, 1982; Parker et al., 2010; Lindsay et al., 2018). Historical accounts report the movement of *A. salmiana* and *A. americana* from central- (Tlaxcala) to northern- Mexico (Saltillo and Durango), right after the conquest of Mexico by the Spaniards, as thousands of náhuatl-speaking peoples, mainly Tlaxcaltecs, colonized the region bringing with them *maguey* to maintain their deep-rooted tradition of *pulque* production (Gentry, 1982). During the XVII century (*circa* 1621) agaves were already under cultivation in the Tequila region of Jalisco (Valenzuela Zapata, 1995) although historical accounts of their origin are still unknown. It remains to be studied how *A. angustifolia* var. *espadin*, *A. tequilana* var. *azul* and *A. rhodacantha* from Alamos, Sonora are interconnected and a more in depth study of natural populations of the Rigidae group in the Pacific coast of Mexico is needed.

Variation of some characters within the *A. angustifolia complex* is so large that separation at the species level is very difficult (Gentry, 1982). Our work has provided evidence for the ample genetic diversity within the *A. angustifolia* complex (**Figure 9**) and agrees with previous studies (Rivera-Lugo et al., 2018). We have also provided novel evidence for the large genetic diversity within the *A. rhodacantha* complex and the polyphyletic origin of Rigidae whose members are distributed in seven subclades/clades (**Figure 9**).

The *A. angustifolia* complex has the most ample distribution of agaves in North-America: inhabits diferent plant communities (thorn forests, tropical savannah, and drought-deciduous tropical forests) between sea level to near 1,600 m. a.s.l. It is possible that the long, narrow, rigid leaves typical of *A. angustifolia* and *A. rhodacantha* represent an ecomorphological response of different species to similarities in microenvironmental conditions across their distribution range (i.e. aridity, dew formation at night, etc.), causing confusion on the identification of phylogenetic signals based on pure morphological characters. The phenotypic similarity observed across independent lineages of closely related species has lead to the concept of phylogenetic niche conservatism (PNC) (Bravo et al., 2014). PNC has been observed in different taxa of the plant and animal kingdoms (Losos, 2008) and the *A. angustifolia* and *A. rhodacantha* complexes could be experiencing this process. In agaves and other xerophytic rosette plants, long-narrow leaf morphologies are very efficient for fog-harvesting (Martorell and Ezcurra, 2007). It remains to be studied if, in the *A. angustifolia* and *A. rhodacantha* complexes, such traits relate to their ecological niches along the Pacific and Atlantic coasts of Mexico and Central America. The phenotypic similarity could be the result of similar changes at the genetic level, as leaf growth and morphogenesis proceed through conserved genetic mechanisms (Nelissen et al., 2016). However, it can be due to other mechanisms such as gene flow among taxa, genetic drift, etc. (Bravo et al., 2014). Gentry (Gentry, 1982) also pointed out the dificulty of delimiting species between *A. angustifolia* and some A. *rhodacantha* populations due to overlaps in leaf characters (length, width). As these morphological characters seem to offer little taxonomic value within the Rigidae group in *Agave*, the finding of new diagnostic characters at different levels of complexity, among them DNA sequence markers, and a new taxonomic revision, are pending.

Finally, it is quite evident the hard polytomy at different levels of the tree from a large number of of *A. rhodacantha* and *A. angustifolia* accessions studied (**Figure 9**). Further work is needed to asses if this polytomy is an artifact caused by the lack of phylogenetic resolution of the genetic marker used or whether it represents a true phylogenetic radiaton pattern due perhaps to the great degree of artificial selection and geographical movement caused by *Homo sapiens*.

Possible Causes for Active Site Substitutions in Mayahuelin Within Agavoideae

Are substitutions in the active site of mayahuelin orthologs within Agavoideae the result of domestication, functional specialization or simple evolutionary adaptations for defense against particular viruses, microbes or hervibores? The evidence provided to date is too preliminary to answer with certainty any of these questions. Out of the 34 taxa studied, 13 are cultivated and

21 come from wild populations. The frequency for both Y/Y homocigocity and homozigotic substitutions (D/D and S/S) at aa position 76 was higher in wild accessions relative to cultivars (**Supplementary Figure S10** and **Supplementary Table S5**). In contrast, heterozigocity at position 76 (Y/D, Y/D, and Y/S) was more frequent in cultivars (**Supplementary Figure S10** and **Supplementary Table S5**). Examples of this bias between the Y/Y homocigocity at position 76 and species that have no record of cultivation or human utilization are *A. horrida* and *A. guiengola* as well as wild forms of cultivated species such as *A. angustifolia* and *A. rhodacantha*. Other taxa with a long history of cultivation or utilization as a natural resource since historical or pre-historical times displayed homocigous D/D (*A. parryi, A. vilmoriniana,* and *A. zebra*) or S/S (only found in *A. angustifolia* ssp. *rubescens*) substitutions. The only exceptions of cultivated forms with canonical Y/Y homocigocity at position 76 are *A. rhodacantha* var. *ixtlero amarillo* and *A. tequilana* var. *mano larga* that are used for fiber and mescal production, respectively. In opposition, *A. isthmensis* is a wild species with no history of cultivation or utilization by human beings and is heterozygous for Y/D76. The evolutionary significance of these substitutions should become more clear once a comparative functional study of mayahuelin proteins containing each of the allelic variations is performed.

Our observations could raise the interest in Agavoideae to become great models for in-depth studies on the evolution and functional/structural analysis of RIPs. Moreover, *Mayahuelin* sequences are promising to reconstruct reliable phylogenies within Agavoideae and they could complement information derived using other genetic markers, such as the chloroplast genes or *ITS* sequences used so far.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repositories and accession numbers can be found in **Figure 6** (mass spectrometry) and **Figure 9** (phylogenetic information) of the article as well as in **Supplementary Tables S4, S7** of **Supplementary Material** (DNA sequences).

AUTHOR CONTRIBUTIONS

FL identified and developed purification protocols for the biochemical characterization of mayahuelin and for the isolation of anti-mayahuelin antibodies and prepared figures under supervision of GC. SR, JG, and ES designed and performed all experiments for evaluation of the cytotoxic effects of mayahuelin in yeast. JG and JN-S carried all phylogenetic analyses, performed homology modeling, prepared figures and tables, and went on field trips for Agave collection. JG also evaluated effects of mayahuelin on wheat germ system under supervision of TD. FH-B evaluated levels of mayahuelin protein in tissues from different Agave species and prepared figures. AM-H constructed and characterized EST libraries from *A. tequilana* and evaluated mayahuelin transcript levels by RT-PCR. AG-M is the curator of the Agavaceae Collection of IB-UNAM Botanical Garden that provided most Agave tissues, went on field trips, and helped in their taxonomic identification. JN-S conceived the project, analyzed data, supervised all author's work, and wrote the texts. All authors contributed to the manuscript.

FUNDING

This work was supported by research grants from PAPIIT/DGAPA/UNAM IN214119 to GIC, PAPIIT/DGAPA/UNAM IN215120 to JN-S, and CONACYT-PN247732 grant to GC and JN-S. Funding for the payment of publication fees for this manuscript was received from Instituto de Biología, Universidad Nacional Autónoma de México.

ACKNOWLEDGMENTS

We express our appreciation to Idalia Rojas Barrera, Yajima Osorno, Viridiana Rivas, and Olivia Cabanillas, for technical help and to Maria Eugenia Campos and Luz María Rangel for the managerial assistance of the project. We thank Dr. June Simpson from CINVESTAV-Irapuato for sharing unpublished EST sequence data from *A. tequilana* var. *azul*. We praise Dr. César Batista and Erika Meneses Romero from the *Unidad de Proteómica*, *Instituto de Biotecnología*, *Universidad Nacional Autónoma de México*, and the late Dr. Guillermo Mendoza-Hernández, from the Biochemistry Department, *Facultad de Medicina*, *Universidad Nacional Autónoma de México*, for their fine work in determining mayahuelin amino acid sequence by LC-MS/MS and Edman degradation methods, respectively. We thank *Laboratorio de Secuenciación Genómica de la Biodiversidad y de la Salud* at Instituto de Biología, UNAM and *Unidad de Síntesis y Secuenciación de ADN* at Instituto de Biotecnología, UNAM for DNA sequencing. We thank Elizabeth Mata in charge of Instituto de Biotecnología's-UNAM Bioterium. We are indebted to Daniel Sandoval for his help in the collection and identification of *Agave* specimens in the field and Ivonne Olalde for her support in taking care of our *Agave* specimens in the greenhouse. We appreciate the support of Dr. Ignacio del Real, Ing. Ramón Rubio, and Ing. Rafael Ramos for donating ramets and leaf samples from different *A. tequilana* cultivars from the Botanical Garden at Casa Sauza. We thank Dr. Miguel Ángel Gruintal and Samyr Corona for their generous donation of leaf samples from three *A. angustifolia* and one *A. horrrida* specimens from their *Agave* collection. We also thank Francisco Javier Martínez por aid on the identification of *A. tequilana* cultivars.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00573/full#supplementary-material

REFERENCES

Abascal, F., Zardoya, R., and Telford, M. J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* 38, 7–13. doi: 10.1093/nar/gkq291

Arizaga, S., and Ezcurra, S. (2002). Propagation mechanisms in *Agave macroacantha* (Agavaceae), a tropical arid-land succulent rosette. *Am. J. Bot.* 89, 632–641. doi: 10.3732/ajb.89.4.632

Bagga, S., Seth, D., and Batra, J. (2003). The cytotoxic activity of ribosome-inactivating protein Saporin-6 is attributed to its rRNA N-glycosidase and internucleosomal DNA fragmentation activities. *J. Biol. Chem.* 278, 4813–4820. doi: 10.1074/jbc.M207389200

Barbieri, L., Valbonesi, P., Bonora, E., Gorini, P., Bolognesi, A., and Stirpe, F. (1997). Polynucleotide: adenosine glycosidase activity of ribosome-inactivating proteins: effect on DNA, RNA and poly(A). *Nucleic Acids Res.* 25, 518–522. doi: 10.1093/nar/25.3.518

Barreto, R., Nieto-Sotelo, J., and Cassab, G. I. (2010). Influence of plant growth regulators and water stress on ramet induction, rosette engrossment, and fructan accumulation in *Agave tequilana* Weber var. Azul. *Plant Cell Tissue Organ. Cult.* 103, 93–101.

Berger, S., Mitchell-Olds, T., and Stotz, H. U. (2002). Local and differential control of vegetative storage protein expression in response to hervibore damage in *Arabidopsis thaliana*. *Physiol. Plant.* 114, 85–91. doi: 10.1046/j.0031-9317.2001.1140112.x

Bogler, D. J., Pires, J. C., and Francisco-Ortega, J. (2006). Phylogeny of Agavaceae based on nndhF, rbcL, and ITS sequences: implications of molecular data for classification. *Aliso* 22, 313–328.

Bogler, D. J., and Simpson, B. B. (1996). Phylogeny of Agavaceae based on ITS rDNA sequence variation. *Am. J. Bot.* 83, 1225–1235.

Bonnes, M. S., Ready, M. P., Irvin, J. D., and Marby, T. J. (1994). Pokeweed antiviral protein inactivates pokeweed ribosomes; implications for the antiviral mechanism. *Plant J.* 5, 173–183. doi: 10.1046/j.1365-313x.1994.05020173.x

Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4:e1660. doi: 10.7717/peerj.1660

Bousios, A., Saldana-Oyarzabal, I., Valenzuela-Zapata, A. G., Wood, C., and Pearce, S. R. (2007). Isolation and characterization of Ty1-copia retrotransposon sequences in the blue agave (*Agave tequilana* Weber var. azul) and their development as SSAP markers for phylogenetic analysis. *Plant Sci.* 172, 291–298.

Bravo, G. A., Remsen, J. V. Jr., and Brumfield, R. T. (2014). Adaptive processes drive ecomorphological convergent evolution in antwrens (Thamnophilidae). *Evolution* 68, 2757–2774. doi: 10.1111/evo.12506

Castetter, E. F., Bell, W. H., and Grove, A. R. (1938). The early utilization and the distribution of agave in the American southwest. *Univ. N. Mex. Biol. Ser.* 5:335.

Chase, M. W., Reveal, J. L., and Fay, M. F. (2009). A subfamilial classification for the expanded asparagalean families Amaryllidaceae, Asparagaceae and Xanthorrhoeaceae. *Bot. J. Linn. Soc.* 161, 132–136.

da Fonseca, R., Smith, B., Wales, N., Cappellini, E., Skoglund, P., Fumagalli, M., et al. (2015). The origin and evolution of maize in the Southwestern United States. *Nat. Plants* 1:14003. doi: 10.1038/nplants.2014.3

Day, P., Ernst, S., Frankel, A., Mozingo, A., Pascal, M., and Robertus, J. (1996). Structure and activity of an active site substitution of ricin A chain. *Biochemistry* 35, 11098–11103. doi: 10.1021/bi960880n

De Virgilio, M., Lombardi, A., Caliandro, R., and Fabbrini, M. (2010). Ribosome-inactivating proteins: from plant defense to tumor attack. *Toxins* 2, 2699–2737. doi: 10.3390/toxins2112699

Di Maro, A., Citores, L., Russo, R., Iglesias, R., and Ferreras, J. M. (2014). Sequence comparison and phylogenetic analysis by the Maximum Likelihood method of ribosome-inactivating proteins from angiosperms. *Plant Mol. Biol.* 85, 575–588.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–97. doi: 10.1093/nar/gkh340

Eguiarte, L. E., Aguirre-Planter, E., Aguirre, X., Colín, R., González, A., Rocha, M., et al. (2013). From isozymes to genomics: population genetics and conservation of *Agave* in México. *Bot. Rev.* 79, 483–506. doi: 10.1007/s12229-013-9123-x

Escamilla-Treviño, L. L. (2011). Potential of plants from the genus Agave as bioenergy crops. *Bioenerg. Res.* 5, 1–9.

Fernández, R., Kallal, R. J., Dimitrov, D., Ballesteros, J. A., Arnedo, M. A., Giribet, G., et al. (2018). Phylogenomics, diversification dynamics, and comparative transcriptomics across the spider tree of life. *Curr. Biol.* 28, 1489–1497. doi: 10.1016/j.cub.2018.06.018

Ferreras, J. M., Barbieri, L., Girbés, T., Batelli, M. G., Rojo, A., Arias, F. J., et al. (1993). Distribution and properties of major ribosome-inactivating proteins (28 S rRNA N-glycosidases) of the plant *Saponaria officinalis* L. (Caryophyllaceae). *Biochim. Biophys. Acta* 1216, 31–42. doi: 10.1016/0167-4781(93)90034-b

Fuchs, H. (2019). Dianthin and its potential in targeted tumor therapies. *Toxins* 11:592. doi: 10.3390/toxins11100592

García, E. (2003). Distribución de la precipitación en la República Mexicana. *Bol. Inst. Geog. Univ. N.* 50, 67–76.

García-Mendoza, A. (1998). *Con Sabor a Maguey. Guía de la Colección Nacional de Agaváceas y Nolináceas del Jardín Botánico, Instituto de Biología - UNAM*. Mexico: UNAM-SIGSA.

García-Mendoza, A. J., and Chávez-Rendón, C. (2013). Agave kavandivi (Agavaceae: grupo Striatae), una especie nueva de Oaxaca, México. *Rev. Mex. Biodivers.* 84, 1070–1076. doi: 10.7550/rmb.35241

Gentry, H. S. (1982). *Agaves of continental North America*. Tucson, AZ: University of Arizona Press.

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

Gil-Vega, K., González-Chavira, M., Martínez de la Vega, O., Simpson, J., and Vandemark, G. (2001). Analysis of genetic diversity in *Agave tequilana* var. Azul using RAPD markers. *Euphytica* 119, 335–341.

Good-Avila, S., Souza, V., Gaut, B. S., and Eguiarte, L. (2006). Timing and rate of speciation in Agave (Agavaceae). *Proc. Natl. Acad. Sci. U.S.A.* 103, 9124–9129. doi: 10.1073/pnas.0603312103

Granick, E. B. (1944). A karyosystematic study of the genus Agave. *Am. J. Bot.* 31, 283–298.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate Maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Harley, S. M., and Beevers, H. (1982). Ricin inhibition of in vitro protein synthesis by plant ribosomes. *Proc. Natl. Acad. Sci. U.S.A.* 79, 5935–5938. doi: 10.1073/pnas.79.19.5935

Hernández-Valdepeña, M. A., Pedraza-Chaverri, J., Gracia-Mora, I., Hernández-Castro, R., Sánchez-Bartez, F., Nieto-Sotelo, J., et al. (2016). Suppression of the tert-butylhydroquinone toxicity by its grafting onto chitosan and further cross-linking to agavin toward a novel antioxidant and prebiotic material. *Food Chem.* 199, 485–491. doi: 10.1016/j.foodchem.2015.12.042

Heyduk, K., McKain, M. R., Lalami, F., and Leebens-Mack, J. (2016). Evolution of a CAM anatomy predates the origins of Crassulacean acid metabolism in the Agavoideae (Asparagaceae). *Mol. Phylogenet. Evol.* 105, 102–113. doi: 10.1016/j.ympev.2016.08.018

Howell, D. J., and Roth, B. S. (1981). Sexual reproduction in agaves: the benefits of bats; the cost of semelparous advertising. *Ecology* 62, 1–7.

Huang, X., Wang, B., Xi, J., Zhang, Y., He, C., Zheng, J., et al. (2018).). Transcriptomic comparison reveals distinct selection patterns in domesticated and wild Agave species, the important CAM plants. *Int. J. Genomics* 2018:5716518. doi: 10.1155/2018/5716518

Hur, Y., Hwang, D., Zoubenko, O., Coetzer, C., and Uckun, F. (1995). Isolation and characterization of pokeweed antiviral protein mutations in Saccharomyces cerevisiae: identification of residues important for toxicity. *Proc. Natl. Acad. Sci. U.S.A.* 92, 8448–8452. doi: 10.1073/pnas.92.18.8448

Jiang, S., Bhalla, R., Ramamoorthy, R., Luan, H., Venkatesh, P., Cai, M., et al. (2012). Over-expresion of OSRIP18 increases drought and salt tolerance in transgenic rice plants. *Transgenic Res.* 21, 785–795. doi: 10.1007/s11248-011-9568-9

Kim, Y., and Robertus, J. (1992). Analysis of several key active site residues of ricin A chain by mutagenesis and X ray crystallography. *Protein Eng.* 5, 775–779. doi: 10.1093/protein/5.8.775

Korchowiec, B., Gorczyca, M., Wojszko, K., Janikowska, M., Henry, M., and Rogalska, E. (2015). Impact of two different saponins on the organization of model lipid membranes. *Biochim. Biophys. Acta* 1848, 1963–1973. doi: 10.1016/j.bbamem.2015.06.007

Kumar, D., Verma, H. N., Tuteja, N., and Tewari, K. K. (1997). Cloning and characterisation of a gene encoding an antiviral protein from *Clerodendrum aculeatum* L. *Plant Mol. Biol.* 33, 745–751. doi: 10.1023/a:1005716103632

Lefort, V., Longueville, J. E., and Gascuel, O. (2017). SMS: Smart Model Selection in PhyML. *Mol. Biol. Evol.* 34, 2422–2424. doi: 10.1093/molbev/msx149

Lemmon, E. M., and Lemmon, A. R. (2013). High-throughput genomic data in systematics and phylogenetics. *Annu. Rev. Ecol. Evol. Syst.* 44, 99–121.

Li, X., Baricevic, M., Saidasan, H., and Tumer, N. (2007). Ribosome depurination is not sufficient for ricin-mediated cell death in Saccharomyces cerevisiae. *Infect. Immun.* 75, 417–428. doi: 10.1128/IAI.01295-06

Lindsay, D. L., Swift, J. L., Lance, R. F., and Edwards, C. E. (2018). A comparison of patterns of genetic structure in two co-occurring Agave species (Asparagaceae) that differ in the patchiness of their geographical distributions and cultivation histories. *Bot. J. Linn. Soc.* 186, 361–373.

Lledías, F., Hernández, F., Rivas, V., García-Mendoza, A., Cassab, G. I., and Nieto-Sotelo, J. (2017a). A rapid and reliable method for total protein extraction from succulent plants for proteomic analysis. *Protein J.* 36, 308–321. doi: 10.1007/s10930-017-9720-3

Lledías, F., Hernández, F., Rivas, V., García-Mendoza, A., Cassab, G. I., and Nieto-Sotelo, J. (2017b). Erratum to: a rapid and reliable method for total protein extraction from succulent plants for proteomic analysis. *Protein J.* 36:523. doi: 10.1007/s10930-017-9739-5

Losos, J. B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol. Lett.* 11, 995–1007. doi: 10.1111/j.1461-0248.2008.01229.x

Luján, R., Lledías, F., Martínez, L., Barreto, R., Cassab, G., and Nieto-Sotelo, J. (2009). Small heat-shock proteins and leaf cooling capacity account for the unusual heat tolerance of the central spike leaves in *Agave tequilana* var. Weber. *Plant Cell Environ.* 32, 1791–1803. doi: 10.1111/j.1365-3040.2009.02035.x

Martínez-Hernández, A., Mena-Espino, M. E., Herrera-Estrella, A. H., and Martínez-Hernández, P. (2010). Construcción de bibliotecas de ADNc y análisis de expresión génica por RT-PCR en agaves. *Rev. Latinoamer. Quím.* 38, 21–44.

Martorell, C., and Ezcurra, E. (2007). The narrow-leaf syndrome: a functional and evolutionary approach to the form of fog-harvesting rosette plants. *Oecologia* 151, 561–573. doi: 10.1007/s00442-006-0614-x

May, K., Yan, Q., and Nilgun, E. T. (2013). Targetin ricin to the ribosome. *Toxicon* 69, 143–151.

Minnis, P. E., and Plog, S. E. (1976). A Study of the site specific distribution of *Agave parryi* in east central Arizona. *Kiva* 41, 299–308.

Monzingo, A., and Robertus, J. (1992). X-ray analysis of substrate analogs in the ricin A-chain active site. *J. Mol. Biol.* 227, 1136–1145. doi: 10.1016/0022-2836(92)90526-p

Nelissen, H., Gonzalez, N., and Inze', D. (2016). Leaf growth in dicots and monocots: so different yet so alike. *Curr. Opin. Plant Biol.* 33, 72–76. doi: 10.1016/j.pbi.2016.06.009

Nobel, P. S. (1988). *Environmental Biology of Agaves and Cacti*. Cambridge: Cambridge University Press.

Nobel, P. S. (1992). Annual variations in flowering percentage, seedling establishment, and ramet production for a desert perennial. *Int. J. Plant Sci.* 153, 102–107.

Parker, K. C., Trapnell, D. W., Hamrick, J. L., Hodgson, W. C., and Parker, A. J. (2010). Inferring ancient Agave cultivation practices from contemporary genetic patterns. *Mol. Ecol.* 19, 1622–1637. doi: 10.1111/j.1365-294X.2010.04593.x

Peumans, W., Hao, Q., and Van Damme, J. (2001). Ribosome-inactivating proteins from plants: more then RNA N-glycosidases. *FASEB J.* 15, 1493–1506. doi: 10.1096/fj.00-0751rev

Puri, M., Kaur, I., Perugini, A., and Gupta, R. (2012). Ribosome-inactivating proteins: current status and biomedical applications. *Drug Discov. Today* 17, 774–783. doi: 10.1016/j.drudis.2012.03.007

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

Ready, M., Kim, Y., and Robertus, J. (1991). Site-directed mutagenesis of ricin A-chain and implications for the mechanism of action. *Proteins* 10, 270–278. doi: 10.1002/prot.340100311

Reisbig, R. R., and Bruland, Ø. S. (1983). Dianthin 30 and 32 from *Dianthus caryophyllus*: two inhibitors of plant protein synthesis and their tissue distribution. *Arch. Biochem. Biophys.* 224, 700–706. doi: 10.1016/0003-9861(83)90258-8

Rivera-Lugo, M., García-Mendoza, A., Simpson, J., Solano, E., and Gil-Vega, K. (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 Syst. Evol.* 304, 969–979.

Rojas-Martínez, A., Valiente-Banuet, A., Arizmendi, M. C., Alcántara-Eguren, A., and Arita, H. T. (1999). Seasonal distribution of the long-nosed bat (Leptonycteris curasoae) in North America: does a generalized migration pattern really exist? *J. Biogeogr.* 26, 1065–1077.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Rosengaus Moshinsky, M., Jiménez, E., and Vázquez Conde, M. T. (2002). *Atlas Climatológico de Ciclones Tropicales en México*. México: Centro Nacional de Prevención de Desastres.

Sánchez-Salas, J., Flores, J., Jurado, E., Sáenz-Mata, J., Orozco-Figueroa, P., and Muro Pérez, G. (2017). Hydrochory in seeds of Agave victoriae-reginae T. Moore endangered species: morphology and anatomy as facilitators of hydro-dispersion and germination. *Gayana Bot.* 74, 251–261.

Santos-Zea, L., Leal-Díaz, A. M., Cortés-Ceballos, E., and Gutiérrez-Uribe, J. A. (2012). Agave (Agave spp.) and its traditional products as a source of bioactive compounds. *Curr. Bioac. Comd.* 8, 218–231.

Savino, C., Federici, L., Ippoliti, R., Lendaro, E., and Tsernoglou, D. (2000). The crystal structure of saporin SO6 from Saponaria officinalis and its interaction with the ribosome. *FEBS Lett.* 470, 239–243. doi: 10.1016/s0014-5793(00)01325-9

Shen, M., and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Sci.* 15, 2507–2524. doi: 10.1186/1471-2105-15-307

Simpson, J., Martínez-Hernández, A., Abraham Juárez, M. J., Delgado Sandoval, S., Sánchez Villarreal, A., and Cortés Romero, C. (2011). Genomic resources and transcriptome mining in *Agave tequilana*. *GCB Bioenergy* 3, 25–36. doi: 10.1111/j.1757-1707.2010.01079.x

Smith, M. E. (2010). "Trading patterns, ancient American," in *The Berkshire Encyclopedia of World History*, 2nd Edn, ed. W. H. McNeill, (Great Barrington, MA: Berkshire Publishing Group).

Spackova, N., and Sponer, J. (2006). Molecular dynamics simulations of sarcin-ricin rRNA motif. *Nucleic Acids Res.* 34, 697–708. doi: 10.1093/nar/gkj470

Staswick, P. E. (1990). Novel regulation of vegetative storage protein genes. *Plant Cell* 2, 1–6. doi: 10.1105/tpc.2.1.1

Stirpe, F. (2013). Ribosome-inactivating proteins: from toxins to useful proteins. *Toxicon.* 67, 12–16. doi: 10.1016/j.toxicon.2013.02.005

Stirpe, F., Barbieri, L., Gorini, P., Valbonesi, P., Bolognesi, A., and Polito, L. (1996). Activities associated with the presence of ribosome-inactivating proteins increase in senescent and stressed leaves. *FEBS Lett.* 382, 309–312. doi: 10.1016/0014-5793(96)00188-3

Stirpe, F., and Batelli, M. (2006). Ribosome-inactivating proteins: progress and problems. *Cell. Mol. Life Sci.* 66, 1850–1866. doi: 10.1007/s00018-006-6078-7

Szewcsak, A., and Moore, P. (1995). The sarcin/ricin Loop, a modular RNA. *J. Mol. Biol.* 247, 81–98. doi: 10.1006/jmbi.1994.0124

Tena Martínez, R. (ed.) (2002). "Mitos e historias de los antiguos nahuas," in *Consejo Nacional Para la Cultura y las Artes, Dirección General de Publicaciones* (Mexico: Consejo Nacional para la Cultura y las Artes).

The Angiosperm Phylogeny Group, (2009). An update of the Angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121. doi: 10.1016/j.jep.2015.05.035

Torres, I., Casas, A., Vega, A., Martínez-Ramos, M., and Delgado-Lemus, A. (2015). Population dynamics and sustainable management of mescal agaves in central Mexico: Agave potatorum in the Tehuacán valley. *Econ. Bot.* 69, 26–41. doi: 10.1186/1746-4269-10-63

Touloupakis, E., Gessmann, R., Kavelaki, K., Christofakis, E., and Petratos, K. (2006). Isolation, characterization, sequencing and cristal structure of charybdin, a type 1 ribosome-inactivating protein from *Charybdis maritima* agg. *FEBS J.* 273, 2684–2692. doi: 10.1111/j.1742-4658.2006.05287.x

Trejo, L., Limones, V., Peña, G., Scheinvar, E., Vargas-Ponce, O., Zizumbo-Villarreal, D., et al. (2018). Genetic variation and relationships among agaves

related to the production of Tequila and Mezcal in Jalisco. *Ind. Crop. Prod.* 125, 140–149.

Valenzuela Zapata, A. (1995). La agroindustria del agave tequilero *Agave tequilana* Weber. *Bol. Soc. Bot. Méx.* 57, 15–25.

Van Damme, E. J. M., Barre, A., Barbieri, L., Valbonesi, P., Rouge, P., Van Leuven, F., et al. (1997). Type 1 ribosome-inactivating proteins are the most abundant proteins in iris (Iris hollandica var. Professor Blaauw) bulbs: characterization and molecular cloning. *Biochem. J.* 324, 963–970. doi: 10.1042/bj3240963

Watson, J. T., and García, C. (2016). Postclassic expansion of Mesoamerican biocultural characteristics into Sonora, Mexico. *J. Field Archaeol.* 41, 222–235.

Wilcox, D. R., Wiegand, P. C., Wood, J. S., and Howard, J. B. (2008). Ancient cultural interplay of the American Southwest in the Mexican Northwest. *J. Southwest* 50, 103–206.

Zhang, F., Ding, Y., Zhu, C.-D., Zhou, X., Orr, M. C., Scheu, S., et al. (2018). Phylogenomics from low-coverage whole-genome sequencing. *Methods Ecol. Evol.* 10, 507–517.

Zhu, F., Yuan, S., Zhang, Z., Qian, K., Feng, J., and Yang, Y. (2016). Pokeweed antiviral protein (PAP) increases plant systemic resistance to Tobacco mosaic virus infection in *Nicotiana bethamiana*. *Eur. J. Plant Pathol.* 146, 541–549.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 28 May 2020
doi: 10.3389/fpls.2020.00685

Hybridization Between Yuccas From Baja California: Genomic and Environmental Patterns

Maria Clara Arteaga[1], Rafael Bello-Bedoy[2] and Jaime Gasca-Pineda[1,3]*

[1] Departamento de Biología de la Conservación, Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), Ensenada, Mexico, [2] Facultad de Ciencias, Universidad Autónoma de Baja California, Ensenada, Mexico, [3] Unidad de Biotecnología y Prototipos (UBIPRO), Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Mexico City, Mexico

Hybridization can occur when two geographically isolated species are reproductively compatible and have come into sympatry due to range shifts. Yucca and yucca moths exhibit obligate pollination mutualism; yucca moths are responsible for the gene flow mediated by pollen among yucca populations. In the Baja California Peninsula, there are two yucca sister species, *Y. capensis* and *Y. valida*, that have coevolved with the same pollinator, *Tegeticula baja*. Both yucca species are endemic to the peninsula, and their current distributions are allopatric. Based on their morphological characteristics, it has been suggested that some plants growing in the southern part of the Magdalena flatland, a spatially disjunct part of *Yucca valida*'s range, have hybrid origins. We conducted genomic and climatic analyses of the two yucca species as well as the putative hybrid populations. We genotyped 3,423 single nucleotide polymorphisms in 120 individuals sampled from 35 localities. We applied Bayesian tests and geographic cline analyses to the genomic data. Using climatic information from the occurrence sites, we projected species distribution models in different periods to assess changes in the distributional range, and we performed a statistical test to define the niche divergence between the paternal species and the putative hybrid populations. Structure analysis revealed mixed ancestry in the genome of hybrid populations, and the Bayesian models supported a scenario of post-divergence gene flow between the yucca species. Our species distribution models reveal that the geographical ranges of the parental species overlapped mainly during the Last Glacial Maximum, which could facilitate genetic admixture between those species. Finally, we found that most of the assessed environmental axes between the parents and hybrid populations are divergent, indicating that the climatic niche of the hybrid populations is shifting from that of the populations' progenitors. Our results show that the populations in the southern part of the Magdalena flatland are the result of combination of the genetic components of two species. Hybrid individuals with this novel genomic combination arose in a different habitat than their parental species, and they exhibit ecological divergence, which contributes to reproductive isolation through spatial and temporal barriers.

Keywords: endemism, climatic change, hybridization, mutualism, *Yucca valida*, *Yucca capensis*

OPEN ACCESS

Edited by:
Karolina Heyduk,
University of Hawaii, United States

Reviewed by:
Jeremy D. Rentsch,
Francis Marion University, United States
David Althoff,
Syracuse University, United States
Jeremy B. Yoder,
California State University, Northridge, United States

****Correspondence:***
Maria Clara Arteaga
arteaga@cicese.mx

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *28 January 2020*
Accepted: *30 April 2020*
Published: *28 May 2020*

Citation:
Arteaga MC, Bello-Bedoy R and Gasca-Pineda J (2020) Hybridization Between Yuccas From Baja California: Genomic and Environmental Patterns. Front. Plant Sci. 11:685. doi: 10.3389/fpls.2020.00685

INTRODUCTION

Hybridization, which is defined as the reproduction of members of genetically distinct populations that produce offspring of mixed ancestry (Barton and Hewitt, 1985), plays a pivotal role in evolution and diversification (Rieseberg, 1997; Taylor and Larson, 2019). There is ample evidence of hybridization in plants some of which has lead to the subsequent diversification of plant lineages (Arnold, 2015). In a spatial context, hybridization can occur when two geographically isolated species are reproductively compatible and have come into sympatry due to range shifts. Over time, climates have changed and species have responded by expanding or contracting their geographic distributions, causing some of them to become sympatric over parts of their ranges with the potential to interbreed during some periods of their evolutionary histories (e.g., Zinner et al., 2009; Cahill et al., 2018). Commonly, the biogeographic distributions of parents species are allopatric with respect to their hybrid derivatives (Gross and Rieseberg, 2004), which can happen when the environmental conditions of the area occupied by hybrids do not favor the establishment of parental species. Interspecific gene flow can generate new combinations of alleles, which create phenotypes that are able to explore novel habitats. The colonization of new habitats may allow hybrids to avoid competition with their parental species, and reproductive isolation can be achieved through ecological divergence and/or geographical isolation (e.g., Brochmann et al., 2000; Abbott et al., 2013).

Closely related species often hybridize (Abbott et al., 2013) if they share ecological attributes. For example, obligate mutualisms between flowering plants and their pollinators can favor genetic exchange and hybridization (Arnold, 2015), when pollinators do not differentiate between host species and plant reproductive barriers are incomplete (e.g., Leebens-Mack and Pellmyr, 2004; Kawakita and Kato, 2006). This has been the case in the obligate pollination mutualism between yuccas and yucca moths. The major reproductive barriers to gene flow among yucca species are spatial and temporal isolation and pollinator preference, and when these barriers collapse, hybridization can take place (Lenz and Hanson, 2000). There are several examples in which hybridization result from heterospecific pollination among sympatric yucca species (Miles, 1983; Leebens-Mack et al., 1998; Lenz and Hanson, 2000; Rentsch and Leebens-Mack, 2012; Starr et al., 2013). Within this genus, the hybridization has generated high morphological and genomic variation, including hybrids that coexist with their parents (Miles, 1983; Leebens-Mack et al., 1998; Lenz and Hanson, 2000; Starr et al., 2013; Royer et al., 2016) as well as new species (Rentsch and Leebens-Mack, 2012).

The two sister species *Yucca valida* Brandegee and *Yucca capensis* Lenz are long-lived monocot trees endemic to the Baja California Peninsula, located in northwestern México. Those species currently exhibit an allopatric distribution and grow under different environmental conditions. Populations of *Y. valida* grow in arid ecosystems from the Central Desert (30°N) to the Magdalena flatland, and this species show high density of individuals across the landscape (Turner et al., 1995). Plants of *Y. valida* flower between April and July (Turner et al., 1995; personal observation). In turn, populations of *Y. capensis* mainly grow in a remnant of tropical deciduous forest along the mountains located in the southern part of the Baja California Peninsula (Lenz, 1998; De la Luz et al., 2012), and consist of groups of less than 15 individuals separated by considerable distances (Lenz, 1998; Arteaga et al., 2015). Flowering of *Y. capensis*, in part, is determined by rainfall, with the boom during September and October (Lenz, 1998; Arteaga et al., 2015). These two sister Yucca species are pollinated by the same yucca moth, *Tegeticula baja* Pellmyr (Pellmyr et al., 2007, 2008). Briefly, after emerging from the cocoon, the female moth uses specialized mouthparts to collect pollen from yucca flowers and carry it to other flowers. She cuts into the floral ovary with her ovipositor and injects eggs. Then, she walks up to the stigma and actively deposits the pollen to fertilize the flower. The fruit develops, and her progeny feed on a small portion of the seeds (Pellmyr, 2003).

Until the mid-1990s, *Y. valida* and *Y. capensis* populations were considered a single species, *Y. valida*, and the large phenotypic variation was attributed to the ample variation in climatic conditions experienced by members of the species inhabiting the peninsula (Turner et al., 1995). However, based on their morphological characteristics, populations located in the tropical deciduous forest were described as a new species, named *Y. capensis* (Lenz, 1998). Moreover, the yucca plants that grow in the southern part of the Magdalena flatland (23.5°–24.5°N), show phenotypic traits of leaves and stems that resemble those of *Y. valida* and *Y. capensis*, leading to the suggestion that those populations originated through hybridization of both endemic yuccas (Lenz, 1998). Morphological inspection of those plants lends partial support to that hypothesis (personal observation), but the allopatric distribution of *Y. valida* and *Y capensis* makes it difficult to infer whether those plants are hybrids. However, it is possible that secondary contact between endemic yuccas occurred in the past, and the arrangement of new genetic combinations allowed putative hybrids to occupy environmental conditions that were unexplored by the parental species (Arnold, 2015).

In this study, we integrated genomic and climatic analyses to assess whether populations located in the southern part of the Magdalena flatland are *Y. valida* × *Y. capensis* hybrids. We sampled individuals across the geographical distribution of both endemic species as well as the putative hybrid populations, and we genotyped a total of 120 plants derived from 35 localities within the distribution ranges. We applied Bayesian tests and geographic cline analyses to the genomic data, and using climatic data from the occurrence sites, we did species distribution models (SDMs) and a statistical test to define the niche divergence among taxa. Our specific goals were to examine (i) whether populations located in the southern part of the Magdalena flatland originated through hybridization between *Y. capensis* and *Y. valida* and their level of admixture; (ii) whether Quaternary climate change influenced a shift in species distribution that favored hybridization and the establishment of hybrid descendants; and (iii) whether there is ecological divergence among putative hybrid populations and endemic yucca species.

MATERIALS AND METHODS

Sampling Area, DNA Extraction, and Genotyping

Across a transect of 800 km (from latitude 23° to latitude 26° N), we collected 10 g of fresh leaf tissue from 120 individuals from 35 localities (**Figure 1**). Using this sampling scheme, we covered the geographical range of *Y. valida* Brandegee and *Y. capensis* Lenz and the known range of the putative hybrid populations, which were located in the southern part of the Magdalena flatland (23.5°–24.5°N). Based in the geographical locations where we collected the individuals, we assigned them to a putative species or to the hybrid. We dried the tissues using silica gel for preservation. We extracted total genomic DNA from 100 mg of disrupted lyophilized leaf tissue using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). We assessed the quantity and integrity of DNA using 1.5% GelRed stained agarose gels and a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, United States).

We used the genomic DNA to generate nextRAD libraries, following previously described strategies (SNPsaurus, LLC; Russello et al., 2015). This method uses selective PCR primers to amplify genomic loci consistently among samples. Genomic DNA was first fragmented with Nextera reagent (Illumina, Inc., San Diego, CA, United States), which ligates short adapter sequences to the ends of the fragments. The Nextera reaction was scaled to fragment 25 ng of genomic DNA, and 50 ng of genomic DNA was used as an input to compensate for the amount of degraded DNA in the samples and to increase fragment sizes. Then, fragmented DNA was amplified for 27 cycles at 74°C, with one of the primers matching the adapter and nine nucleotides extending into the genomic DNA with the selective sequence GTGTAGAGCC. Thus, only fragments starting with a sequence that can be hybridized by the selective sequence of the primer would be efficiently amplified. The nextRAD libraries were sequenced on a HiSeq 4000 with one lane of 150 bp reads (University of Oregon). Raw reads are available in NCBI under the SRA numbers: SRR11514779, SRR11514778, and SRR11514777.

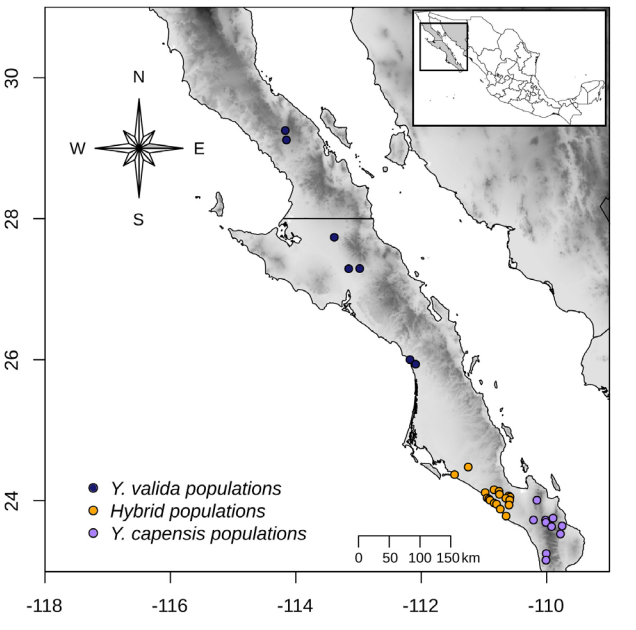


FIGURE 1 | Sampling locations of *Y. valida* (blue), *Y. capensis* (purple), and the putative hybrid populations (orange) in the Baja California Peninsula.

Raw Data Processing and Filtering

We checked the quality of raw sequence reads using FASTQC v0.11.5 (Andrews, 2010) with a threshold of $q > 29$. We removed the adapter sequences and low-quality reads using the *bbduk* module of the BBTools v37.87 software package[1]. All reads were end-trimmed to a length of 99 pb. Afterwards, we rechecked the reads' quality with FASTQC.

We built loci using a *de novo* assembly module of the Stacks v2.4 pipeline (Catchen et al., 2011, 2013). To obtain the optimal set of parameters, we used the procedure proposed by Paris et al. (2017). This procedure involves iterating different ranges of values for the parameters *m* (the minimum number of identical reads required to create a stack), *M* (the number of mismatches allowed between loci on a single individual), and *n* (the number of mismatches allowed between loci when building the catalog). We plotted and evaluated the resultant metrics to select the optimal set of parameters that maximizes the amount of reliable information. The parameter *r* (the minimum percentage of individuals required to process a locus) was fixed at 0.80, as recommended by Paris et al. (2017). The final set of parameters were $m = 3$, $M = 3$, and $n = 4$. When a locus showed more than one segregating site, one variant was randomly sampled from that locus to ensure that the loci were mostly independent. We discarded single-nucleotide polymorphisms (SNPs) with a minimum allele frequency (MAF) smaller than 0.05. Then, we used plink v1.07 (Purcell et al., 2007) to exclude individuals with more than 15% missing loci and loci that failed the Hardy-Weinberg equilibrium (HWE) test at $p < 0.05$. Moreover, to assess the degree of linkage disequilibrium among selected loci, we estimated correlation r^2 index and we considered those loci with $r^2 < 0.8$. Regarding the ploidy of these species, both endemic yuccas as well as the hybrid are diploids.

Genomic Diversity and Structure

We estimated observed and expected heterozygosities as well as inbreeding coefficients (H_E, H_O, F_{IS}) using the R v3.41 R Core Team 2017 packages "adegenet" v2.1 (Jombart and Ahmed, 2011), "hierfstat" v0.4-22 (Goudet and Jombart, 2015), and Genepop v4.6 (Rousset, 2008). We also estimated pairwise differentiation (F_{ST}) among the species and putative hybrid populations, and we evaluated the statistical significance using the gstat.randtest function of the "hierfstat" package in 10,000 simulations. As a first approach to explore the grouping of individuals as a function of their allele composition, we performed a principal component analysis (PCA) at the individual level using the "adegenet" package.

[1]https://sourceforge.net/projects/bbmap/files/

To evaluate the genetic structure of the taxa, we used two approaches. First, we ran Structure v2.3.4 (Pritchard et al., 2000; Falush et al., 2003) using the admixture model with correlated allele frequencies and a burnin period of 150,000 steps and 500,000 iterations. We tested one to eight clusters (K), performing 15 iterations each. The software was executed in parallel using the StrAuto v1.0 script (Chhatre and Emerson, 2017). To determine the most likely value of K, we performed the Evanno test (Evanno et al., 2005), which was implemented in Structure Harvester v0.6.94 (Earl, 2012). We also plotted the likelihood values for each K-value to identify the highest value of K with the lowest variance. Additionally, we used the method for discriminant analysis of principal components (DAPC) proposed by Jombart et al. (2010). This method relies on the use of multivariate individual information (in this case, genotypes) using discriminant functions, which optimize the variance between groups and minimize the variance within clusters. To select the number of retained components as well as the number of discriminant functions, we performed a cross-validation test. The number of possible clusters was assessed using the Bayesian Information Criterion (BIC), which was run 10 times to ensure the stability of the results. DAPC analysis was carried out using the "adegenet" package.

Evaluation of Historical Scenarios Using ABC Analysis

To approximate the evolutionary history of the endemic yuccas and putative hybrid populations, we used DIY-ABC v2.1 (Cornuet et al., 2014). We tested five scenarios related to the origin of hybrid plants located in the Magdalena flatland (**Supplementary Figure S1**). Scenarios A and B proposed that the putative hybrid populations originated from *Y. valida* and *Y. capensis*, respectively. Scenario C proposed that the hybrid populations diverged before the divergence of *Y. valida* and *Y capensis*. Scenario D proposed that *Y. valida* and *Y. capensis* diverged and the putative hybrid populations are a result of mixture of the two genetic pools. Scenario E proposed that the three taxa diverged simultaneously in the past, and they have independent demographic histories. We ran 2×10^6 simulations per scenario, as recommended by the software. We divided the entire data set into three sets and ran them independently, resulting in a total of 6×10^6 simulations per scenario.

We used the proportion of monomorphic loci, the mean genetic diversity, and the variance across polymorphic loci, as well as the mean genetic diversity across all loci as "single sample statistics." For the "two-sample statistics," we used the proportion of loci with null F_{ST} and Nei distances, the mean and variance across loci with non-null F_{ST} and Nei distances, and the mean and variance across loci of F_{ST} and Nei distances. In addition, we used the mean of admixture across all loci, the mean and variance of non-null admixture estimates, and the proportion of loci with null admixture as the "three sample statistics. To check if the combination of scenarios produced simulated data close to the actual data set (pre-evaluation), we implemented a PCA using 10,000 simulations. Then, we estimated the posterior probabilities of different scenarios by fitting a multinomial logistic regression using the 1% of simulated dataset closest to the observed data, followed by linear discriminant analysis. The goodness of fit of the scenario ("Model checking" option) was carried out by simulating 1,000 pseudo-observed data sets with the posterior model's distribution combination with parameter values drawn from 1,000 sets of the posterior sample. The summary statistics of the actual data were ranked with the posterior distribution of the scenarios' summary statistics. Finally, we estimated type-I and type-II errors for all scenarios implementing the "evaluate the confidence in scenario choice" option.

Geographic Cline Analyses

We performed sigmoid cline analyses of the allele frequencies for each SNP and the admixture index to identify genomic signals of hybridization across the geographical distribution of the yucca populations (hybrid zone, Abbott et al., 2013). We used the population average of the individual assignment probabilities (q-value) obtained with Structure for $K = 2$ as admixture index of the two putative parental species, *Y. valida*, and *Y. capensis.* Clines were fitted using the R package "hzar" v0.2-5 (Derryberry et al., 2014). First, we calculated the geographic distance between all sampling localities against the northernmost sampled locality using the function *pointDistance* in the R package "raster" v2.6-7 (Hijmans et al., 2017). We tested three models: (i) pmin/pmax set to the observed values without fitted exponential decay curves (i.e., tails; model I), (ii) estimated pmin/pmax with no fitted tails (model II), and (iii) estimated pmin/pmax and both tails fitted (model III). Models were evaluated based on the Akaike Information Criterion corrected for small sample size (AICc). The model with the lowest AICc was considered to have the best fit.

Environmental Data, Testing for Niche Divergence, and Species Distribution Modeling

To test for niche divergence, we employed the multivariate method introduced by McCormack et al. (2010), which compares niche divergence to a null hypothesis of divergence in available background environments on several orthogonal axes of environmental space. The method also uses PCA to reduce the raw GIS data into a smaller, uncorrelated set of axes. The general idea behind McCormack et al. (2010) method is that a pattern of divergence in GIS data can be attributed to either meaningful niche divergence between species or a strong spatial autocorrelation between GIS data. Therefore, a robust test of niche divergence or conservatism must compare niche divergence between taxa to the baseline levels of divergence drawn from the background of the available habitat within each taxon's geographic range. The null hypothesis is rejected when niches are more similar (niche conservatism) or more different (niche divergence) between taxa than the null model of background divergence. If the null hypothesis is not rejected, this does not mean that there is no niche divergence between the taxa, but that divergence between taxa (whether meaningful or due to spatial autocorrelation) is plausible.

To conduct the multivariate test for niche divergence, we used 19 BioClim layers (Bio1-Bio19)[2] that describe aspects of temperature, precipitation, and seasonality as well as potentially biologically limiting extremes of these variables. BioClim layers have a resolution of 1 km^2. To describe the climatic niche used by each taxon, we extracted raw data from 428 unique records (localities), 385 of which were from *Y. valida*, 18 of which were from *Y. capensis*, and 25 of which were from the putative hybrid populations. The latitude and longitude coordinates for individuals sampled and observed in the field were obtained from a global positioning system tracker. To generate the background predictions for each taxon, we developed a distribution polygon by drawing 5 km circles around each individual and merged them to obtain a continuous area. Then, we placed 2,000 random points inside each polygon and extracted raw data from them. We used the R packages "sp" (Pebesma and Bivand, 2005) and "dismo" (Hijmans et al., 2017). Because we were interested in determining whether putative hybrid populations ecologically diverge from their parents, we performed two pairwise analyses. One was conducted between *Y. valida* and putative hybrid populations, and the other was conducted between *Y. capensis* and putative hybrid populations. For each pairwise analysis, we joined the climatic data from records and background polygons of the two taxa, and we conducted a PCA. We extracted the first three principal component (niche) axes for further consideration since they comprised the bulk of the variation and were readily interpretable (see the Results section). Niche divergence or conservatism was evaluated on each niche axis by comparing the observed difference between the means for each yucca species and the putative hybrid populations on that axis to the mean difference in their background environments on the same axis. A null distribution of background divergence was created by recalculating the background divergence score over 1,000 jackknife replicates with 75% replacement. Significance for rejecting the null was evaluated at the 95% level. All analyses were conducted using Stata v10.

To evaluate the influence of past environmental conditions on possible distribution overlap of the taxa, we built species distribution models (SDMs) using BioClim layers[2] (Hijmans et al., 2005) for the present, mid-Holocene (MIROC-ESM; Watanabe et al., 2011), Last Glacial Maximum (MIROC-ESM; Watanabe et al., 2011), and Last Interglacial (Otto-Bliesner et al., 2006) periods. To avoid possible bias due to highly correlated variables, we extracted data from the 19 BioClim layers and conducted paired Pearson correlation tests with a threshold of >0.75. From each pair of correlated variables, we selected the variable that had more than one significant correlation with another variable. Additionally, we estimated the variance inflation factor using the vifcor function of the R package "usdm" (Naimi et al., 2014). Using both criteria, we retained twelve bioclimatic variables: *BIO1* (annual mean temperature), *BIO2* (mean diurnal range), *BIO4* (temperature seasonality), *BIO5* (maximum temperature of warmest month), *BIO8* (mean temperature of wettest quarter), *BIO9* (mean temperature of driest quarter), *BIO12* (annual precipitation), *BIO13* (precipitation of wettest month), *BIO14* (precipitation of driest month), *BIO15* (precipitation seasonality), *BIO18* (precipitation of warmest quarter), and *BIO19* (precipitation of coldest quarter).

We constructed the final models based on an ensemble of forecasting models using the committee-averaging criteria (Araújo and New, 2007) in the R package "biomod2" v3.1 (Thuiller et al., 2019). The ensemble uses four algorithms: (1) a generalized linear model (McCullagh and Nelder, 1989), (2) a generalized boosted model (Friedman, 1991), (3) MaxEnt (Phillips et al., 2006), and (4) random forests (Breiman, 2001). To reduce possible autocorrelation bias due to local overrepresentation of records, we generated a grid of 1/10 degree cells for each taxon and then selected one point at random from each cell using the "raster" package. Two independent pseudo-absence sets of 5,000 points were generated at random, and the species records were split (70% for model training and 30% for evaluation of the model's performance). With the 70–30 criterion, we ran five random replicates for all models. We assessed the models' performance using the area under the receiver operating characteristic curve (AUC; Swets, 1988) and the true skill statistic (TSS; Allouche et al., 2006). Ensembles were restricted to models with AUC $>$ 0.9 and TSS $>$ 0.8, and we transformed them into binary data using the evaluation metrics and thresholds obtained by TSS evaluation.

RESULTS

Genomic Diversity and Structure

After quality checking and filtering, the final data set consisted of 3,423 biallelic loci from 103 individuals from 35 localities. The highest value of linkage disequilibrium was $r^2 = 0.72$, and more than 97.5% of paired values were lower than $r^2 = 0.5$, so we keep all loci for further analyses. The two endemic yuccas and hybrid populations had an overall genetic diversity of $H_E = 0.2826$ and $H_O = 0.1255$ for the 3,423 analyzed SNPs. The three taxa show differences in genetic diversity, and the putative hybrid populations had the highest diversity. We also found a significant deficiency of heterozygotes in the three taxa (**Table 1**). In addition, fixed loci were common in *Y. valida* Brandegee and *Y. capensis* Lenz, and polymorphic loci were mostly found in hybrid populations (**Supplementary Figure S2**). Finally, the hybrid populations had only 28 private alleles, while we detected 40 private alleles in *Y. valida*, and 78 in *Y. capensis*.

TABLE 1 | Number of localities and individuals sampled per taxon, and summary statistics of genetic diversity (H_E, expected heterozygosities; H_O, observed heterozygosities; and F_{IS}, inbreeding coefficient).

Taxa	Localities (Ind)	H_E	H_O	F_{IS}
Y. valida	7 (35)	0.1736	0.1119	0.3522
Hybrid pops.	18 (43)	0.2606	0.161	0.3802
Y. capensis	10 (25)	0.2496	0.1036	0.5763
Overall	35 (103)	0.2826	0.1255	0.5338

[2]http://worldclim.org/bioclim

The overall value of genomic differentiation among the three taxa was F_{ST} = 0.1939. While the level of structure between the endemic species (*Y. valida* and *Y. capensis*) was high (F_{ST} = 0.2538), the values of the paired comparisons with hybrid populations were lower (*Y. valida*–hybrid populations: F_{ST} = 0.099; *Y. capensis*–hybrid populations: F_{ST} = 0.102). Moreover, the PCA performed at the individual level showed clear separation of individuals of different yucca populations (**Figure 2**). The clustering analysis, which was performed with Structure using the Evanno test (Evanno et al., 2005), suggested the existence of two groups (K = 2; **Figure 2**). Close inspection of the ln (P) plots showed that individuals with higher assignment probabilities correspond to the localities of *Y. valida* ($q \geq 0.930$) or *Y. capensis* ($q \geq 0.888$) and all individuals with high levels of admixture belonged to the locality of the hybrid populations. Following the PCA analysis, DAPC revealed three genetic groups separating all individuals into their respective taxon (**Supplementary Figure S3**). This result was expected, as multivariate methods optimize the variance between groups without *a priori* assumptions, while Structure searches for the optimal grouping searching a Hardy-Weinberg equilibrium. Based on this, we included the barplot values for K = 3 (**Figure 2**), which allowed us to determine that the third genetic group predominates in individuals from the hybrid populations.

Evaluation of Historical Scenarios and Geographic Cline Analyses

ABC modeling of historical scenarios provided unambiguous support for scenario D (**Supplementary Figure S1**), which suggests that hybrid populations are the result of a mixture of the previously diverged species *Y. valida* and *Y. capensis*. The cline of the admixture accurately described the geographic transition between *Y. valida* and *Y. capensis* (**Figure 3**). We

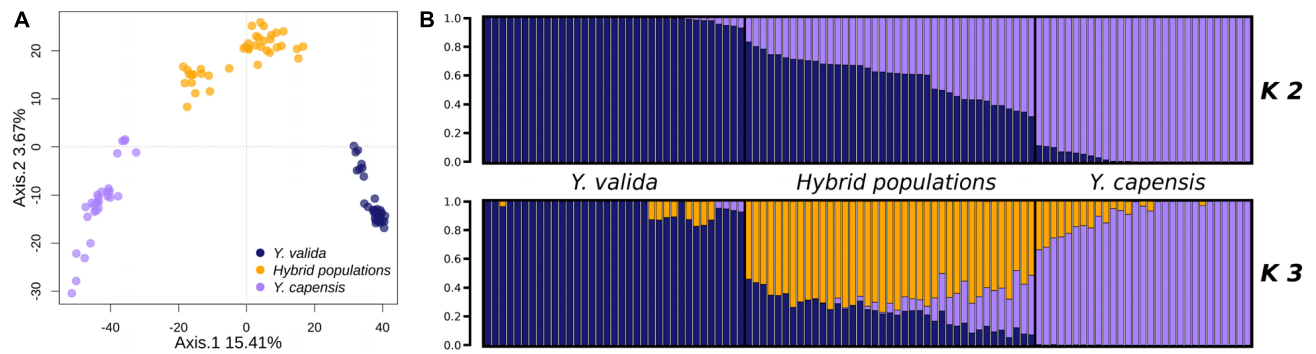


FIGURE 2 | (A) Principal component analysis of the 3,423 SNPs obtained for *Y. valida*, *Y. capensis*, and the hybrid populations. Dots represent individuals. **(B)** Bayesian cluster analysis of *Y. valida*, *Y. capensis*, and the putative hybrid individuals using Structure (K = 2 and K = 3). Genetic clusters are coded with different colors. Individuals were ordered according to the latitude of their locality of origin.

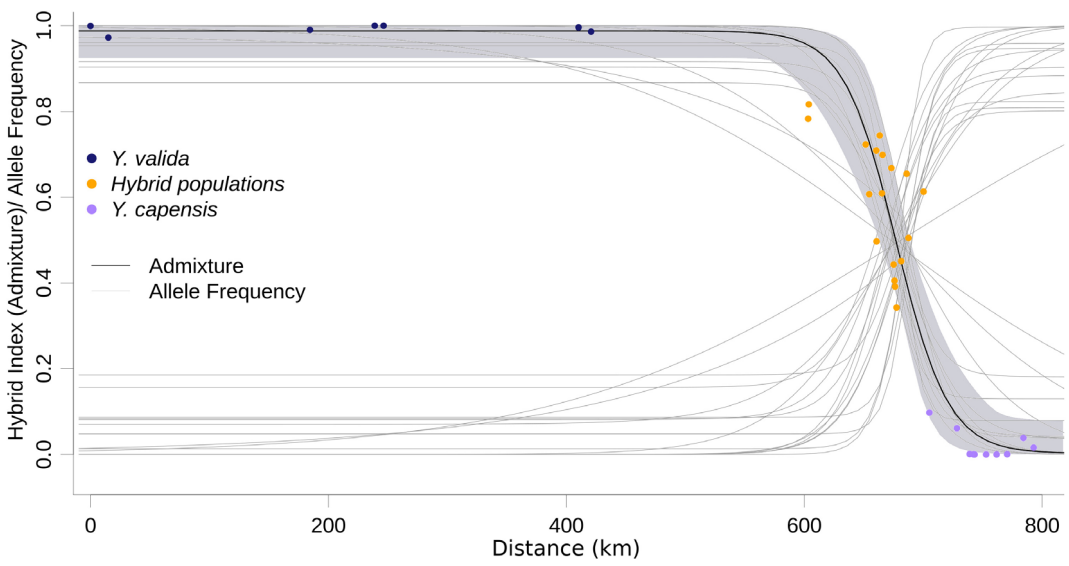


FIGURE 3 | Maximum likelihood geographic cline for the hybrid index. Dots represent the 35 localities, and the shaded area represents the 95% interval of the fitted model. Light gray lines represent the clines for the 174 diagnostic loci.

found 176 diagnostic loci (i.e., loci that fit the sigmoidal cline model), 30 of which exhibited coincidence with the center of the admixture cline.

Testing for Niche Divergence and Species Distribution Modeling

For comparison between *Y. valida* and the hybrid populations, the PCA identified three main niche axes that explained 82% of the variation. The first niche axis was associated with mean temperatures and rain seasonality; the second was associated with the annual range (variation) and seasonality of temperature and the lowest temperatures; and the third was associated with the highest temperature and precipitation and annual rain. The results of the niche divergence test showed evidence of conservatism on the first niche axis and divergence on the second and third axes (**Table 2**). Both taxa have conserved niches related with mean temperatures and rain seasonality (e.g., deserts). However, they diverge in terms of extreme temperature and rain values as well as the range of variation in temperature.

Between *Y. capensis* and the hybrid populations, PCA indicated that three main niche axes explained 92.3% of the variation. The first niche axis was associated with annual precipitation and extreme levels of rain; the second was associated with temperature and rain seasonality; and the third was associated with the minimum temperature of the coldest month. The results of the test indicate divergence on the second and third axes. The first niche axis did not significantly differ from the null expectation of background divergence (**Table 2**). Both taxa have divergent niches related to rain and temperature seasonality and the lowest temperatures.

Finally, the SDMs indicate that the current distribution of both endemic yucca species does not overlap (**Figure 4**). Nevertheless, there are interesting results for past scenarios. The distribution of the suitable habitat for *Y. valida* reached lower latitudes during the mid-Holocene and LGM periods, but apparently, there was not suitable habitat in the center and south of the peninsula for the species during the Last Interglacial period. Although *Y. capensis* had the most restricted distribution in the present period, its suitable habitat was more widely distributed in the past, reaching the middle of the peninsula in the LIG period (around 28°N). Thus, the potential distribution of both endemic species overlapped in the past, mainly around latitudes of 25.5–23.5°N (**Figure 4**), which exhibit environmental suitability for both yucca species and correspond to the area currently occupied by the hybrid populations. A higher level of habitat suitability for the hybrid populations in the present period was found at latitudes of 25–22.8°N in a larger area than where they are currently observed (**Figure 4**).

TABLE 2 | Tests of niche divergence and conservatism.

	PC1	PC2	PC3
***Y. valida*-Hybrid pops.**			
Observed difference	4.3165*	**1.367***	**1.0301***
Null distribution	4.4744–4.6540	0.7710–1.0356	0.5880–0.7087
% variance explained	43%	26%	13%
***Y. capensis*-Hybrid pops.**			
Observed difference	4.4887	**2.9678***	**0.4564***
Null distribution	4.3197–4.5314	2.3443–2.4765	0.0520–0.1471
% variance explained	60%	20%	11%

*Observed differences in the climatic niche of the hybrid populations and parental species on each PC axis are compared to the middle 95th percentile of the null distribution of the differences between their environmental backgrounds. Bold values indicate niche divergence. *Significance level, $p < 0.05$.*

DISCUSSION

Species geographic ranges are dynamic over evolutionary time. The large degree of allopatry among sibling species of yuccas and moth taxa suggest that diversification inside each group has been driven by geographical isolation rather than by reproductive isolation (Althoff et al., 2012). Thus, when two closely related species have come into sympatry due to range shifts, they could interbreed if their pollinators are not highly specialized. Our genomic data and Species Distribution Models support this scenario to endemic yuccas of the Baja California Peninsula. We confirmed the hybrid origin of yucca populations located in the southern part of the Magdalena flatland in the Baja California Peninsula. Climatic changes in the past probably caused geographic overlap of the distribution areas of *Y. capensis* Lenz and *Y. valida* Brandegee, which could facilitate genetic admixture between those species. Environmental analyses show that the hybrid populations are distributed in areas that are slightly different from those of the parental species, which can promote reproductive isolation.

Genomic Patterns and Probable Historical Scenarios of Hybridization

One of the goals of this study was to assess whether populations located in the southern part of the Magdalena flatland have hybrid origins due to genetic exchange between *Y. valida* and *Y. capensis*. We found a clear grouping of two clusters corresponding to the two parental species and a third cluster that corresponded to individuals from the hybrid populations (**Figure 2**). In addition, those populations show the highest levels of genomic admixture compared to the level of admixture observed in individuals from the parental species. Although mixed ancestry in the genome of taxa is an important indicator of hybridization, it can be difficult to distinguish from ancestral polymorphism or continued gene exchange (Abbott et al., 2013). For this reason, we used alternative scenarios (ABC models) to test the most probable history that explains the data. We found the highest support for the model of a hybrid origin of these populations. Our results indicate that populations located in the southern part of the Magdalena flatland originated of hybridization between *Y. capensis* and *Y. valida*.

The occurrence of hybridization has been recorded at least four times among *Yucca* species, and it occurs in zones of host sympatry where yucca moths play an active role in the movement of heterospecific pollen between parental species

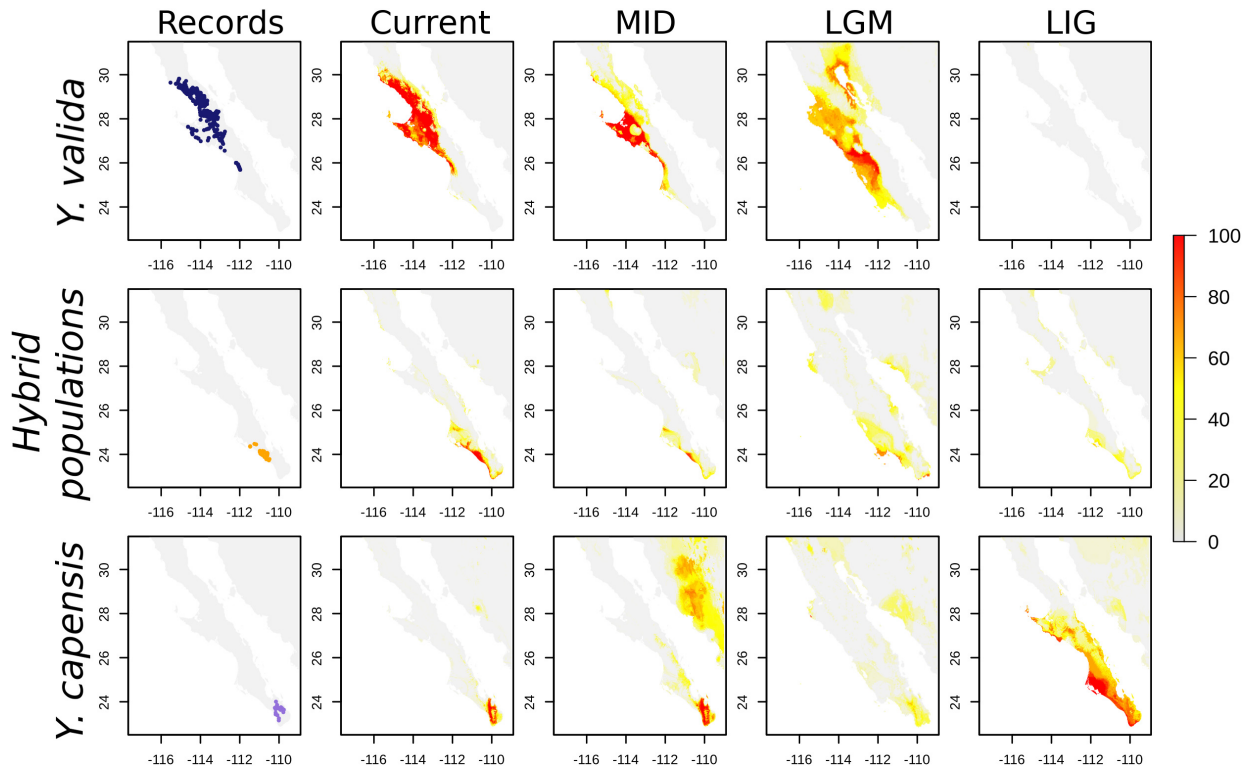


FIGURE 4 | Georeferenced records and Species distribution models (SDMs) for *Y. valida*, *Y. capensis*, and the hybrid populations. The SDMs were generated for current conditions, the mid-Holocene (MID ∼6,000 years), the Last Glacial Maximum (LGM ∼21,000 years), and the Last Interglacial period (LIG ∼120,000 years). The color represents the habitat suitability values.

(Miles, 1983; Lenz and Hanson, 2000; Rentsch and Leebens-Mack, 2012; Starr et al., 2013; Royer et al., 2016). Because yucca moths move pollen across short distances (Marr et al., 2000; Powell, 2013), heterospecific pollination could occur in places where parental species coexist, given that flowering phenologies overlap. The endemic species, *Y. valida* and *Y. capensis*, are pollinated by the same yucca moth species (Pellmyr et al., 2007), which could favor pollen flow between those hosts. However, in the current landscape, they have an allopatric distribution, suggesting that the hybridization events occurred in the past. Our SDMs reveal that the potential geographic distribution of the parental species changed since the Last Interglacial period and that their ranges show overlap mainly during the Last Glacial Maximum (21,000 ka) and mid-Holocene (6,000 ka) periods. This overlap was located between the latitudes of 25.5–23.5°N, where the hybrid populations are distributed (23.5 and 24.5°N). Similar patterns of secondary contact caused by quaternary oscillations have been recorded in other species and resulted in hybridization (Liu et al., 2014; Marques et al., 2016; Cahill et al., 2018). Our results highlight the role of climatic changes in biodiversity and indicate that Quaternary climate change could favor the scenario of hybridization between the endemic yuccas and establishment of hybrid descendants.

Hybridization over multiple generations causes gradual changes, or clines, in allele frequencies over geographical locations, and a balance between gene flow and natural selection maintains these changes (Barton and Hewitt, 1985). Those loci that display large differences in allele frequencies between parental species can be under strong selection, or linked to selected loci (Yeaman and Otto, 2011). Moreover, these differences can be caused by genetic drift. Among our 3,423 loci, only 174 (5%) fitted the cline model, and 30 of those had a concordant cline center as the hybrid index cline. The displacement of the cline center of the other 144 SNPs suggests asymmetrical gene flow from one of the parental species due to geographical proximity, synchronic flowering, or higher environmental affinities. This displacement also can result from selection against parental variants or allele combinations that do not fit well with the local conditions of the hybrid zone. The other 95% of loci show high variance in allele frequencies across the geographic sampling area, and they could result of the retention of ancestral polymorphism. Our cline analysis confirms the geographic area where hybrids are occurring. The allele frequencies behavior detected provide support, as well done by our ABC models, that these populations are result of admixture from parental species, and they are not a different genetic pool. In the future, when yucca genomes will be available, we will map these candidate loci and associate with morphological features under natural selection. Our results suggest semi-permeable boundaries between genomes (Wu, 2001), which allow some loci to freely introgress and not

others. This scenario is common in the early and intermediate stages of divergence among taxa (Baldassarre et al., 2014; De La Torre et al., 2015).

The level of differentiation between parental and hybrid species is influenced by the gene flow rates, and the semi-permeable boundaries between genomes results in differential introgression (Wu, 2001). On a wider scale, the F_{ST} estimations indicate the average level of gene flow between populations. In the current landscape of the peninsula, the geographic distribution of hybrid populations is closer to that of *Y. capensis* than that of *Y. valida*, and it is plausible that higher levels of gene flow exist between hybrids and *Y. capensis* than between hybrids and *Y. valida*. However, the genetic differentiation between the hybrids and both parental species is similar, suggesting similar gene flow rates in both directions. Spatial barriers can limit the gene flow between *Y. valida* and hybrids because there are several hundred kilometers between the geographical ranges of both taxa. In the other hand, temporal barriers can maintain low rates of gene flow between *Y. capensis* and hybrids because the hybrid plants flower in July and August (personal observation) and *Y. capensis* flowers mainly in September and October (Lenz, 1998).

Environmental Patterns

Reproductive isolation of the hybrid and parental species is an important factor in the speciation process (Schumer et al., 2014). Geographic isolation and ecological divergence can serve as prezygotic barriers. In our study area, yucca hybrid populations occur in areas where parental species are not currently present, which leads to a major geographical isolation barrier to gene flow between parental and hybrid individuals. In addition, we found that most of the environmental axes between the parents and hybrid populations are divergent, achieving our third research goal by indicating that the climatic niche of hybrid populations is shifting from that of their progenitors. Hybrid populations can colonize new ecological spaces that are not utilized by the parental taxa. This ecological divergence could result from a new combinations of traits generated by hybridization, or it can be achieved after speciation through the gradual accumulation of new mutations (Gross and Rieseberg, 2004). Further experiments need to be carried to test which of the hypotheses explains the detected pattern (e.g., Brochmann et al., 2000).

Regions with high environmental heterogeneity favor speciation. The endemic species *Y. valida* occurs mainly in the Vizcaino Desert, and the southern populations of this species are located in the northern part of the Magdalena flatland, while the hybrid populations are distributed in the southern part of the flatland. The climatic conditions of both regions (the Magdalena flatland and Vizcaino Desert) are influenced by the California Current and the associated cold surges (González-Abraham et al., 2010). The habitat of *Y. valida* is extreme in terms of some environmental conditions, exhibiting, for example, the lowest temperatures and low levels of rain, which can explain the partial divergence in climatic niches observed between the taxa. On the other hand, *Y. capensis* is endemic to the Cape Region of Baja California Sur, and it is distributed in the undergrowth of the lowland deciduous forest in the mountains, up to 1,000 meters above sea level (Lenz, 1998; De la Luz et al., 2012). The habitat of *Y. capensis* is much more seasonal than the habitat of hybrid populations, supporting the niche divergence between the taxa.

Geographically explicit predictions of climatic niches are a good starting point to explore niche differences and discover regions with suitable climatic conditions for species. Our SDMs for the present accurately described the distribution of the two endemic species. For the hybrid populations, the model revealed a wide area with suitable conditions that extends beyond the regions where we recorded hybrid individuals. In addition, projecting the models to different times allowed us to link current genomic patterns with historical distributions. For example, *Y. capensis* shows the highest genomic variability, although its geographical range is currently the most restricted. Nevertheless, the predicted distribution of this species during the Last Interglacial and Last Glacial Maximum periods was wider than in the present, suggesting that, in the past, there was a larger suitable area that could support a higher effective population size. On the other hand, *Y. valida*, the parental species with wider distribution, had a low level of genetic diversity and a high number of fixed alleles. This is explained in part by a reduced suitable habitat during the Last Interglacial period that could result in a small effective historical population for this species.

This study confirms how the geographical context influences the specificity of the pollination mutualism between yuccas and yucca-moths. We showed that the hybrid populations in the Baja California Peninsula are the result of combination of the genetic components of the two endemic yucca species. Currently, hybrid individuals with this novel genomic combination occur in different habitats than their parental species, and ecological divergence between them, as well as the spatial and temporal barriers, contributes to reproductive isolation. Finally, to consider these populations as homoploid hybrid species, it is necessary to collect new evidence, which may be related to the role of natural selection in maintaining the distinctions of hybrid taxon as well as the mechanisms through which hybridization generates reproductive isolation (Schumer et al., 2014).

DATA AVAILABILITY STATEMENT

Raw reads are available in NCBI under the SRA numbers: SRR11514779, SRR11514778, and SRR11514777.

AUTHOR CONTRIBUTIONS

All work described in this manuscript was original research carried out by the authors. MA and RB-B developed the original research design and field data collection. JG-P and MA carried out the analyses. All authors contributed substantially to the writing and proofreading of the article, and read and approved the final submitted version.

FUNDING

Funding was provided by CONACYT (Grant Nos. CB-2014-01 238843 and infra-20141 226339). The Rufford Foundation also provided financial support for part of this study.

ACKNOWLEDGMENTS

We are grateful with three reviewers for their comments to improve the manuscript. We would like to thank José Luis León de la Luz, Raymundo Domíguez, Alfonso Medel Narváez, and José Delgadillo for their invaluable assistance in the fieldwork and to share their botanic knowledge. We thank to Cynthia Rocío ÁĄlamo, Carlo G. González, and Astrid Luna for their help with data processing and fieldwork. We also thank to Lita Castañeda for the technical assistance in the laboratory. MA thanks CONACYT for funding project CB-2014-01-238843 and the Rufford Foundation for providing financial support to develop the study. RB-B thanks CONACYT for funding project infra-20141-226339 and the Facultad de Ciencias of UABC for providing the facilities to conduct this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00685/full#supplementary-material

FIGURE S1 | Models tested for the origin of putative hybrid populations using Approximate Bayesian Computation (ABC) toolbox. Scenario **(A)** ancestral divergence of *Y. valida*, and *Y. capensis* and the posterior origin of the hybrid populations from *Y. valida*; Scenario **(B)** ancestral divergence of *Y. valida* and *Y. capensis*, and the posterior origin of the putative hybrid populations from *Y. capensis*; Scenario **(C)** the divergence of the hybrid populations predates the divergence of *Y. valida* and *Y. capensis*; Scenario **(D)** ancestral divergence of *Y. valida* and *Y. capensis* and the posterior origin of the putative hybrid populations as a result of the admixture of the two genetic pools; Scenario **(E)** the three taxa diverged simultaneously in the past, and they have independent demographic histories.

FIGURE S2 | Barplots of the proportion of polymorphic loci for *Y. valida*, *Y. capensis*, and the hybrid populations.

FIGURE S3 | Plot of the results of discriminant principal component analysis (DPCA; $K = 3$) for the localities of *Y. valida* (blue), *Y. capensis* (purple), and the hybrid populations (orange). Dots represent individuals.

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J., Bierne, N., et al. (2013). Hybridization and speciation. *J. Evol. Biol.* 26, 229–246. doi: 10.1111/j.1420-9101.2012.02599.x

Allouche, O., Tsoar, A., and Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* 43, 1223–1232. doi: 10.1111/j.1365-2664.2006.01214.x

Althoff, D. M., Segraves, K. A., Smith, C. I., Leebens-Mack, J., and Pellmyr, O. (2012). Geographic isolation trumps coevolution as a driver of Yucca and Yucca moth diversification. *Mol. Phylogenet. Evol.* 62, 898–906. doi: 10.1016/j.ympev.2011.11.024

Andrews, S. (2010). *FastQC: a Quality Control Tool for High Throughput Sequence Data*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed October 2018).

Araújo, M. B., and New, M. (2007). Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22, 42–47. doi: 10.1016/j.tree.2006.09.010

Arnold, M. L. (2015). *Divergence with Genetic Exchange*. Oxford: OUP.

Arteaga, M. C., Bello-Bedoy, R., León-de la Luz, J. L., Delgadillo, J., and Dominguez, R. (2015). Phenotypic variation of flowering and vegetative morphological traits along the distribution for the endemic species Yucca capensis (Agavaceae). *Bot. Sci.* 93, 765–770. doi: 10.17129/botsci.214

Baldassarre, D. T., White, T. A., Karubian, J., and Webster, M. S. (2014). Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. *Evolution* 68, 2644–2657. doi: 10.1111/evo.12457

Barton, N. H., and Hewitt, G. M. (1985). Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* 16, 113–148. doi: 10.1146/annurev.es.16.110185.000553

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32.

Brochmann, C., Borgen, L., and Stabbetorp, O. E. (2000). Multiple diploid hybrid speciation of the canary island endemic *Argyranthemum sundingii* (Asteraceae). *Plant Syst. Evol.* 220, 77–92. doi: 10.1007/bf00985372

Cahill, J. A., Heintzman, P. D., Harris, K., Teasdale, M. D., Kapp, J., Soares, A. E., et al. (2018). Genomic evidence of widespread admixture from polar bears into brown bears during the last ice age. *Mol. Biol. Evol.* 35, 1120–1129. doi: 10.1093/molbev/msy018

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1111/mec.12354

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3* 1, 171–182. doi: 10.1534/g3.111.000240

Chhatre, V. E., and Emerson, K. J. (2017). StrAuto: Automation and parallelization of STRUCTURE analysis. *BMC Bioinformatics*, 18, 192. doi: 10.1186/s12859-017-1593-0

Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., et al. (2014). DIYABC v2. 0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* 30, 1187–1189. doi: 10.1093/bioinformatics/btt763

De la Luz, J. L., Domínguez-Cadena, R., and Medel-Narváez, A. (2012). Florística de la selva baja caducifolia de la península de Baja California, México. *Bot. Sci.* 90, 143–162.

De La Torre, A., Ingvarsson, P. K., and Aitken, S. N. (2015). Genetic architecture and genomic patterns of gene flow between hybridizing species of Picea. *Heredity* 115, 153–164. doi: 10.1038/hdy.2015.19

Derryberry, E. P., Derryberry, G. E., Maley, J. M., and Brumfield, R. T. (2014). HZAR: hybrid zone analysis using an R software package. *Mol. Ecol. Resour.* 14, 652–663. doi: 10.1111/1755-0998.12209

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294x.2005.02553.x

Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *Ann. Stat.* 19, 1–67.

González-Abraham, C. E., Garcillán, P. P., and Ezcurra, E. (2010). Ecorregiones de la península de Baja California: una síntesis. *Bol. Soc. Bot. Mex.* 87, 69–82.

Goudet, J., and Jombart, T. (2015). *hierfstat: Estimation and Tests of Hierarchical F-Statistics. R package version 0.04-22*. Available online at: https://CRAN.R-project.org/package=hierfstat (accessed December 4, 2015).

Gross, B. L., and Rieseberg, L. H. (2004). The ecological genetics of homoploid hybrid speciation. *J. Hered.* 96, 241–252. doi: 10.1093/jhered/esi026

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. doi: 10.1002/joc.1276

Hijmans, R. J., Phillips, S., Leathwick, J., and Elith, J. (2017). *dismo: Species Distribution Modeling. R package version 1.1-4*. Available online at: https://CRAN.R-project.org/package=dismo (accessed January 9, 2017).

Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070–3071. doi: 10.1093/bioinformatics/btr521

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 1:94. doi: 10.1186/1471-2156-11-94

Kawakita, A., and Kato, M. (2006). Assessment of the diversity and species specificity of the mutualistic association between Epicephala moths and Glochidion trees. *Mol. Ecol.* 15, 3567–3581. doi: 10.1111/j.1365-294x.2006.03037.x

Leebens-Mack, J., Pellmyr, O., and Brock, M. (1998). Host specificity and the genetic structure of two Yucca moth species in a Yucca hybrid zone. *Evolution* 52, 1376–1382. doi: 10.1111/j.1558-5646.1998.tb02019.x

Leebens-Mack, J., and Pellmyr, O. L. L. E. (2004). Patterns of genetic structure among populations of an oligophagous pollinating Yucca moth (*Tegeticula yuccasella*). *J. Hered.* 95, 127–135. doi: 10.1093/jhered/esh025

Lenz, L. E. (1998). Yucca capensis (Agavaceae, Yuccoideae), a new species from Baja California Sur, Mexico. *Cact. Succ. J.* 70, 289–296.

Lenz, L. W., and Hanson, M. A. (2000). Yuccas (Agavaceae) of the international four corners: Southwestern USA and northwestern Mexico. *Aliso* 19, 165–179. doi: 10.5642/aliso.20001902.04

Liu, B., Abbott, R. J., Lu, Z., Tian, B., and Liu, J. (2014). Diploid hybrid origin of *Ostryopsis intermedia* (Betulaceae) in the Qinghai-Tibet plateau triggered by quaternary climate change. *Mol. Ecol.* 23, 3013–3027. doi: 10.1111/mec.12783

Marques, I., Draper, D., López-Herranz, M. L., Garnatje, T., Segarra-Moragues, J. G., and Catalán, P. (2016). Past climate changes facilitated homoploid speciation in three mountain spiny fescues (Festuca, Poaceae). *Sci. Rep.* 6:36283. doi: 10.1038/srep36283

Marr, D. L., Leebens-Mack, J., Elms, L., and Pellmyr, O. (2000). Pollen dispersal in *Yucca filamentosa* (Agavaceae): the paradox of self-pollination behavior by *Tegeticula yuccasella* (Prodoxidae). *Am. J. Bot.* 87, 670–677. doi: 10.2307/2656853

McCormack, J. E., Zellmer, A. J., and Knowles, L. L. (2010). Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecological speciation?: insights from tests with niche models. *Evolution* 64, 1231–1244. doi: 10.1111/j.1558-5646.2009.00900.x

McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall.

Miles, N. J. (1983). Variation in host specificity in the Yucca moth, *Tegeticula yuccasella* (Incurvariidae): a morphometric approach. *J. Lepid. Soc.* 37, 207–216.

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., and Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography* 37, 191–203. doi: 10.1111/j.1600-0587.2013.00205.x

Otto-Bliesner, B. L., Marshall, S. J., Overpeck, J. T., Miller, G. H., and Hu, A. (2006). Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science* 311, 1751–1753. doi: 10.1126/science.1120808

Paris, J. R., Stevens, J. R., and Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. *Methods Ecol. Evol.* 8, 1360–1373. doi: 10.1111/2041-210x.12775

Pebesma, E. J., and Bivand, R. S. (2005). *Classes and Methods for Spatial Data in R. R News*. Available online at: https://cran.r-project.org/doc/Rnews/ (accessed February 1, 2019).

Pellmyr, O. (2003). Yuccas, Yucca moths, and coevolution: a review. *Ann. Mo. Bot. Gard.* 90, 35–55.

Pellmyr, O., BalcAZar-Lara, M., Segraves, K. A., Althoff, D. M., and Littlefield, R. J. (2008). Phylogeny of the pollinating Yucca moths, with revision of Mexican species (Tegeticula and Parategeticula; Lepidoptera, Prodoxidae). *Zool. J. Linn. Soc. Lond.* 152, 297–314. doi: 10.1111/j.1096-3642.2007.00361.x

Pellmyr, O., Segraves, K. A., Althoff, D. M., Balcázar-Lara, M., and Leebens-Mack, J. (2007). The phylogeny of yuccas. *Mol. Phylogenet. Evol.* 43, 493–501. doi: 10.1016/j.ympev.2006.12.015

Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259. doi: 10.1016/j.ecolmodel.2005.03.026

Powell, J. A. (2013). Longevity and individual activity of the Yucca moth, *Tegeticula maculata* extranea (Prodoxidae), based on mark-release monitoring. *J. Lepid. Soc.* 67, 187–196.

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

Purcell, S., Benjamin, N., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Rentsch, J. D., and Leebens-Mack, J. (2012). Homoploid hybrid origin of *Yucca gloriosa*: intersectional hybrid speciation in *Yucca* (Agavoideae, Asparagaceae). *Ecol. Evol.* 2, 2213–2222. doi: 10.1002/ece3.328

Rieseberg, L. H. (1997). Hybrid origins of plant species. *Annu. Rev. Ecol. Syst.* 28, 359–389. doi: 10.1146/annurev.ecolsys.28.1.359

Rousset, F. (2008). Genepop'007: a complete re-implementation of the genepop software for windows and Linux. *Mol. Ecol. Resour.* 8, 103–106. doi: 10.1111/j.1471-8286.2007.01931.x

Royer, A. M., Streisfeld, M. A., and Smith, C. I. (2016). Population genomics of divergence within an obligate pollination mutualism: selection maintains differences between Joshua tree species. *Am. J. Bot.* 103, 1730–1741. doi: 10.3732/ajb.1600069

Russello, M. A., Waterhouse, M. D., Etter, P. D., and Johnson, E. A. (2015). From promise to practice: pairing non-invasive sampling with genomics in conservation. *PeerJ* 3:e1106. doi: 10.7717/peerj.1106

Schumer, M., Rosenthal, G. G., and Andolfatto, P. (2014). How common is homoploid hybrid speciation? *Evolution* 68, 1553–1560. doi: 10.1111/evo.12399

Starr, T. N., Gadek, K. E., Yoder, J. B., Flatz, R., and Smith, C. I. (2013). Asymmetric hybridization and gene flow between Joshua trees (Agavaceae: *Yucca*) reflect differences in pollinator host specificity. *Mol. Ecol.* 22, 437–449. doi: 10.1111/mec.12124

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science* 24, 1285–1293. doi: 10.1126/science.3287615

Taylor, S. A., and Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat. Ecol. Evol.* 3, 170–177. doi: 10.1038/s41559-018-0777-y

Thuiller, W., Georges, D., Engler, R., and Breiner, F. (2019). *biomod2: Ensemble Platform for Species Distribution Modeling. R package version 3.3-7.1*. Available online at: https://CRAN.R-project.org/package=biomod2 (accsessed March 3, 2019).

Turner, R. M., Bowers, J. E., and Burgess, T. L. (1995). *Sonoran Desert Plants: an Ecological Atlas*. Tucson: The University of Arizona Press.

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., et al. (2011). MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments. *Geosci. Model Dev.* 4, 845–872. doi: 10.5194/gmd-4-845-2011

Wu, C. I. (2001). The genic view of the process of speciation. *J. Evol. Biol.* 14, 851–865. doi: 10.1046/j.1420-9101.2001.00335.x

Yeaman, S., and Otto, S. P. (2011). Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution* 65, 2123–2129. doi: 10.1111/j.1558-5646.2011.01277.x

Zinner, D., Groeneveld, L. F., Keller, C., and Roos, C. (2009). Mitochondrial phylogeography of baboons (*Papio* spp.)–indication for introgressive hybridization? *BMC Evol. Biol.* 9:83. doi: 10.1186/1471-2148-9-83

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 11 June 2020
doi: 10.3389/fpls.2020.00654

Tissue Composition of *Agave americana* L. Yields Greater Carbohydrates From Enzymatic Hydrolysis Than Advanced Bioenergy Crops

*Alexander M. Jones[1], Yadi Zhou[2], Michael A. Held[2] and Sarah C. Davis[1,3]**

[1] Voinovich School of Leadership and Public Affairs, Ohio University, Athens, OH, United States, [2] Department of Chemistry and Biochemistry, Ohio University, Athens, OH, United States, [3] Department of Environmental and Plant Biology, Ohio University, Athens, OH, United States

Agave americana L. is a highly productive, drought-tolerant species being investigated as a feedstock for biofuel production. Some *Agave* spp. yield crop biomass in semi-arid conditions that are comparable to C_3 and C_4 crops grown in areas with high rainfall. This study evaluates the bioethanol yield potential of *A. americana* by (1) examining the relationship between water use efficiency (WUE) and plant carbohydrates, (2) quantifying the carbohydrate and energy content of the plant tissue, and (3) comparing the products of enzymatic hydrolysis to that of other candidate feedstocks (*Miscanthus x giganteus* Greef et Deuter, *Sorghum bicolor* (L.) Moench, and *Panicum virgatum* L.). Results indicate that (1) WUE does not significantly affect soluble and insoluble (i.e., structural) carbohydrate composition per unit mass in *A. americana*; (2) without pretreatment, *A. americana* biomass had the lowest gross heat of combustion, or higher heating/calorific value, compared to high yielding C_4 crops; and (3) after separation of soluble carbohydrates, *A. americana* cellulosic biomass was most easily hydrolyzed by enzymes with greater sugar yield per unit mass compared to the other biomass feedstocks. These results indicate that *A. americana* can produce substantial yields of soluble carbohydrates with minimal water inputs required for cultivation, and fiber portions of the crop can be readily deconstructed by cellulolytic enzymes for subsequent biochemical fermentation.

Keywords: CAM, energy, bioethanol, sorghum, switchgrass, miscanthus, crassulacean acid metabolism, biofuel

OPEN ACCESS

Edited by:
Karolina Heyduk, University of Hawai'i, United States

Reviewed by:
Rachel Burton, The University of Adelaide, Australia
June Simpson, Instituto Politécnico Nacional de México (CINVESTAV), Mexico

***Correspondence:**
Sarah C. Davis
daviss6@ohio.edu

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *13 December 2019*
Accepted: *28 April 2020*
Published: *11 June 2020*

Citation:
Jones AM, Zhou Y, Held MA and Davis SC (2020) Tissue Composition of Agave americana L. Yields Greater Carbohydrates From Enzymatic Hydrolysis Than Advanced Bioenergy Crops. Front. Plant Sci. 11:654. doi: 10.3389/fpls.2020.00654

INTRODUCTION

Agave americana has the potential to produce commercially viable biomass yields in semi-arid climates (Davis et al., 2011, 2014, 2016). *Agave* species use Crassulacean Acid Metabolism (CAM), a photosynthetic pathway that is associated with both greater water use efficiency (WUE) and greater soluble carbohydrate concentrations in leaf tissue compared to most C_4 and C_3 plants (Nobel, 1991; Borland et al., 2009; Davis et al., 2014). The CAM photosynthetic pathway also has

greater theoretical maximum energy conversion efficiency than other photosynthetic pathways (Davis et al., 2014). This study evaluates the tissue composition of an obligate CAM species, *A. americana*, grown as a field crop with variable water inputs, and directly compares the tissue composition, hydrolytic degradation, and products of enzymatic reactions with those of high-yielding C_4 biomass crops.

Advanced lignocellulosic biofuel is produced from either crop residues or stem and leaf biomass of dedicated crops (*i.e.*, structural polysaccharides from plant cell walls), unlike first generation biofuels that are exclusively produced from water soluble saccharides (WSS) in grains or seeds (DOE, 2006; Hess et al., 2007; Epa, 2013, 2015; Stock, 2015). In the case of maize, theoretical dry biomass of stalks (stover) have nearly double the yield of grain biomass and a carbon content per unit mass that is similar to grain (Latshaw and Miller, 1924). Although the first advanced lignocellulosic biorefineries are using maize stover as feedstock due to the abundance of this crop, there are alternative feedstock crops that have greater yields and lower input requirements (e.g., tillage, fertilizer, water, pest management) for cultivation (Somerville et al., 2010).

Advanced biofuel crops, including *Agave* spp., *Miscanthus* x. *giganteus* Greef et Deu. (miscanthus), *Opuntia ficus-indica* (L.) Mill., *Panicum virgatum* (L.) Moench (switchgrass), and *Sorghum bicolor* L. (sorghum) have the potential to reduce greenhouse gas (GHG) emissions associated with biofuel production (Davis et al., 2009, 2013; Somerville et al., 2010; Yan et al., 2011; Cushman et al., 2015). Harvestable dry biomass for *A. americana* is projected to yield up to 9.3 Mg ha^{-1} y^{-1} in semi-arid conditions (Davis et al., 2016). Average biomass yields from the perennial grasses miscanthus and switchgrass are ca. 23.4 Mg ha^{-1} y^{-1} and 10.0 Mg ha^{-1} y^{-1}, respectively (Arundale et al., 2015), while yields of sorghum biomass are 22.0 Mg ha^{-1} y^{-1} on average (Grennell, 2014). Both perennial grasses require less fertilizer compared to conventional agricultural crops because of the ability to efficiently recycle nitrogen (Christian et al., 2006; Heaton et al., 2009; Davis et al., 2010) and the root systems of perennial species allow for greater carbon sequestration compared to annual species (Sartori et al., 2006; Heaton et al., 2009; Davis et al., 2010).

Plants with CAM have the greatest theoretical WUE, and in some cases species have high annual productivity coupled with high concentrations of soluble carbohydrates stored in plant tissue (Nobel, 1990, 1991; Borland et al., 2009; Davis et al., 2011, 2019). Recent research emphasizes the potential for CAM species, such as *Agave* spp. and *Opuntia* spp. (prickly pear), to be grown as biofuel feedstock on marginal or arid lands that are deemed unsuitable for food crops or pasture lands (Smith, 2008; Borland et al., 2009; Chambers et al., 2010; Davis et al., 2011, 2019; Li et al., 2014; Stewart, 2015; Yang et al., 2015; Cushman et al., 2015). *Agave americana* has not been grown in commercial agriculture in the past, but the first field trials and models of productivity indicate it is a viable feedstock crop in semi-arid conditions (Davis et al., 2016; Niechayev et al., 2018). Here, we assess the difference in plant tissue composition of *A. americana* grown with different water inputs to determine if soluble carbohydrates vary as WUE increases in drier conditions.

In addition to potentially high concentrations of WSS, *A. americana* also has lignocellulosic biomass that may be useful in fuel production. In advanced bioethanol production that makes use of lignocellulosic biomass, lignin must be separated from cellulose and hemicellulose to make the carbohydrates available for hydrolysis (Moorhead and Reynolds, 1993; Roche et al., 2009; Humbird et al., 2011); this can be accomplished biologically using specialized enzymes (Evans et al., 1994). Aromatic substances within the primary wall may act as nucleation sites for lignin biosynthesis and may act to cross-link lignin polymers and other polysaccharides to pectin, glycoprotein and/or hemicellulosic matrices. Aromatic residues, such as ferulic acid, are present adjacent to arabinose units of type II xylans and some glucose and arabinose units of pectic polysaccharides (Markwalder and Neukom, 1976; Grabber et al., 2000; Lam et al., 2001), and this effect can increase biomass recalcitrance to enzymatic hydrolysis and exacerbate the persistence of inhibitors (Kim et al., 2011; Jönsson et al., 2013; Jönsson and Martín, 2016).

Both the plant tissue composition and hydrolytic degradation of *A. americana* are evaluated in this study and compared to three advanced biofuel feedstocks that use the C_4 photosynthetic pathway (i.e., miscanthus, switchgrass, and sorghum). Previous work to increase fermentable products from lignocellulosic biomass includes improving pretreatment methods to increase enzymatic hydrolysis (Faik, 2013; Camesasca et al., 2015; Davis R. et al., 2015; Jin et al., 2016), biochemical engineering to improve enzyme activity (Zhang et al., 2006; Raghuwanshi et al., 2014), discovering novel cellulolytic enzymes (Zhang et al., 2006; Rani et al., 2014), optimizing enzyme blends for specific feedstocks (Gao et al., 2010), and decreasing the cost of enzymes (Klein-Marcuschamer et al., 2012; Culbertson et al., 2013; Johnson, 2016). This study specifically evaluated acid hydrolysis, enzymatic hydrolysis with a commercially available enzyme mixture, and saccharification of plant tissue from each of the four feedstocks to quantify the difference in conversion efficiency with difference tissue compositions.

Gross heat of combustion (GH), also referred to as the calorific value or the Higher Heating Value (HHV), relates to biofuel yield because it is negatively correlated with enzymatic digestion efficiency of biomass, and high GH values are associated with high lignin and mineral contents (Demirbaş, 2001; Sheng and Azevedo, 2005; Godin et al., 2013). The energy and chemical inputs required for separating and degrading cellulose in *Agave* spp. may be reduced because of lower lignin concentrations relative to other feedstock crops (Davis et al., 2011). Little research has thus far described the lignocellulosic conversion of *A. americana* and how it directly compares to other lignocellulosic feedstocks of interest (Saucedo-Luna et al., 2011; Corbin et al., 2015; Mielenz et al., 2015; Yang et al., 2015; Pérez-Pimienta et al., 2018; Wang et al., 2019).

This study evaluates the potential bioethanol yield from *A. americana* by resolvingGH was adjusted for moistur e(1) carbohydrate concentrations, (2) overall differences in plant tissue composition relative to other advanced biofuel feedstocks, and (3) enzymatic conversion efficiencies of cellulose and hemicellulose. The potential energy yield of *A. americana*, a vigorous CAM plant, is also compared with the potential

energy yield of high-yielding C_4 crop species (miscanthus, switchgrass, and sorghum).

MATERIALS AND METHODS

Plant tissue compositional analysis, energy content analysis, and enzymatic hydrolysis were completed to determine differences in biomass quality and hydrolytic conversion efficiency of *A. americana* relative to other advanced biofuel crops. Plant tissue samples from two field sites, one with an *A. americana* crop and the other with switchgrass, miscanthus, and sorghum crops, were analyzed using identical Laboratory Analytical Procedures (LAP) developed by the National Renewable Energy Laboratory (NREL) to determine the percentage of carbohydrates within biomass. Carbohydrate concentrations of *A. americana* grown in a field experiment with different irrigation treatments were resolved, compared, and then regressed against the WUE associated with different water inputs, as determined in previous work (Davis et al., 2016). Both GH and sugar yield from enzymatic hydrolysis were measured in all four feedstocks and compared. Detailed methodologies are provided in the following sections.

Field Characteristics for *A. americana*

A 3-acre site, located at the University of Arizona Maricopa Agricultural Center (elevation, 363.77 m; 33° 05′ N, 111° 97′ W), was divided into eight plots, each containing six 15 m^2 subplots (Davis et al., 2016). In 2012, forty-nine individual *A. americana* were planted in two randomly selected subplots within each plot; individuals were spaced 2 m apart, within and along rows. The two planted subplots in each plot were designated as one of four different irrigation treatments for the duration of the experiment (4 years). The plots were flood irrigated at different levels (100, 260, 330, and 580 mm y^{-1}), and the field received ∼200 mm mean annual precipitation over four years (2012–2016). Mean annual water inputs (MWI) for the treatments (including both irrigation and precipitation) totaled 300, 460, 530, and 780 mm y^{-1}, respectively. The soils at this site are characterized as sandy clay loam. For more information pertaining to field conditions at this site, see Davis et al. (2016). Samples used in this study were harvested and dried in the oven as per Hames et al. (2008a) in January 2016 and stored in paper bags at room temperature in a dry, dark place until analysis was accomplished.

Field Characteristics for Temperate Grasses

Experimental plots of advanced cellulosic bioenergy crops were established at the Ridges Land Lab at Ohio University (elevation 222 m; 39° 19′ N, 82° 06′ W) on abandoned agricultural land; one site (site "A") was historically used for pasture, and the other site (site "B") was historically managed for hay. Soils in this area are mostly Aquic Hapludalfs and Typic Hapludalfs, and the land was unmanaged from 1992 until the time this experiment was established in 2013. Three replicate 10 m × 10 m plots for each of the advanced biofuel feedstock species were randomly assigned within each site with 2 m spacing between plots that were mowed regularly (**Supplementary Figure 1**).

Miscanthus rhizomes were donated by Tom Voigt from the University of Illinois at Urbana-Champaign (UIUC), and planted May 8, 2013. Sorghum seed (CHR-FS4; Chromatin; New Deal, TX, United States) was donated by Dr. Pat Brown from UIUC, planted in rows during the first week of June spaced 1 m apart and seeds within a row spaced 5 cm, and fertilized the second week of June (2013, 2014, and 2015). Switchgrass seed ('Timber' var.) was sourced from Ernst Conservation Seed and planted April 25, 2013. Sorghum was harvested in November 2015 following senescence, and air-dried for 6 weeks. Switchgrass and miscanthus were harvested in February 2016 following senescence; above-ground biomass was harvested, bundled and air-dried in the field for 4 weeks as would be expected in biomass crop management. Subsamples were oven dried for tissue composition analysis as described in subsequent sections.

Analysis of GH

Gross heat of combustion was measured using a Parr 6200EA isoperibol (oxygen bomb) calorimeter (Parr Instrument Company; Moline, IL, United States) in conjunction with a Parr 6510 water handling system to maintain constant temperature (Parr Instrument Company) per method ASTM D5865. Benzoic acid (GH = 6318 cal g^{-1}) was used as a standard for calibration. GH was adjusted for moisture content (GH_{OD}) using equation 1 (Boundy et al., 2011).

$$GH_{OD} = \frac{GH_{sample}}{\left(\frac{\%\ Total\ \ Solids}{100}\right)} \quad (1)$$

The GH value was converted to MJ kg^{-1} from cal g^{-1} using 0.004187 as a conversion factor. Measured GH of samples was compared to theoretical values calculated from lignin contents using equations 2a and 2b from Demirbaş (2001).

$$GH_1 = 0.0889\,(\%L) + 16.8218 \quad (2a)$$

$$GH_2 = 0.0877\,(\%L) + 16.4951 \quad (2b)$$

where GH_n is the theoretical GH value (MJ kg^{-1}) and %L is percent lignin. The percent lignin was measured according to methods described in section "Determination of Lignin."

Composition Analysis

Composition analysis of field grown *A. americana* was performed per LAPs developed by NREL and Moorhead and Reynolds (1993) wherein sequential fractionation was used to determine extractives, minerals, cellulose, lignin, and ash contents. The review by Sluiter and Sluiter (2011) provides mass closure for the LAPs published by NREL, and the herbaceous feedstocks calculation spreadsheet published by NREL was used in this work. A workflow is given in **Figure 1**. Two randomly selected replicate samples were analyzed from each subplot ($n_{subplot}$ = 1), which equates to eight replicates per irrigation treatment ($n_{treatment}$ = 4), and a total N = 16 for all four treatments. In the case of the C_4 grass feedstocks,

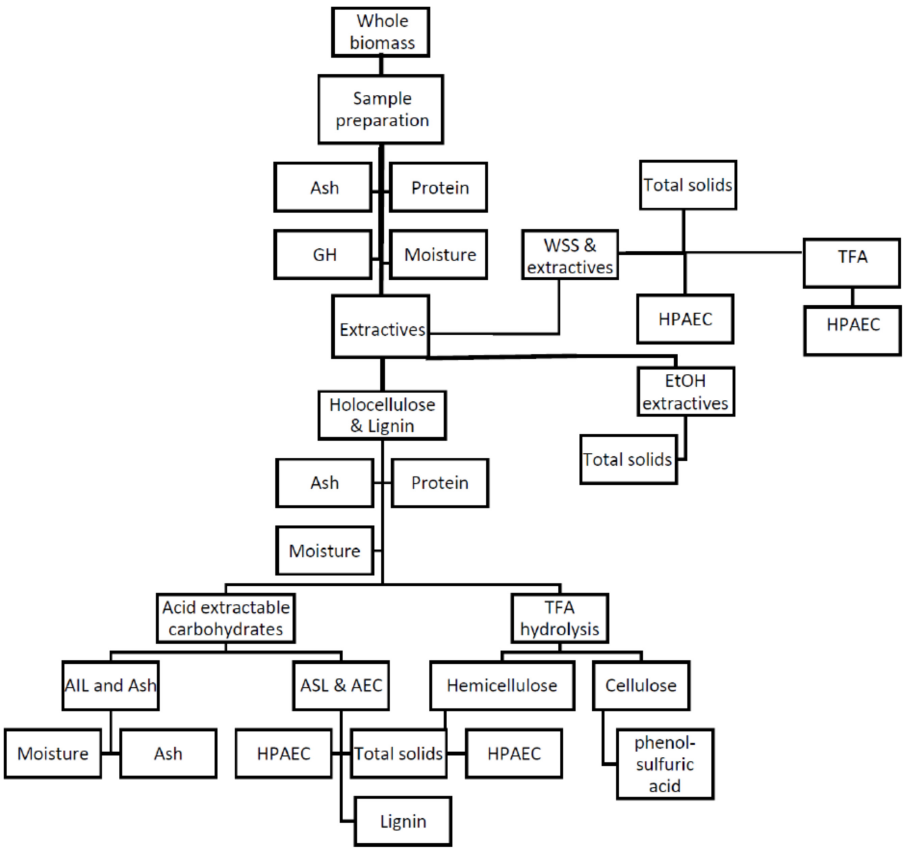


FIGURE 1 | Flowchart outlining the composition analysis procedures for *A. americana* adapted from Sluiter and Sluiter (2011). Whole biomass sample is prepared, extracted with water to remove water-soluble carbohydrates (WSS), extracted with ethanol to remove inhibitors from residue containing holocellulose and lignin, which is then hydrolyzed to solubilize some lignin (ASL) and acid extractable carbohydrates (AEC) (*i.e.*, structural carbohydrates) and leave a residue containing insoluble lignin (AIL) and minerals. Water soluble extract and holocellulose fraction undergo TFA hydrolysis and analyzed by HPAEC, and TFA insoluble is analyzed for cellulose content by phenol-sulfuric acid assay. Lignocellulosic grasses were only be analyzed for GH, WSS, and AEC. Procedure short titles are provided in non-shaded boxes, and biomass fraction name is provided in shaded ovals.

WSS and acid-extractable carbohydrates were analyzed per NREL protocols for comparison with *Agave,* and then also analyzed for GH.

Tissue Sample Preparation

Leaves from *A. americana* were harvested from the plants at the end of January 2016, dried in an oven as per Hames et al. (2008a), then shipped to Ohio University. Plant tissue samples from the grass species were collected following field drying, and oven-dried to remove residual moisture that remained following field drying. Oven dried samples were stored in paper bags at ambient room temperature and processed as per Hames et al. (2008a). In summary, samples were ground to uniform particle size using a Wiley mill (Thomas Scientific, 3383L10, United States) with a 20 mesh (0.85 mm) screen. Once samples were ground, they were kept in 3.79 L airtight polyethylene zipper storage bags at 2°C.

Determination of % Total Solids

Total solids and moisture present in biomass and liquid samples was determined gravimetrically from methods from Sluiter et al. (2008a). Total solids were calculated as a percentage of mass (%), given in equation 3.

$$\%Total\ Solids = \frac{M_{dry}}{M_{wet}} \times 100 \quad (3)$$

Determination of Protein Content

Protein content was determined by quantifying percent N via combustion and calculated using an appropriate N factor for the plant amino acid profile. Percent N content of biomass was determined by combustion on a Costech ECS 4010 (Costech Analytical Technologies, Inc., Valencia, CA, United States). Samples (5.0 ± 0.5 mg and 7.5 ± 0.5 mg for whole biomass and extractive-free biomass (EFBM), respectively) were combusted and detected in the gas phase using a thermal conductivity detector and a Colibrick A/D transducer (U32; DataApex→ Ltd., Prague, Czechia) with Clarity→ Elemental Analysis Software (C50; DataApex→ Ltd.). The measured %N was used to calculate percent protein with a protein factor (PF) of 4.6, as determined suitable for biomass without a

characterized amino acid profile by Hames et al. (2008b), using equation 4.

$$\%Protein = PF \times \%N \quad (4)$$

Protein content was determined in whole biomass to calculate whole protein content in the sample and EFBM, and was not determined in the post-hydrolysis solid residue. It was assumed that minimal protein co-condensation occurred within the post-hydrolysis residue because dried hydrolysate had very low %N (as observed when determining acid-soluble lignin).

Determination of Ash

Percent ash in biomass was determined by methods that were adapted from the LAP by Sluiter et al. (2008b). Ash content was measured in whole biomass, EFBM, and the post-hydrolysis residue. Percent ash was estimated based on oven dry mass of the sample (M_{dry}), as follows:

$$\%Ash = \frac{(M_{crucible+ash} - M_{crucible})}{M_{dry}} \times 100 \quad (5)$$

Determination of Water-Soluble Extractives and Carbohydrates

A water extraction was performed to remove WSS and other water-soluble compounds by adding 15 mg ± 2.5 mg biomass ($M_{WholeBM}$) to preweighed 25 mL centrifuge tubes (W_{tube}). 12.5 mL distilled water (dH_2O) was added and the tube was sealed and placed into a water bath at 60°C for 30 min. The mixture was centrifuged (∼125 × *g*) for 15 min and the supernatant collected and stored. Two more water extractions were performed and combined for a total of 37.5 mL extract. A subsample (5 mL) was used for determination of total dissolved solids in the water extract.

Water soluble saccharides were quantitatively determined from a 0.75 ml subsample. Samples were centrifuged and dried under a steady stream of air. Dried extracts were dissolved in 0.75 mL cellobiose (20 mM in dH_2O) (C7257; Sigma-Aldrich Co., LLC.), filtered through 0.22 μm nylon spin filters (8169; Corning Inc., Corning, NY, United States), and appropriately diluted (1:1000 and 1:100 for *A. americana* and grasses, respectively) with de-ionized water. Aliquots (25 μL) were fractionated by high-pH anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) (Dionex Corp., Sunnyvale, CA, United States) fitted with a Carbo-Pac PA20 anion-exchange column (Dionex Corp.) and similar guard column (50 × 4 mm i.d.; Dionex Corp.). Elution proceeded as an isocratic flow (0.5 ml min^{-1}, for all samples) in 40 mM sodium hydroxide (NaOH) for 4 min, followed by a linear increase to 40 mM NaOH and 40 mM sodium acetate over 5 min. Peak areas of the fractionated sugars were integrated and quantified by molar response factors generated from the peak areas measured for sugar standards D-(+)-glucose (Glc) (G8270; Sigma-Aldrich Co., LLC.; St. Louis, MO, United States), D-(−)-fructose (Fru) (F0127; Sigma-Aldrich Co., LLC.), and sucrose (Suc) (S7903; BioXtra; Sigma-Aldrich Co., LLC.) using cellobiose as an internal standard (0.5 nmol, each).

Determination of Ethanol-Soluble Extractives

Removal of extractives to achieve EFBM was necessary to avoid inhibitors that can negatively impact biomass analysis and acid hydrolysis of structural oligomeric carbohydrates (Thammasouk et al., 1997; Sluiter et al., 2008c). Samples were extracted from plant tissues using a method adapted from Moorhead & Reynolds (1993).

Percent extractives were calculated from equation 6.

$$\%Extractives = \frac{M_{Whole\ BM} - (M_{EFBM+tube} - M_{tube})}{W_{Whole\ BM}} \times 100 \quad (6)$$

where $M_{WholeBM}$ is the oven dry weight of whole biomass and M_{tube} is the mass of the preweighed tube.

Determination of Acid-Extractable (Structural) Carbohydrates

An acid hydrolysis using sulfuric acid (H_2SO_4) was performed to solubilize structural carbohydrates and acid soluble lignin and separate these components from the residue containing insoluble lignin and structural inorganic material. Dry EFBM (0.50g) were hydrolyzed using methods adapted from Moorhead and Reynolds (1993) and Sluiter and Hames (2012). Oven dry EFBM mass (ODM_{sample}) was calculated from equation 7.

$$ODM_{sample} = \frac{M_{EFBM} \times \%\ TS}{100} \quad (7)$$

where % total solids was calculated using equation 3.

Acid insoluble residue (AIR) was calculated by equation 8:

$$\%AIR = \frac{M_{filter+sample(HBM)} - M_{filter}}{ODM_{sample}} \times 100 \quad (8)$$

AIR was prepared from *A. americana* and the grasses following Held et al. (2011) to determine composition of hemicellulose. Hemicellulose sugar compositions were determined on AIR from *A. americana* and grass samples as per York et al. (1986). AIR samples (25 μg) and standards (100 nmol) were hydrolyzed in 2N trifluoroacetic acid (TFA) at 121°C for 90 min. Samples were cooled to room temperature then centrifuged (3000 × *g*) for 3 min. TFA was evaporated under a steady stream of nitrogen at 45°C. All samples were then dissolved again in 500 μL deionized water, filtered using Corning Costar Spin-X centrifuge tube filters (CLS8160), and then stored at −20°C. Aliquots (25 μL) were fractionated by HPAEC (Dionex Corp., Sunnyvale, CA, United States) fitted with a Carbo-Pac PA20 anion-exchange column (Dionex Corp.) and similar guard column (50 × 4 mm i.d.; Dionex Corp.) as per Øbro et al. (2004). Peak areas of the fractionated sugars were integrated and quantified by molar response factors generated from the peak areas of external sugar standards (4-point calibrations). Following TFA hydrolysis, an additional phenol-sulfuric acid hydrolysis was performed on the TFA-insoluble pellet as per

DuBois et al. (1956) to determine the proportion of cellulose using Glc as a standard.

Additionally, TFA hydrolysis was performed on the water extract of *A. americana* to resolve the amount of oligomeric sugar present as starch and fructans within the liquid. The TFA water extract hydrolysate was analyzed by HPAEC as described above to measure Glc, Fru, and Suc from oligomers.

Determination of Lignin

Some lignin was removed from the material during acid extraction, while some remained insoluble within the cell wall matrix. Acid soluble lignin present within the hydrolysate was analyzed within six hours and determined from the concentration of dissolved organic carbon (DOC) (Browning, 1967; Lin and Dence, 1992; Weishaar et al., 2003; Sluiter and Hames, 2012). An aliquot (5 mL) of hydrolysate was used for DOC quantification. Acid soluble lignin was determined by comparing absorbance at 320 nm (A_{320}) against a blank containing 4% H_2SO_4; for lignin from maize, absorbance is greatest between 260–380 nm with relative maxima at 280 and 320 nm (Müse et al., 1997). ASL was calculated from equation 9:

$$\%ASL = \frac{A_{320} \times V_{filtrate} \times D}{SUVA_{320} \times ODM_{sample}} \times 100 \quad (9)$$

where A_{320} is the average ultraviolet (UV) absorbance at 320 nm, $V_{filtrate}$ is the hydrolysate volume (86.73 mL), D is a dilution factor if used, $SUVA_{320}$ (L g^{-1} cm^{-1}) is the Specific UV Absorbance (SUVA) at 320 nm ($SUVA_{320}$), ODM_{sample} is the mass of the sample (mg) from equation 7. SUVA, also described as the specific absorptivity or the molar absorbance coefficient for a given concentration of dissolved carbon (Browning, 1967; Lin and Dence, 1992), was used to estimate the percent lignin by colorimetrically measuring aromaticity; results of analysis of dissolved organic carbon (DOC) and humic substances by Weishaar et al. (2003) showed 97% correlation between SUVA and ^{13}C nuclear magnetic resonance (NMR) values for aromaticity. $SUVA_{320}$ was calculated from equation 10,

$$SUVA_{320} = \frac{A_{320}}{[DOC]} \times L_{path} \quad (10)$$

where A_{320} is the measured absorbance at 320nm. [DOC] was determined by combustion of dried down hydrolysate on a Costech ECS using the same method as described previously for protein determination, but percent carbon (%C) was used in lieu of %N, and L_{path} was the path length (cm) of the spectrophotometer cell; [DOC] was calculated by drying down 5mL of H_2SO_4 hydrolysate at 105°C for 48 h and combusting a sample (5.5 ± 0.5 mg) as previously described, from equation 11,

$$[DOC]\,(g\,L^{-1}) = \%C \times \left(\frac{M_{sample+pan} - M_{pan}}{V_{sample}}\right) \quad (11)$$

where %C was determined from combustion, M_{pan} is the mass of the drying pan (g), $M_{sample+pan}$ is the mass of the pan and dried sample (g), and V_{sample} is the volume (L) of liquid sample dried down. Acid insoluble lignin (AIL), or Klason lignin, was calculated using equation 12.

$$\%AIL = \frac{M_{sample(HBM)} - M_{ash}}{ODW_{sample}} \times 100 \quad (12)$$

Percent lignin on an extractives free basis can be calculated using equation 13,

$$\%Lignin_{Ext.\ Free} = \%AIL + \%ASL \quad (13)$$

where %AIL was calculated from equation 12 and %ASL from equation 9. Total lignin, based on whole biomass was calculated from equation 14,

$$\%Lignin_{whole} = \left(\%Lignin_{Ext.\ Free}\right) \times \frac{(100 - \%Extractives)}{100} \quad (14)$$

where $\%Lignin_{Ext.Free}$ was calculated from equation 13, and %Extractives is from equation 6.

Comparison of Enzymatic Digestibility

Enzymatic hydrolysis conversion efficiency was evaluated for each of the four feedstocks to determine biomass recalcitrance using purified and lyophilized enzymes. Purified cellulase from *Trichoderma reesei* ATCC 26921 (C8546; Sigma-Aldrich), xylanase from *Trichoderma viride* (X3876; Sigma-Aldrich), and beta-glucosidase from almonds (49290; Sigma-Aldrich) were used for hydrolysis. Prior to comparing enzymatic digestibility of a biomass substrate, protein content of the enzyme solution was determined using a Pierce bicinchoninic acid (BCA) protein quantification kit (BCA1; Sigma-Aldrich Co., LLC.). The BCA quantification assay was recommended for fungus-derived enzymes because cellulase produced by the species *T. reesei* contains roughly 50% cellobiohydrolase I (CBH I; Cel7A), and this cellulolytic enzyme and others within the same glycoside hydrolase family (GHF) 7 may react poorly to other calorimetric assays (Resch et al., 2015). Protein content of the enzyme cocktail determined from the Pierce BCA protein assay was 14.47 mg mL^{-1} ± 1.17 mg mL^{-1} for the mixed enzymes. Three samples from each of the four plant species studied here were subjected to enzymatic hydrolysis treatments using the mixed enzymes.

Low Solids Loading Enzymatic Hydrolysis

Enzymatic saccharification of EFBM using low solids loading was performed according to Resch et al. (2015). EFBM was washed three times with 30mM sodium citrate buffer (pH 5.0) containing 0.002% (w v^{-1}) sodium azide to remove soluble sugars and inhibitors, and the residue was suspended in the same buffer at a loading rate of 2.0% (w v^{-1}). A sample containing 0.014 g EFBM, corrected for moisture, was quantitatively transferred to a screwcap test tube, and 42 µL 1.0 M sodium citrate buffer (pH 5.0) was added. 5.6 µL of 5.0% (w v^{-1}) sodium azide was added to each vial to inhibit microbial contamination. The volume

of enzyme solution used was calculated from equation 15, so 20.0 mg protein from enzyme was available per g glucan,

$$Enzyme\ volume = \frac{1.0\ mL}{X mg\ protein} \times \frac{20.0\ mg\ protein}{1.0\ g\ glucan} \times g\ glucan \quad (15)$$

where X was mg protein as calculated from 1.0mL enzyme solution using BCA kit. Released sugars (*e.g.*, cellobiose, Glc, and Xyl) were analyzed using HPAEC, as previously described in the determination of structural carbohydrates.

Determination of Liberated Sugars

Resulting sugar concentrations (mg mL^{-1}) from each digestion mixture were used to determine percent conversion during hydrolysis as the amount of sugar liberated (i.e., concentration hydrolyzed oligomeric sugar) divided by the initial mass of EFBM, as in equation 16,

$$\%\ hydrolysis = \frac{\frac{mg\ sugar}{mL} \times 14\ mL}{0.14\ g\ EFBM} \times 100 \quad (16)$$

where sugar (mg mL^{-1}) is quantified by HPAEC.

Statistical Analysis

Analysis of Variance (ANOVA) was used to statistically compare tissue composition, GH of biomass, and products of enzymatic digestion from *A. americana* plants that were grown with different MWI, and to compare these characteristics of biomass between the four crop species. *Post hoc* Tukey's honest significant difference (HSD) was used to resolve significant differences between treatment groups or species with $\alpha = 0.05$.

RESULTS

Composition Analysis

Composition analysis of the feedstocks measured in this study indicated that the leaf tissue of *A. americana* differed substantially from the grass feedstocks (**Table 1**). *Agave americana* had the largest mean proportion of water soluble saccharides compared to miscanthus, sorghum, and switchgrass (0.25, 1.14, and 0.28%, respectively; $p = 7.42 \times 10^{-9}$) and much lower hemicellulose (5.63%) relative to miscanthus (24.53%), sorghum (21.83%), and switchgrass (24.22%) ($p = 1.11 \times 10^{-16}$; **Table 1**). Cellulose was also significantly different ($p = 5.22 \times 10^{-15}$; **Table 1**), with *A. americana* having the lowest cellulose (5.55%) relative to the other feedstocks, and miscanthus having the greatest cellulose content (32.43%). Total holocellulose contents for the four feedstocks in descending order were 59.96% for miscanthus, 46.73% for switchgrass, 41.50% for sorghum, and 11.18% for *A. americana* (**Table 1**). The C_4 crops had similar tissue compositions to one another, and miscanthus and switchgrass were the most similar (**Table 1**).

TABLE 1 | Composition analyses from experimental and literature values for dry untreated biomass for *A. americana (Ag), M. x giganteus (Mi), S. bicolor (So)*, and *P. virgatum (Sw)*. Water soluble saccharides (WSS), hemicellulose (Hc), cellulose, holocellulose (HoC), lignin, and carbon (C) are reported as a percentage of dry biomass.

Species	%WSS	%Hc	%Cellulose	%HoC	%Lignin	%C
Ag	43.95*a*	5.63*a*	5.55*a*	11.18*a*	9.06	38.86*a*
Mi	0.25*b*	24.53*b*	32.43*c*	56.96*c*	–	45.59*c*
So	1.14 *b*	21.84*b*	19.66*b*	41.50*b*	–	42.38*b*
Sw	0.28 *b*	24.22*b*	22.51*b*	46.73*b*	–	42.99*b*

Lowercase italicized letters designate significant differences within a column as resolved from one-way ANOVA and Tukey's HSD test ($p < 0.05$).

TABLE 2 | Mean annual water input (MWI) in mm y^{-1}, water-soluble saccharides (WSS) as % of dry biomass, WSS yield (Mg ha^{-1} y^{-1}) calculated based on biomass yield reported in Davis et al. (2016), and WSS WUE (WUE_{WSS}) expressed as WSS yield per MWI (kg WSS ha^{-1} mm^{-1}) for *A. americana*.

MWI[1] (mm y^{-1})	%WSS	WSS Yield (Mg ha^{-1} y^{-1})	WUE_{WSS} (kg WSS ha^{-1} mm^{-1})
300	42.45 ± 6.08	1.70 ± 0.24*a*	5.66 ± 0.81*ab*
460	46.49 ± 5.69	3.16 ± 0.39*b*	6.87 ± 0.84*ab*
530	46.53 ± 3.67	4.33 ± 0.34*b*	8.17 ± 0.64*a*
780	45.37 ± 4.07	3.31 ± 0.30*b*	4.25 ± 0.38*b*

Standard error indicated after each mean; n = 4. Lowercase italicized letters designate significant differences within a column as resolved from one-way ANOVA and Tukey's HSD test ($p < 0.05$). Where no letters are indicated, there was no significant difference ($p > 0.05$).

WSS in *A. americana*

Water soluble extracts in *A. americana* ranged from 42.21% to 45.37% of dry leaf biomass and contained soluble mono-, di-, and oligosaccharides (**Table 2**). Water extracts contained 16.3% total mono- and disaccharides Glc, Fru, and Suc in a ratio of 10:6:6, respectively ($n = 16$). There was no significant difference between total mono- and disaccharide content present in the water-extracted fraction from plants grown with different irrigation treatments ($p = 0.335$). Mean percent mono- and disaccharides were 7.46, 4.42, and 4.42% for Glc, Fru, and Suc, respectively ($n = 16$). There was no significant difference for individual monosaccharide contents of the water extract between plants from different irrigation treatments ($p = 0.435$, 0.769, 0.218 for Glc, Fru, and Suc, respectively; **Figure 2a**).

There was also a large percentage of water soluble material (∼35%) that was not identified as monosaccharide by HPAEC. TFA hydrolysis of the water extract yielded mostly monosaccharides derived from water-soluble oligosaccharides and the unresolved fraction was subjected to a phenol-sulfuric acid assay to determine the unknown oligosaccharides in glucose equivalents. Oligosaccharides present within the water extractable fraction of *A. americana* leaf tissue represented nearly 10% of dry biomass (**Figure 2b**). There was no significant difference in Glc nor Suc from oligomers across irrigation treatment ($p = 0.18$ and 0.55, respectively). There was a significant difference between Fru from oligomers across differing water inputs ($p = 0.0008$); maximum Fru content was measured in biomass from crops receiving MWI of 460 mm y^{-1} (6.19 ± 0.61%) while the lowest Fru

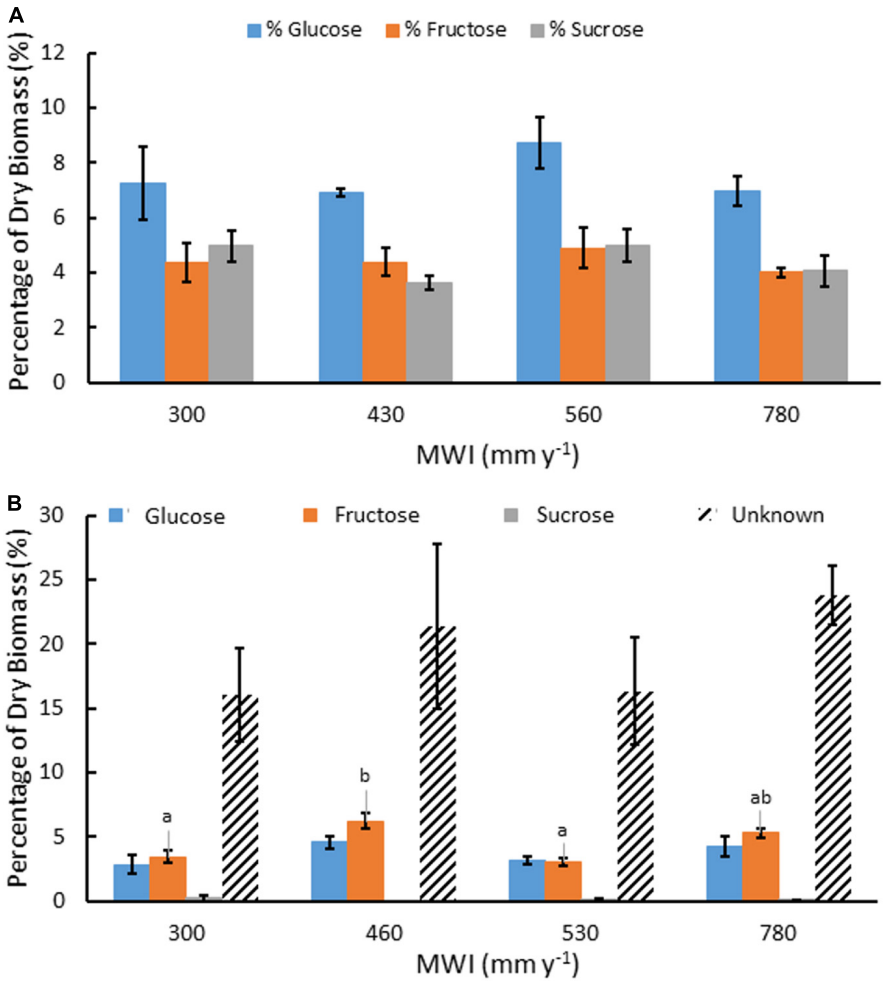


FIGURE 2 | Mean concentration of water soluble Glc, Fru and Suc as a percentage of dry biomass **(A)** and saccharides produced after TFA hydrolysis of water soluble extracts including Glc, Fru, Suc, and unknown sugars from oligosaccharides **(B)**, for *A. americana* under different MWI (mm y^{-1}). Error bars indicate standard error of the mean ($n = 4$). Letters indicate significant differences between species in Fru ($p < 0.05$) from one-way ANOVA and Tukey's HSD test. No significant difference in the other sugars were observed between species ($p > 0.05$).

content was 3.06% (± 0.30%) in plants receiving MWI of 530 mm y^{-1}.

In *A. americana*, WUE was found to be positively, but weakly, correlated with percent water soluble mono- and disaccharides ($y = 0.1949x + 42.526$, $R^2 = 0.12018$; **Figure 3b**). The total percentage of WSS was found to be correlated with dry yield (kg ha^{-1}), using a positive second-order polynomial equation ($y = -0.1736x^2 + 3.0531x + 33.059$; $R^2 = 0.90728$) (**Figure 3a**). Water soluble monosaccharide yields were largest in individuals that were grown with 560 mm MWI (**Figure 3d**). WSS yield was highly correlated ($y = 0.4952x - 0.2675$; $R^2 = 0.99848$) with dry biomass yield (**Figure 3c**) because there was no significant difference in the concentration of WSS across irrigation treatment. Similarly, MWI and WSS per unit water (WUE_{WSS}; kg WSS ha^{-1} mm^{-1}) exhibited a similar trend with a maximum WUE *ca.* 500mm y^{-1} (**Figure 3e**). There was no significant difference associated with WSS concentrations across irrigation treatment, and there was a strong linear correlation between WUE_{WSS} and WUE ($y = 0.4855x - 0.4422$; $R^2 = 0.97796$; **Figure 3f**).

Comparison of WSS Between Feedstocks

Water soluble extracts derived from *A. americana* and the temperate grasses differed by total percent recovered and carbohydrate composition. For *A. americana* measured in this study, water soluble components comprised 52.07% of dry biomass while water soluble compounds only contributed to 6.82, 11.60, and 2.79% of dry biomass in miscanthus, sorghum, and switchgrass, respectively. Water soluble mono- and disaccharides also represented a much larger fraction of total biomass in *A. americana* compared to the temperate grass species (**Figure 4**). Individual mono- and disaccharide contents greatly differed between the temperate grass species miscanthus, sorghum, and switchgrass ($p = 0.001$), having Glc:Fru:Suc ratios of 6:5:1, 1.3:1.7:1, and 4:1:1, respectively (**Figure 4**).

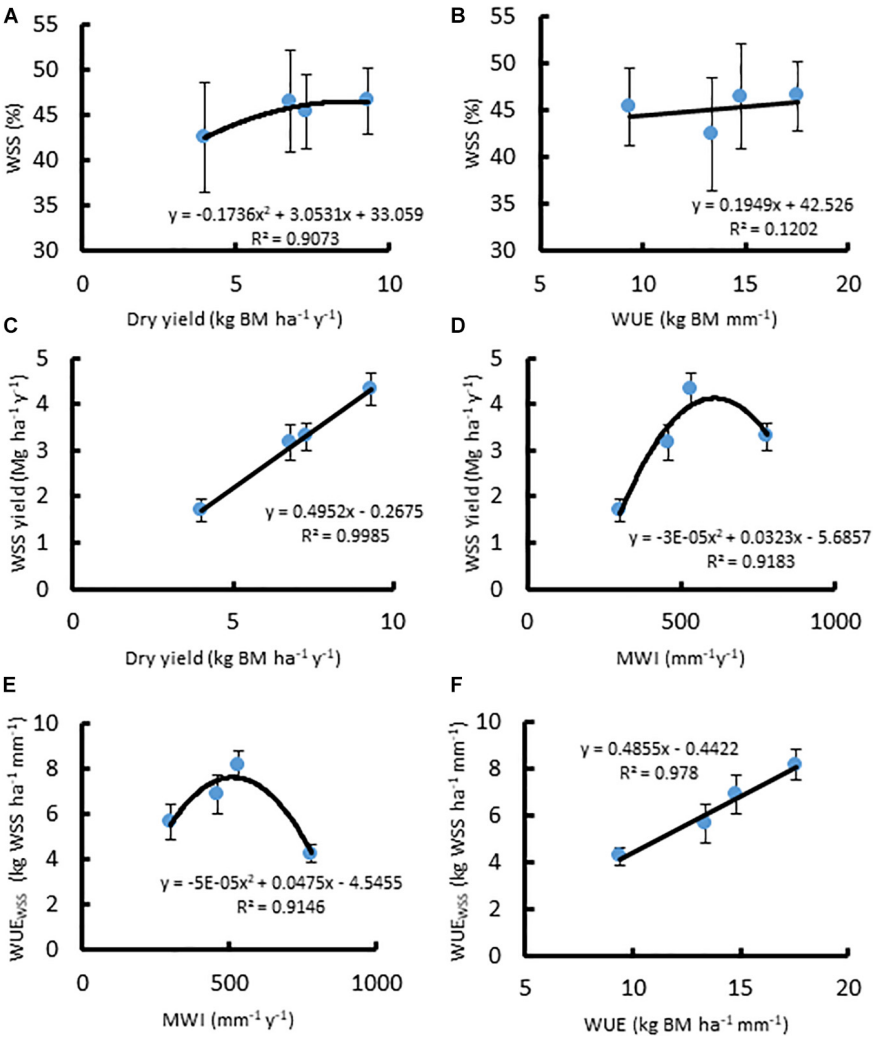


FIGURE 3 | Relationships between **(A)** dry biomass yield (Mg ha^{-1} y^{-1}) and WSS (%), **(B)** WUE (kg ha^{-1} mm^{-1}) and WSS (%), **(C)** dry biomass yield and WSS yield (Mg ha^{-1} y^{-1}), **(D)** MWI (mm y^{-1}) and WSS yield, **(E)** MWI and WUE_{WSS} (kg WSS ha^{-1} mm^{-1}), and **(F)** WUE and WUE_{WSS}, for *A. americana*. Error bars indicate standard error of the mean; $n = 4$.

Structural Polysaccharide Composition in *A. americana*

Composition of *A. americana* leaf biomass was examined to determine the relationship between WUE and structural carbohydrates in the CAM species *A. americana*. Results from TFA hydrolysis of *A. americana* leaf EFBM yielded a mean (n = 16) of 5.63% fermentable monosaccharides. There was no significant difference among irrigation treatments in uronic acid content (p = 0.59), or monosaccharides liberated by TFA (p = 0.54) besides Glc; per one-way ANOVA and Tukey's HSD test, there was a significant difference (n = 4; p = 0.0266) between the Glc content in samples from growing conditions with the lowest (300 mm) and highest (780 mm) MWIs (1.0275% ± 0.092% and 0.605% ± 0.094%, respectively) (**Supplementary Table 1** and **Figure 5**). Uronic acids in the form of GalA, a component of pectins HGA and xylogalacturonan, were present in the largest proportion across all treatments (2.76%) compared to other hemicellulose components liberated by TFA hydrolysis.

Phenol-sulfuric acid hydrolysis of the insoluble pellet yielded an average percent cellulose in *A. americana* of 5.55% ± 0.47% (n = 16). There was no significant difference in cellulose content across irrigation treatment per one-way ANOVA and Tukey's HSD test (p = 0.168). Cellulose content was 6.74% ± 0.83%, 4.11% ± 0.70%, 6.33% ± 0.50%, 5.01% ± 1.22% for individuals receiving water inputs of 300, 460, 530, and 780 mm y^{-1}, respectively (n = 4).

Comparison of Structural Polysaccharide Composition Between Feedstocks

Compared to *A. americana*, the grasses had a larger percentage of dry biomass in structural polysaccharides (**Figure 6**). TFA hydrolysis of EFBM yielded more total monosaccharides in the grass feedstocks (24.53% ± 0.16%, 21.84% ± 0.32, and

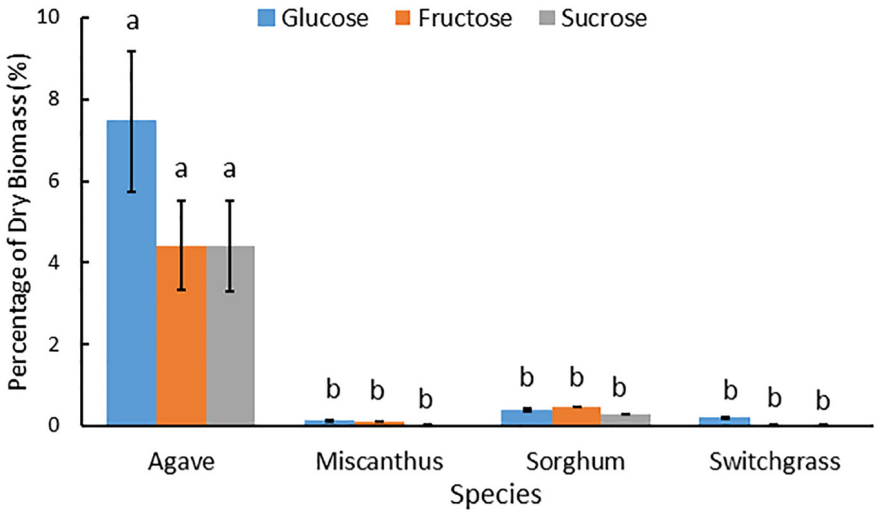


FIGURE 4 | Mean Glc, Fru and Suc as a percentage of dry biomass composition for miscanthus, sorghum, and switchgrass and then all feedstocks including *A. americana* at a different scale to show relative differences. Error bars indicate standard error of the mean (n = 16 for *A. americana* and n = 3 for grasses). All three sugar concentrations were significantly greater in *A. americana* than in the perennial grasses ($p < 0.05$) from one-way ANOVA and Tukey's HSD test of differences between species within each sugar type, as indicated by lowercase letters.

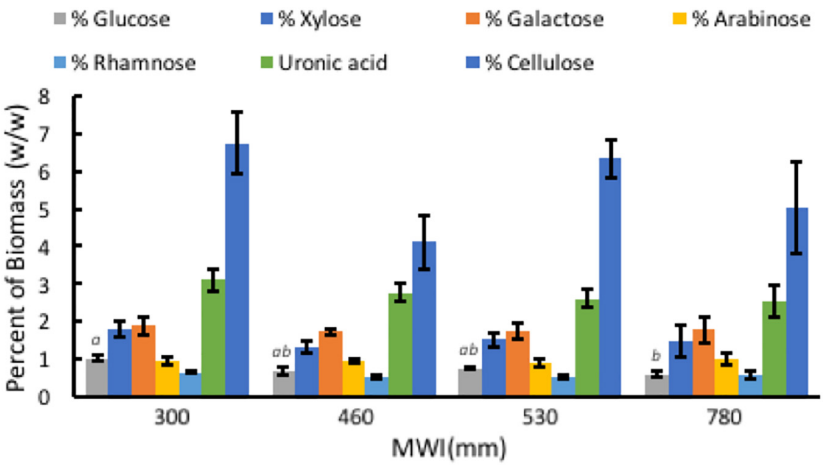


FIGURE 5 | Relationships associated with MWI (mm y^{-1}) and polysaccharide composition in dry biomass (%) from TFA hydrolysis and phenol-sulfuric acid hydrolysis of EFBM showing percent glucose, xylose, galactose, arabinose, rhamnose, uronic acid, and cellulose for *A. americana*. Error bars indicate standard error of the mean; n = 4. Lowercase letters above bars indicate significant differences in glucose ($p < 0.05$) from one-way ANOVA and Tukey's HSD test. No significant difference in the other sugars were observed between species ($p > 0.05$).

24.22% ± 2.04 for miscanthus, sorghum, and switchgrass, respectively) compared to *A. americana* (5.63% ± 1.23%), however, *A. americana* had a larger mean uronic acid content (2.76% ± 0.59) compared to the other species (0.64% ± 0.05, 0.55% ± 0.03, and 1.04% ± 0.24, respectively) (**Figure 6**). For all grasses, the largest proportion of monosaccharides from hemicellulose in descending order of abundance were xylose, glucose, arabinose, galactose, and rhamnose. This differs from *A. americana* with galactose as the most abundant fermentable monosaccharide present in hemicellulose (**Figure 6**). Results from phenol-sulfuric acid hydrolysis of the grasses showed a greater percentage of cellulose contents compared to *A. americana* ($p = 5.22 \times 10^{-15}$); cellulose content was 32.43% ± 2.67%, 19.66% ± 0.00%, and 22.51% ± 0.50% for miscanthus, sorghum, and switchgrass, respectively, and was 5.55% ± 0.47% for *A. americana* (**Figure 6**).

Analysis of GH

Measured GH of *A. americana* EFBM leaf tissue was 15.44 MJ kg^{-1} ± 0.15 MJ kg^{-1} (n = 16), and there was no significant difference between groups receiving differing MWIs as per one-way ANOVA (p = 0.096). Of the different feedstocks tested, *A. americana* leaves had the lowest mean measured GH from oxygen bomb calorimetry (15.44 MJ kg^{-1} ± 0.15 MJ

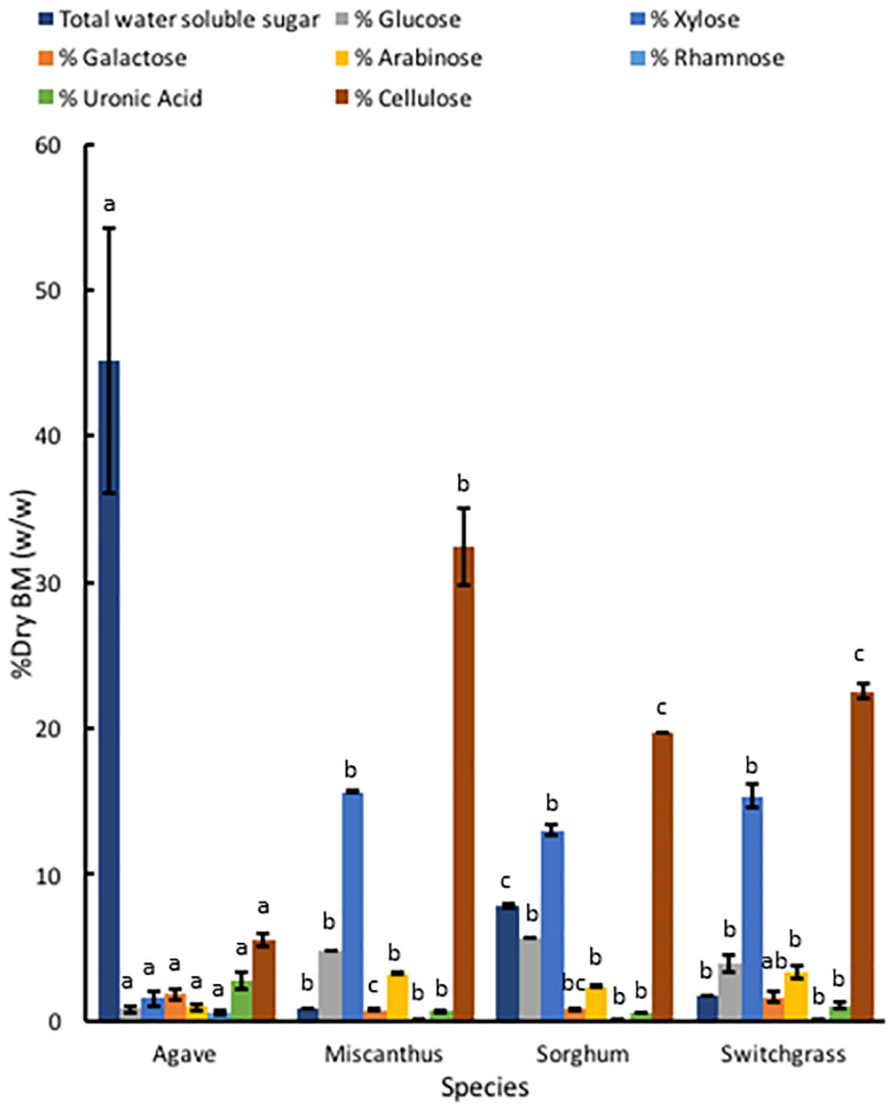


FIGURE 6 | Polysaccharide composition of dry biomass (%) from TFA hydrolysis and phenol-sulfuric acid hydrolysis of EFBM showing mean percent total water-soluble carbohydrates, glucose, xylose, galactose, arabinose, rhamnose, uronic acid, and cellulose for *A. americana*, miscanthus, sorghum, and switchgrass (n = 16 for *A. americana* and n = 3 for grasses). Lowercase letters above bars indicate significant differences between species ($p < 0.05$) from one-way ANOVA and Tukey's HSD test of each soluble sugar.

kg^{-1}; p = 0.0065) compared to whole plant tissue samples of miscanthus, sorghum, and switchgrass (**Figure 7**). Theoretical GH, calculated from equations 2a and 2b, ranged from 17.29 MJ kg^{-1} to 17.63 MJ kg^{-1} for *A. americana*.

Comparison of Enzymatic Digestibility

After the hydrolysis treatment, *A. americana* had the greatest amount of sugar products compared to the other feedstocks (p = 0.0005); Glc and Xyl released from 0.014 g extractive free biomass was found to be 1.45 mg ± 0.21 mg and 0.44 mg ± 0.09 mg for *A. americana* (n = 3; **Figure 8**). Glc and Xyl released from enzymatic hydrolysis of extractive free biomass from *A. americana* were significantly greater compared to the other feedstocks tested (p = 0.00028 and 0.013, respectively), and sorghum had the greatest amount of sugar products compared to the grasses ($p = 7.67 \times 10^{-5}$; **Figure 8**).

Percent hydrolysis was determined for each Glc and Xyl with respect to contents of each monosaccharide within EFBM. *Agave americana* had the greatest mean percentage of total Glc and Xyl liberated by enzymatic hydrolysis (21.28% ± 3.20%). The percentage of monosaccharides in the other feedstocks, in descending order, were 13.16% ± 0.53% (sorghum), 6.19% ± 0.44% (switchgrass), and 3.93% ± 0.47% (miscanthus). Percent Glc after enzymatic hydrolysis of EFBM was statistically similar in samples from switchgrass and miscanthus (p = 0.66), and percent hydrolysis of *A. americana* and sorghum were

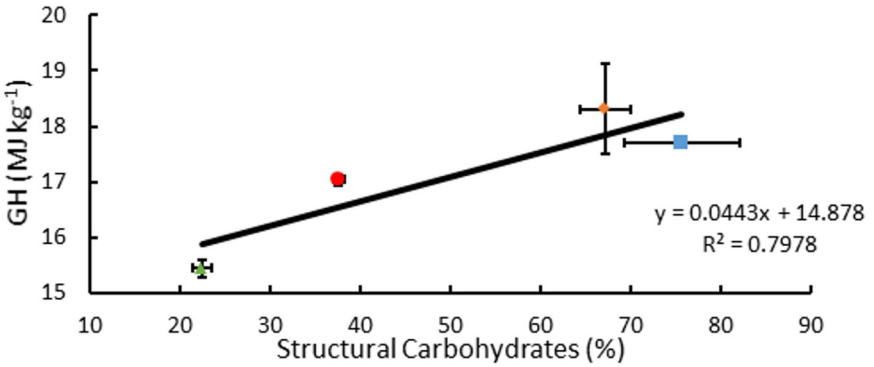


FIGURE 7 | Relationships associated with percentage structural carbohydrates per unit dry biomass (%) and GH for *A. americana* (▲), *Miscanthus* (■), *Sorghum* (●), and switchgrass (◆). Error bars indicate standard error of the mean; $n = 3$ for GH; n for composition is reported in **Supplementary Table 1**.

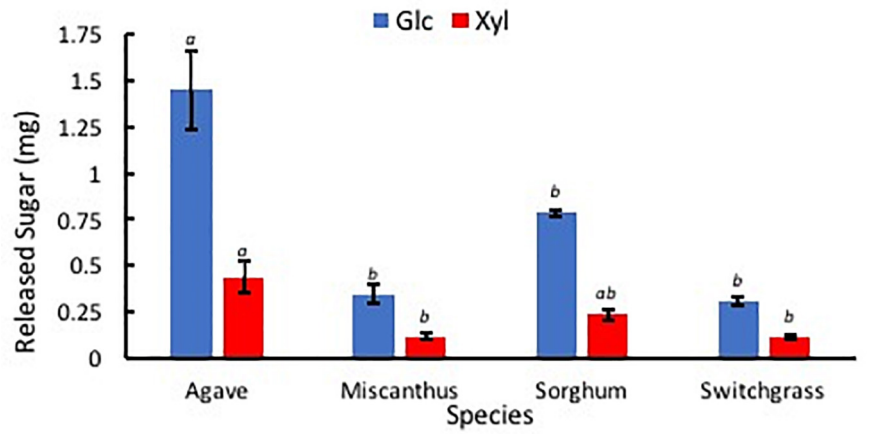


FIGURE 8 | Sugars Glc (blue) and Xyl (red) released from low solids loading enzymatic hydrolysis, expressed in mg, for each of the four species *A. americana* (Agave), miscanthus, sorghum, and switchgrass. Error bars indicate standard error; $n = 3$. Lowercase italic letters indicate significant differences between species by sugar type as per Tukey's HSD test ($p < 0.05$).

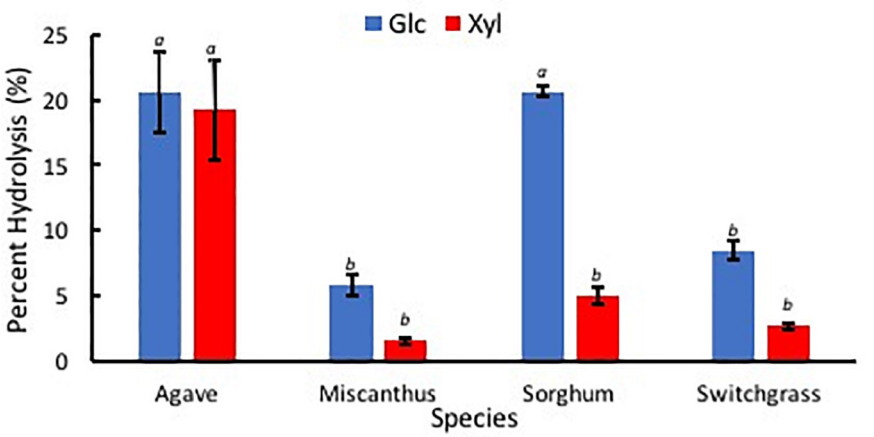


FIGURE 9 | Percent sugars Glc (blue) and Xyl (red) released from low solids loading enzymatic hydrolysis for each of the four species *A. americana* (Agave), miscanthus, sorghum, and switchgrass. Error bars indicate standard error; $n = 3$. Lowercase italic letters indicate significant differences between species by sugar type as per Tukey's HSD test ($p < 0.05$).

significantly greater than that of switchgrass and miscanthus as per ANOVA and Tukey's HSD test (p = 0.0017). Glc monomers from *A. americana* (20.58% ± 3.04%) and sorghum (20.64% ± 0.40%) were similar (p = 0.90), but more Xyl was released from *A. americana* (19.21% ± 3.77%) compared to sorghum (5.03% ± 0.66%) (p = 0.0035; **Figure 9**). Glc and Xyl released from switchgrass (8.49% ± 0.72% and 2.67% ± 0.28%, respectively) and miscanthus (5.84% ± 0.79% and 1.56% ± 0.22%, respectively) were the lowest of the feedstocks tested (n = 3; **Figure 9**). Percent Xyl released from *A. americana* was significantly greater than that released from the other feedstocks as per ANOVA and Tukey's HSD test (p = 0.0025).

DISCUSSION

This study evaluated *A. americana* as a potentially less recalcitrant advanced biofuel feedstock to be cultivated in the southwestern United States by analyzing the chemical composition of *A. americana* plants grown in a field trial. Results here indicate that the high proportion of WSS present in *A. americana* leaf tissue can be more easily extracted and fermented even without pretreatment processes to remove hemicelluloses and lignin. Literature values for biomass yield (Grennell, 2014; Arundale et al., 2015; Davis et al., 2016) were used to calculate fuel yield based on WSS and holocellulose content resolved in this study. The results indicate that *A. americana* crops can produce fuel amounts on arid lands that are comparable to fuel yield from temperate grass feedstocks (**Table 3**).

Overall, annual biomass yields of *A. americana* are lower compared annual biomass yields of temperate grass feedstocks analyzed in this study (**Table 3**), but *A. americana* leaf tissue contains WSS that are readily fermentable without pretreatments. *Agave americana* can therefore yield fuel quantities comparable to the temperate grass species even in conditions with low water input because of the large calculated WUE for this species (**Table 3**). Here, the direct measurements of carbohydrates in *A. americana* indicate biofuel yield from this species would be 6647 L ha^{-1} y^{-1}, and the ratio of fuel yield to water input for each of the four species examined in this study was highest for *A. americana* (12.54 L ha^{-1} mm^{-1}) (**Table 3**).

Composition Analysis

For *A. americana*, experimentally quantified WSS (43.95%), hemicellulose (5.63%), and lignin content (9.06%) were greater than prior literature values, while cellulose (5.55%) and holocellulose content (11.18%) were lower than literature values (Corbin et al., 2015). Miscanthus cellulose content measured in this study was lower than estimates from prior literature (Hodgson et al., 2011; Hayes, 2012; Haffner et al., 2013). Switchgrass had lower water-soluble mono- and disaccharide content in this study relative to prior literature (Samuel et al., 2010 and Yan et al., 2010) that may be due to the short term storage outdoors after harvest, as would be expected in real-world management of the crop. There were comparable percentages of total mono- and disaccharides between miscanthus and switchgrass, while past literature did not report any water-soluble mono- and disaccharides in miscanthus (Samuel et al., 2010; Yan et al., 2010; Hodgson et al., 2011; Hayes, 2012; Haffner et al., 2013; **Figure 4**). Sorghum composition was also different from literature values (Zhao et al., 2009), as tissue originated from a biomass variety (*i.e.,* CHR-FS4).

This study confirms the findings of past literature that indicate there are high concentrations of WSS within the leaves of *Agave* spp. and other CAM plants (Sánchez-Marroquín and Hope, 1953; Mancilla-Margalli and López, 2006; Borland et al., 2009; Bouaziz et al., 2014; Zacarías-Toledo et al., 2016). Prior collection and analysis of pressed juice from *A. americana* showed leaves contained high concentrations of free glucose (12.7 g L^{-1}) (Li et al., 2012). Yet, TFA hydrolysis of the water extract in this study did not yield as much Fru in *A. americana* as would be expected by the large amount of material present in the water extract (**Supplementary Table 1**). Previous analysis of pressed juice from *A. americana* leaves by Li et al. (2012) showed the presence of unidentified water-soluble oligomeric sugars (4.2 g L^{-1}). Past literature has reported high content of uniquely structured fructans (*i.e.*, neofructan) present in CAM species, including *A. americana* (Li et al., 2012; Zacarías-Toledo et al., 2016). These neofructans were identified in past studies of *Agave* spp. and are structurally different from inulin isolated from *Chicorium intybus* L., are heterosubstituted and highly branched, and have been shown to have a wholly unique structure, termed agavin, that is more akin to fructans isolated from *Allium* spp. and *Asparagus officinalis* (Mancilla-Margalli and López, 2006; López and Mancilla-Margalli, 2007; Ravenscroft et al., 2009; Arrizon

TABLE 3 | Biomass yield (dry) and calculated bioethanol (EtOH) yield[a] expressed in L (ha y)$^{-1}$ for *A. americana (Ag)*, miscanthus (Mi), sorghum (So), and switchgrass (Sw) with respect to water input (mm y^{-1}), and water use efficiency (WUE) expressed in kg (ha mm)$^{-1}$ showing fuel yield to water input (Fuel:H_2O) ratio.

Species	Biomass yield (Mg ha^{-1} y^{-1})	EtOH yield[a] (L ha^{-1} y^{-1})	Water input (mm y^{-1})	WUE (kg ha^{-1} mm^{-1})	Fuel:H_2O (L ha^{-1} mm^{-1})
Ag	9.3[1]	6647	530[1]	17.55	12.54
Mi	23.4[2]	8653	1030[2]	22.72	8.40
So	22.0[3]	5998	1019[3]	21.59	5.89
Sw	10.0[2]	3039	1030[2]	10.10	2.95

[1]Davis et al., 2016; [2]Arundale et al. (2015); [3]Grennell (2014). [a]Ethanol (EtOH) yield calculated from biomass yield reported in first column, summed WSS and holocellulose, 0.51 conversion of sugar to EtOH (w w^{-1}), and the density of EtOH (d_{EtOH} = 0.789 kg L^{-1}) as described in Li et al. (2012).

et al., 2010; García-Curbelo et al., 2016). Structural carbohydrates were composed of TFA-soluble hemicelluloses and TFA-insoluble cellulose. The structural fraction of *A. americana* leaf biomass contains mostly cell wall carbohydrates from the residual fiber and cuticle (Li et al., 2012; Pérez-Pimienta et al., 2013). The fiber, which constitutes approximately 25% of dry biomass, is predominately cellulose and contains little lignin compared to other feedstocks of interest (Mylsamy and Rajendran, 2010; Li et al., 2012; Corbin et al., 2015; Yang et al., 2015).

The amount of GalA in *A. americana* tissue indicated that a large proportion of structural pectic oligosaccharide was present within the cell wall (**Figure 6**). Uronic acids are of interest with respect to biofuel production because they act as inhibitors to fermentation (Jönsson and Martín, 2016). The proportion of rhamnose liberated from TFA hydrolysis was also indicative of a pectic saccharides present in the *A. americana* cell wall which are most abundant in the RG I molecule, however, the smaller proportion of rhamnose compared to the other monosaccharides indicates these may be from RG II, the more substituted pectic polysaccharide that contains a galacturonic acid backbone with highly substituted sidechains. Large proportions of arabinose, galactose, and xylose are indicative of RG I and may be from the associated sidechains containing branched arabinan, unbranched galactan, and type I arabinogalactan (Øbro et al., 2004). The large proportion of arabinose and galactose may also be indicative of arabinogalactan proteins, which contain mostly type II arabinogalactan. The amount of glucose and some of the xylose present in *A. americana* leaf tissue can be attributed to XyG present in non-commelinoid monocot species. However, fucose, which is present in the fucogalacto-XyG, was minimally detectable (values not reported). These results are consistent with *A. americana*, from the order Asparagales, having a Type I cell wall that is rich in pectins and contains XyGs and xylans.

The large amounts of TFA-soluble xylose, glucose, and arabinose that were measured in the grasses are consistent with the presence of xylans and mixed linkage glucans found in commelinoid species; these polymers contain xylan and glucan backbones, respectively, and are largely present in commelinoid monocot species, with mixed linkage glucans occurring uniquely in Poales (Buckeridge et al., 2004). Chromatogram traces of the grasses showed minimal differences between them except miscanthus had an unidentified peak *ca.* 21 min and had a lower response for rhamnose compared to galactose (**Supplementary Figure 2**). The proportion of arabinose to xylose is consistent with the presence of arabinoxylan. Uronic acid was present as GalA in all three temperate grass species but was a much lower proportion of total hemicellulosic material compared to *A. americana*, which is consistent with the grasses having a pectin-poor cell wall matrix. It is well-known that grasses have Type II cell walls.

Total Carbohydrates, WUE and Fuel Yield

Due to the greater WSS and greater WUE, the fuel yield from *A. americana* per unit of water input is greater than the fuel yield from the other feedstocks per unit of water input (**Table 3**). In the first field trial examining *A. americana* productivity on arid lands, Davis et al. (2016) found dry biomass yield to be 2.5–9.3 Mg (ha y) $^{-1}$ with maximum productivity and WUE [17.55 kg ha^{-1} mm water (H_2O^{-1})] with 530 mm H_2O input. Yan et al. (2011) showed bioethanol derived from *A. tequilana* has a greater GHG offset compared to maize and switchgrass feedstocks, and *A. tequilana* may produce more ethanol per unit area, compared to maize and switchgrass, when grown under favorable conditions. However, this study is the first to examine how productivity and associated WUE of field grown *A. americana* relates to the concentration of fermentable carbohydrates and final biofuel yield.

Crassulacean Acid Metabolism physiology has a principle advantage over C_3 and C_4 photosynthesis because nocturnal carbon assimilation by the enzyme phosphoenolpyruvate carboxylase (PEPC), activity that is temporally separated from the daytime activity of the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RUBISCO), leads to lessened evapotranspiration and greater WUE (Nobel, 1991, 2003; Kluge and Ting, 2012). The WUE of CAM species has been observed to be 4- to 8-fold higher than C_3 plants (Neales, 1973; Nobel, 1991, 2003; Kluge and Ting, 2012) and CAM species can therefore be more productive than C_3 and C_4 plant species in drought conditions (Ehrler, 1983; Cushman et al., 2015; Davis S.C. et al., 2015; Yang et al., 2015). Many CAM species can avoid hydraulic limitations due to adaptations (*e.g.* drought induced abscission of roots/leaves, succulence, rosette morphology, sunken stomata, decreased stomatal density) that allow them to maintain turgor even in a xeric environment (Woerner and Martin, 1999; Nobel, 2003; Borland et al., 2009). *Agave* spp. can resume physiological function after long periods of drought by quickly achieving comparable transpiration rates when water becomes available (Ehrler, 1983).

Agave spp. also exhibit high photosynthetic productivity across a range of environmental conditions (Neales, 1973; Gentry, 1982; Nobel, 2003; García-Moya et al., 2011). In Australia and the semi-arid southwestern US, field trials have been established to test *Agave* spp. as potential biofuel feedstocks (Chambers et al., 2010; Holtum et al., 2011; Davis et al., 2016). Li et al. (2012) compared theoretical maximum ethanol yields derived from *A. americana*, poplar and switchgrass cropping systems, and projected bioethanol yields of 3645–12390, 4819, and 5311 L ha^{-1} y^{-1}, respectively. Davis et al. (2014) found biofuel yields for *A. fourcroydes*, *A. tequilana* and *A. sisalana* to be 3300, 9700, and 4700 L ha^{-1} y^{-1}.

Energy Content of Biomass

Gross heat of combustion was measured in an oxygen bomb calorimeter and expressed in MJ kg^{-1} to determine total recoverable energy from EFBM for each of the four crop species (*A. americana*, miscanthus, sorghum, and switchgrass). Literature has previously shown GH values to be highly correlated with lignin and extractives values (Demirbaş, 2001), but results from this study contradict that finding in the case of *A. americana*. In the case of sorghum and switchgrass, equations

2a and 2b accurately predicted the range of measured GH value, however in the case of *A. americana* and miscanthus, results were overestimated by 11.30 and 1.74%, respectively, compared to the lower of the calculated theoretical GH values for that species. Notably, *A. americana* had a measured GH value that was less than the y-intercept of both equations 2a and 2b. Similarly low values for GH have been reported in the literature for CAM species like *A. tequilana* and *O. ficus-indica*, with GH of 17.50 and 16.95 MJ kg^{-1}, respectively (Yang et al., 2015).

Composition of biomass as reported in previous literature for species examined within this study contradict the correlation of GH with% lignin, in the case of *A. americana*. From composition data compiled from literature values (Thammasouk et al., 1997; Zhao et al., 2009; Yan et al., 2010; Hodgson et al., 2011; Hayes, 2012; Haffner et al., 2013; Pérez-Pimienta et al., 2013; Corbin et al., 2015), GH was found to be weakly correlated with percent lignin (R^2 = 0.516) but was found to be more correlated with percent structural carbohydrates (R^2 = 0.798) (**Figure 7**). Friedl et al. (2005) propose that GH is best correlated with elemental composition analysis with respect to carbon (C), hydrogen (H), and nitrogen (N) contents. These results indicate that estimating GH using algebraic models that solely rely on lignin may not be appropriate for some species, including *A. americana* where lignin content may not be related to the biomass recalcitrance during enzymatic digestion.

Comparison of Enzymatic Digestibility

It was hypothesized that *A. americana* will be more susceptible to enzymatic hydrolysis and produce a higher-quality hydrolysate with a higher concentration of WSS compared to other candidate feedstocks because of the lower percentage of lignin; lignin and hydrolyzed residues are known to inhibit fermentation and enzymatic hydrolysis (Berlin et al., 2006; Ximenes et al., 2010). The results of this study confirmed that *A. americana* was most susceptible to enzymatic hydrolysis compared to the other feedstocks tested.

CONCLUSION

Despite the lower overall energy content of *A. americana* relative to the advanced cellulosic C_4 crops, potential liquid fuel yields from *A. americana* are greater because fermentable sugars are more easily produced from biomass feedstocks. Total amount of sugars liberated from enzymatic hydrolysis of *A. americana* were almost double that of sorghum, more than triple that of switchgrass, and more than five-fold that of miscanthus. These results suggest that *A. americana* is a less recalcitrant feedstock, and a lower quantity of enzymes are necessary to achieve comparable depolymerization of structural polysaccharides to fermentable monosaccharides than other lignocellulosic feedstocks. Results from comparison of the four feedstocks indicated that *A. americana* is the most easily hydrolyzed relative to sorghum, switchgrass, and miscanthus.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**. Raw datasets are available from authors by request.

AUTHOR CONTRIBUTIONS

AJ, MH, and SD designed the study and research objectives. AJ collected biomass, calculated yield data, and performed statistical analysis. AJ and YZ processed biomass samples. AJ, YZ, SD, and MH conducted laboratory experiments and data analysis and contributed section content and edits to subsequent drafts. AJ and SD drafted the initial manuscript. SD and MH financially supported the project work and manuscript publication.

FUNDING

This work was supported financially by Ohio University, Voinovich School of Leadership and Public Affairs. A grant from the Energy Bioscience Institute with financial assistance from BP and Maricopa Agricultural Center supported the cultivation of *Agave americana*.

ACKNOWLEDGMENTS

The authors thank Misako Hata, Edison Biotechnology Institute at Ohio University, and Dr. Brian McCarthy, Department of Environmental and Plant Biology at Ohio University, for the use of materials, equipment, and laboratory space to complete some of the analyses. Acknowledgment is also given to Dr. Ahmed Faik for input on an early draft of this manuscript. All GH samples were performed at the Center for Electrochemical Engineering Research (CEER) at Ohio University. The authors would also like to thank all the many folks that assisted with planting, maintaining, and harvesting the biomass crops in Athens, Ohio and Maricopa, Arizona.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00654/full#supplementary-material

REFERENCES

Arrizon, J., Morel, S., Gschaedler, A., and Monsan, P. (2010). Comparison of the water-soluble carbohydrate composition and fructan structures of *Agave tequilana* plants of different ages. *Food Chem.* 122, 123–130. doi: 10.1016/j.foodchem.2010.02.028

Arundale, R., Bauer, S., Haffner, F., Mitchell, V., Voigt, T., and Long, P. (2015). Environment Has little effect on biomass biochemical composition of *Miscanthus* × giganteus across soil types, nitrogen fertilization, and times of harvest. *BioEnergy Res.* 8, 1636–1646.

Berlin, A., Balakshin, M., Gilkes, N., Kadla, J., Maximenko, V., Kubo, S., et al. (2006). Inhibition of cellulase, xylanase and β-glucosidase activities by softwood lignin preparations. *J. Biotechnol.* 125, 198–209. doi: 10.1016/j.jbiotec.2006.02.021

Borland, A. M., Griffiths, H., Hartwell, J., and Smith, J. A. C. (2009). Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. *J. Exp. Bot.* 60, 2879–2896. doi: 10.1093/jxb/erp118

Bouaziz, M. A., Rassaoui, R., and Besbes, S. (2014). Chemical composition, functional properties, and effect of inulin from tunisian *Agave americana* L. Leaves on textural qualities of pectin gel. *J. Chem.* 2014, 1–11. doi: 10.1155/2014/758697

Boundy, B., Diegel, S. W., Wright, L., and Davis, S. C. (2011). *Biomass Energy Data Book*, 4th Edn. Oak Ridge, TN: Oak Ridge National Laboratory.

Browning, B. L. (1967). *Methods of Wood Chemistry*, Vol. 1. Hoboken, NJ: John Wiley & Sons, Incorporated.

Buckeridge, M. S., Rayon, C., Urbanowicz, B., Tiné, M. A. S., and Carpita, N. C. (2004). Mixed linkage (1/3),(1/4)-bD-glucans of grasses. *Cereal Chem.* 81, 115–127.

Camesasca, L., Ramírez, M. B., Guigou, M., Ferrari, M. D., and Lareo, C. (2015). Evaluation of dilute acid and alkaline pretreatments, enzymatic hydrolysis and fermentation of napiergrass for fuel ethanol production. *Biomass Bioenergy* 74, 193–201. doi: 10.1016/j.biombioe.2015.01.017

Chambers, D., Holtum, J. A. M., and Rural Industries Research and Development Corporation (Australia). (2010). *Feasibility of Agave as a Feedstock for Biofuel Production in Australia*. Barton, A.C.T.: RIRDC.

Christian, D., Poulton, P., Riche, A., Yates, N., and Todd, A. (2006). The recovery over several seasons of 15N-labelled fertilizer applied to *Miscanthus* × *giganteus* ranging from 1 to 3 years old. *Biomass Bioenergy* 30, 125–133. doi: 10.1016/j.biombioe.2005.11.002

Corbin, K. R., Byrt, C. S., Bauer, S., DeBolt, S., Chambers, D., Holtum, J. A. M., et al. (2015). Prospecting for energy-rich renewable raw materials: agave leaf case study. *PLoS ONE* 10:e0135382. doi: 10.1371/journal.pone.0135382

Culbertson, A., Jin, M., Sousa, L., da, C., Dale, B. E., and Balan, V. (2013). In-house cellulase production from AFEXTM pretreated corn stover using *Trichoderma reesei* RUT C-30. 3, 25960–25969. doi: 10.1039/C3RA44847A

Cushman, J. C., Davis, S. C., Yang, X., and Borland, A. M. (2015). Development and use of bioenergy feedstocks for semi-arid and arid lands. *J. Exp. Bot.* 66, 4177–4193. doi: 10.1093/jxb/erv087

Davis, R., Tao, L., Scarlata, C., Tan, E. C. D., Ross, J., Lukas, J., et al. (2015). *Process Design and Economics for the Conversion of Lignocellulosic Biomass to Hydrocarbons: Dilute-Acid and Enzymatic*. Available at: http://www.nrel.gov/docs/fy15osti/62498.pdf.

Davis, S. C., Ming, R., LeBauer, D. S., and Long, S. P. (2015). Toward systems-level analysis of agricultural production from crassulacean acid metabolism (CAM): scaling from cell to commercial production. *New Phytol.* 208, 66–72. doi: 10.1111/nph.13522

Davis, S. C., Anderson-Teixeira, K. J., and DeLucia, E. H. (2009). Life-cycle analysis and the ecology of biofuels. *Trends Plant Sci.* 14, 140–146. doi: 10.1016/j.tplants.2008.12.006

Davis, S. C., Boddey, R. M., Alves, B. J. R., Cowie, A. L., George, B. H., Ogle, S. M., et al. (2013). Management swing potential for bioenergy crops. *GCB Bioenergy* 5, 623–638. doi: 10.1111/gcbb.12042

Davis, S. C., Dohleman, F. G., and Long, S. P. (2011). The global potential for Agave as a biofuel feedstock. *GCB Bioenergy* 3, 68–78. doi: 10.1111/j.1757-1707.2010.01077.x

Davis, S. C., Kuzmick, E. R., Niechayev, N., and Hunsaker, D. J. (2016). Productivity and water use efficiency of Agave americana in the first field trial as bioenergy feedstock on arid lands. *GCB Bioenergy* 9, 314–325. doi: 10.1111/gcbb.12324

Davis, S. C., LeBauer, D. S., and Long, S. P. (2014). Light to liquid fuel: theoretical and realized energy conversion efficiency of plants using crassulacean acid metabolism (CAM) in arid conditions. *J. Exp. Bot.* 65, 3471–3478. doi: 10.1093/jxb/eru163

Davis, S. C., Parton, W. J., Dohleman, F. G., Smith, C. M., Grosso, S. D., Kent, A. D., et al. (2010). Comparative biogeochemical cycles of bioenergy crops reveal nitrogen-fixation and low greenhouse gas emissions in a *Miscanthus* × *giganteus* agro-ecosystem. *Ecosystems* 13, 144–156.

Davis, S. C., Simpson, J., Gil-Vega, K. D. C., Niechayev, N. A., Tongerlo, E. V., Castano, N. H., et al. (2019). Undervalued potential of crassulacean acid metabolism (CAM) for current and future agricultural production. *J. Exp. Bot.* 70, 6521–6537. doi: 10.1093/jxb/erz223

Demirbaş, A. (2001). Relationships between lignin contents and heating values of biomass. *Energy Conv. Manage.* 42, 183–188. doi: 10.1016/S0196-8904(00)00050-9

DOE, (2006). *Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda (DOE/SC-0095)*. Rockville, MD: U.S. Department of Energy.

DuBois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. T., and Smith, F. (1956). Colorimetric method for determination of sugars and related substances – Analytical chemistry (ACS Publications). *Anal. Chem.* 28, 350–356.

Ehrler, W. L. (1983). The transpiration ratios of *Agave americana* L. and *Zea mays* L. as affected by soil water potential. *J. Arid Environ.* 6, 107–113.

Epa (2013). Regulation of fuels and fuel additives: 2013 renewable fuel standards; final rule. *Federal Register* 78, 49794–49830.

Epa. (2015). Renewable fuel standard program: standards for 2014 2015, and 2016 and biomass-based diesel volume for 2017; final rule. *Federal Register* 80, 77420–77518.

Evans, C. S., Dutton, M. V., Guillén, F., and Veness, R. G. (1994). Enzymes and small molecular mass agents involved with lignocellulose degradation. *FEMS Microbiol. Rev.* 13, 235–239.

Faik, A. (2013). "Plant cell wall structure-pretreatment" the critical relationship in biomass conversion to fermentable sugars," in *Green Biomass Pretreatment for Biofuels Production*, ed. T. Gu, (Dordrecht: Springer), 1–30.

Friedl, A., Padouvas, E., Rotter, H., and Varmuza, K. (2005). Prediction of heating values of biomass fuel from elemental composition. *Anal. Chim. Acta* 544, 191–198. doi: 10.1016/j.aca.2005.01.041

Gao, D., Chundawat, S. P. S., Krishnan, C., Balan, V., and Dale, B. E. (2010). Mixture optimization of six core glycosyl hydrolases for maximizing saccharification of ammonia fiber expansion (AFEX) pretreated corn stover. *Bioresour. Technol.* 101, 2770–2781. doi: 10.1016/j.biortech.2009.10.056

García-Curbelo, Y., López, M. G., Bocourt, R., Collado, E., Albelo, N., and Nuñez, O. (2016). Structural characterization of fructans from Agave fourcroydes (Lem.) with potential as prebiotic. *Cuban J. Agric. Sci.* 49, 75–80.

García-Moya, E., Romero-Manzanares, A., and Nobel, P. S. (2011). Highlights for agave productivity: agave productivity. *GCB Bioenergy* 3, 4–14. doi: 10.1111/j.1757-1707.2010.01078.x

Gentry, H. S. (1982). *Agaves of Continental North America*. Tucson, AZ: University of Arizona Press.

Godin, B., Lamaudière, S., Agneessens, R., Schmit, T., Goffart, J.-P., Stilmant, D., et al. (2013). Chemical characteristics and biofuel potential of several vegetal biomasses grown under a wide range of environmental conditions. *Ind. Crops Prod.* 48, 1–12. doi: 10.1016/j.indcrop.2013.04.007

Grabber, J. H., Ralph, J., and Hatfield, R. D. (2000). Cross-Linking of maize walls by ferulate dimerization and incorporation into lignin. *J. Agric. Food Chem.* 48, 6106–6113. doi: 10.1021/jf0006978

Grennell, J. (2014). *Yield and Carbon Exchange of Sorghum Grown as Advanced Biofuel Feedstock on Abandoned Agricultural Land in Southeastern Ohio*. Athens, OH: Ohio University.

Haffner, F. B., Mitchell, V. D., Arundale, R. A., and Bauer, S. (2013). Compositional analysis of Miscanthus giganteus by near infrared spectroscopy. *Cellulose* 20, 1629–1637. doi: 10.1007/s10570-013-9935-1

Hames, B., Ruiz, R., Scarlata, C., Sluiter, A., Sluiter, J., and Templeton, D. (2008a). *Preparation of Samples for Compositional Analysis (Technical Report No.*

NREL/TP-510-42620). Golden, CO: National Renewable Energy Laboratory, 1–9.

Hames, B., Scarlata, C., and Sluiter, A. (2008b). *Determination of Protein Content in Biomass (Technical Report No. NREL/TP-510-42625)*. Golden, CO: National Renewable Energy Laboratory, 1–8.

Hayes, D. J. M. (2012). Development of near infrared spectroscopy models for the quantitative prediction of the lignocellulosic components of wet *Miscanthus* samples. *Bioresour. Technol.* 119, 393–405. doi: 10.1016/j.biortech.2012.05.137

Heaton, E. A., Dohleman, F. G., and Long, S. P. (2009). Seasonal nitrogen dynamics of *Miscanthus* × *giganteus* and *Panicum virgatum*. *GCB Bioenergy* 1, 297–307.

Held, M. A., Be, E., Zemelis, S., Withers, S., Wilkerson, C., and Brandizzi, F. (2011). CGR3: a Golgi-localized protein influencing homogalacturonan methylesterification. *Mol. Plant* 4, 832–844. doi: 10.1093/mp/ssr012

Hess, J. R., Wright, C. T., and Kenney, K. L. (2007). Cellulosic biomass feedstocks and logistics for ethanol production. *Biofuels Bioprod. Biorefin.* 1, 181–190. doi: 10.1016/j.biotechadv.2019.03.002

Hodgson, E. M., Nowakowski, D. J., Shield, I., Riche, A., Bridgwater, A. V., Clifton-Brown, J. C., et al. (2011). Variation in Miscanthus chemical composition and implications for conversion by pyrolysis and thermo-chemical bio-refining for fuels and chemicals. *Bioresour. Technol.* 102, 3411–3418. doi: 10.1016/j.biortech.2010.10.017

Holtum, J. A. M., Chambers, D., Morgan, T., and Tan, D. K. Y. (2011). Agave as a biofuel feedstock in Australia: agave in australia. *GCB Bioenergy* 3, 58–67. doi: 10.1111/j.1757-1707.2010.01083.x

Humbird, D., Davis, R., Tao, L., Kinchin, C., Hsu, D., Aden, A., et al. (2011). *Process Design and Economics for Biochemical Conversion of Lignocellulosic Biomass to Ethanol: Dilute-Acid Pretreatment and Enzymatic Hydrolysis of Corn Stover*. Golden, CO: National Renewable Energy Lab. (NREL).

Jin, M., da Costa Sousa, L., Schwartz, C., He, Y., Sarks, C., et al. (2016). Toward lower cost cellulosic biofuel production using ammonia based pretreatment technologies. *Green Chem.* 18, 957–966. doi: 10.1039/C5GC02433A

Johnson, E. (2016). Integrated enzyme production lowers the cost of cellulosic ethanol. *Biofuels Bioprod. Biorefin.* 10:51.

Jönsson, L. J., Alriksson, B., and Nilvebrant, N.-O. (2013). Bioconversion of lignocellulose: inhibitors and detoxification. *Biotechnol. Biofuels* 6:16. doi: 10.1186/1754-6834-6-16

Jönsson, L. J., and Martín, C. (2016). Pretreatment of lignocellulose: formation of inhibitory by-products and strategies for minimizing their effects. *Bioresour. Technol.* 199, 103–112. doi: 10.1016/j.biortech.2015.10.009

Kim, Y., Ximenes, E., Mosier, N. S., and Ladisch, M. R. (2011). Soluble inhibitors/deactivators of cellulase enzymes from lignocellulosic biomass. *Enzyme Microb. Technol.* 48, 408–415. doi: 10.1016/j.enzmictec.2011.01.007

Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B. A., and Blanch, H. W. (2012). The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnol. Bioeng.* 109, 1083–1087. doi: 10.1002/bit.24370

Kluge, M., and Ting, I. P. (2012). *Crassulacean Acid Metabolism: Analysis of an Ecological Adaptation*. Berlin: Springer Science & Business Media.

Lam, T. B. T., Kadoya, K., and Iiyama, K. (2001). Bonding of hydroxycinnamic acids to lignin: ferulic and p-coumaric acids are predominantly linked at the benzyl position of lignin, not the β-position, in grass cell walls. *Phytochemistry* 57, 987–992. doi: 10.1016/s0031-9422(01)00052-8

Latshaw, W. L., and Miller, E. C. (1924). Elemental composition of the corn plant. *J. Agric. Res.* 27, 845–860.

Li, H., Foston, M. B., Kumar, R., Samuel, R., Gao, X., Hu, F., et al. (2012). Chemical composition and characterization of cellulose for Agave as a fast-growing, drought-tolerant biofuels feedstock. *RSC Adv.* 2:4951. doi: 10.1039/c2ra20557b

Li, H., Pattathil, S., Foston, M. B., Ding, S.-Y., Kumar, R., Gao, X., et al. (2014). Agave proves to be a low recalcitrant lignocellulosic feedstock for biofuels production on semi-arid lands. *Biotechnol. Biofuels* 7:1. doi: 10.1186/1754-6834-7-50

Lin, S. Y., and Dence, C. W. (1992). "Ultraviolet spectroscopy," in *Methods in Lignin Chemistry*, eds S. Y. Lin, and C. W. Dence, (Berlin: Springer), 217–232.

López, M. G., and Mancilla-Margalli, N. A. (2007). "The nature of fructooligosaccharides in agave plants," in *Recent Advances in Fructooligosaccharides Research*, eds S. Norio, B. Noureddine, and O. Shuichi, (Trivandrum: Research Signpost).

Mancilla-Margalli, N. A., and López, M. G. (2006). Water-soluble carbohydrates and fructan structure patterns from *Agave* and *Dasylirion* species. *J. Agric. Food Chem.* 54, 7832–7839. doi: 10.1021/jf060354v

Markwalder, H. U., and Neukom, H. (1976). Diferulic acid as a possible crosslink in hemicelluloses from wheat germ. *Phytochemistry* 15, 836–837.

Mielenz, J. R., Rodriguez, M., Thompson, O. A., Yang, X., and Yin, H. (2015). Development of Agave as a dedicated biomass source: production of biofuels from whole plants. *Biotechnol. Biofuels* 8:79. doi: 10.1186/s13068-015-0261-8

Moorhead, D. L., and Reynolds, J. F. (1993). Changing carbon chemistry of buried creosote bush litter during decomposition in the northern chihuahuan desert. *Am. Midland Natur.* 130:83. doi: 10.2307/2426277

Müse, G., Schindler, T., Bergfeld, R., Ruel, K., Jacquet, G., Lapierre, C., et al. (1997). Structure and distribution of lignin in primary and secondary cell walls of maize coleoptiles analyzed by chemical and immunological probes. *Planta* 201, 146–159.

Mylsamy, K., and Rajendran, I. (2010). Investigation on physio-chemical and mechanical properties of raw and alkali-treated *Agave americana* Fiber. *J. Reinforced Plast. Compos.* 29, 2925–2935. doi: 10.1177/0731684410362817

Neales, T. F. (1973). The effect of night temperature on CO2 assimilation, transpiration, and water use efficiency in *Agave americana* L. *Austr. J. Biol. Sci.* 26, 705–714.

Niechayev, N. A., Jones, A. M., Rosenthal, D. M., and Davis, S. C. (2018). A model of environmental limitations on production of *Agave americana* L. grown as a biofuel crop in semi-arid regions. *J. Exp. Bot.* 70, 6549–6559. doi: 10.1093/jxb/ery383

Nobel, P. S. (1990). Environmental influences on CO2 uptake by agaves, CAM plants with high productivities. *Econ. Bot.* 44, 488–502.

Nobel, P. S. (1991). Achievable productivities of certain CAM plants: basis for high values compared with C3 and C4 plants. *New Phytol.* 119, 183–205. doi: 10.1111/j.1469-8137.1991.tb01022.x

Nobel, P. S. (2003). *Environmental Biology of Agaves and Cacti*. Cambridge: Cambridge University Press.

Øbro, J., Harholt, J., Scheller, H. V., and Orfila, C. (2004). Rhamnogalacturonan I in *Solanum tuberosum* tubers contains complex arabinogalactan structures. *Phytochemistry* 65, 1429–1438. doi: 10.1016/j.phytochem.2004.05.002

Pérez-Pimienta, J. A., Lopez-Ortega, M. G., Varanasi, P., Stavila, V., Cheng, G., Singh, S., et al. (2013). Comparison of the impact of ionic liquid pretreatment on recalcitrance of agave bagasse and switchgrass. *Bioresour. Technol.* 127, 18–24. doi: 10.1016/j.biortech.2012.09.124

Pérez-Pimienta, J. A., Mojica-Álvarez, R. M., Sánchez-Herrera, L. M., Mittal, A., and Sykes, R. W. (2018). Recalcitrance assessment of the agro-industrial residues from five agave species: ionic liquid pretreatment, saccharification and structural characterization. *BioEnergy Res.* 11, 551–561.

Raghuwanshi, S., Deswal, D., Karp, M., and Kuhad, R. C. (2014). Bioprocessing of enhanced cellulase production from a mutant of *Trichoderma asperellum* RCK2011 and its application in hydrolysis of cellulose. *Fuel* 124, 183–189. doi: 10.1016/j.fuel.2014.01.107

Rani, V., Mohanram, S., Tiwari, R., Nain, L., and Arora, A. (2014). Beta-glucosidase: key enzyme in determining efficiency of cellulase and biomass hydrolysis. *J. Bioprocess. Biotech.* 5. doi: 10.4172/2155-9821.1000197

Ravenscroft, N., Cescutti, P., Hearshaw, M. A., Ramsout, R., Rizzo, R., and Timme, E. M. (2009). Structural analysis of fructans from agave americana grown in south africa for spirit production. *J. Agric. Food Chem.* 57, 3995–4003. doi: 10.1021/jf8039389

Resch, M. G., Baker, J. O., and Decker, S. R. (2015). *Low Solids Enzymatic Saccharification of Lignocellulosic Biomass (Technical Report No. NREL/TP-5100-63351)*. Golden, CO: National Renewable Energy Laboratory.

Roche, C. M., Dibble, C. J., Knutsen, J. S., Stickel, J. J., and Liberatore, M. W. (2009). Particle concentration and yield stress of biomass slurries during enzymatic hydrolysis at high-solids loadings. *Biotechnol. Bioeng.* 104, 290–300. doi: 10.1002/bit.22381

Samuel, R., Pu, Y., Foston, M., and Ragauskas, A. J. (2010). Solid-state NMR characterization of switchgrass cellulose after dilute acid pretreatment. *Biofuels* 1, 85–90. doi: 10.4155/bfs.09.17

Sánchez-Marroquín, A., and Hope, P. H. (1953). Agave juice, fermentation and chemical composition studies of some species. *J. Agric. Food Chem.* 1, 246–249.

Sartori, F., Lal, R., Ebinger, M. H., and Parrish, D. J. (2006). Potential soil carbon sequestration and CO 2 offset by dedicated energy crops in the USA. *Crit. Rev. Plant Sci.* 25, 441–472.

Saucedo-Luna, J., Castro-Montoya, A. J., Martinez-Pacheco, M. M., Sosa-Aguirre, C. R., and Campos-Garcia, J. (2011). Efficient chemical and enzymatic saccharification of the lignocellulosic residue from Agave tequilana bagasse to produce ethanol by *Pichia caribbica*. *J. Ind. Microbiol. Biotechnol.* 38, 725–732. doi: 10.1007/s10295-010-0853-z

Sheng, C., and Azevedo, J. L. T. (2005). Estimating the higher heating value of biomass fuels from basic analysis data. *Biomass Bioenergy* 28, 499–507. doi: 10.1016/j.biombioe.2004.11.008

Sluiter, A., and Hames, B. (2012). *Determination of Structural Carbohydrates and Lignin in Biomass (Technical Report No. NREL/TP-510-42618)*. Golden, CO: National Renewable Energy Laboratory, 1–18.

Sluiter, A., Hames, B., Hyman, D., Payne, C., Ruiz, R., Scarlata, C., et al. (2008a). *Determination of Total Solids in Biomass and Total Dissolved Solids in Liquid Process Samples (Technical Report No. NREL/TP-510-42621)*. Golden, CO: National Renewable Energy Laboratory, 1–9.

Sluiter, A., Hames, B., Ruiz, R., Scarlata, C., Sluiter, J., and Templeton, D. (2008b). *Determination of Ash in Biomass (Technical Report No. NREL/TP-510-42622)*. Golden, CO: National Renewable Energy Laboratory, 1–8.

Sluiter, A., Ruiz, R., Scarlata, C., Sluiter, J., and Templeton, D. (2008c). *Determination of Extractives in biomass (Technical Report No. NREL/TP-510-42619)*. Golden, CO: National Renewable Energy Laboratory, 1–12.

Sluiter, J., and Sluiter, A. (2011). *Summative mass closure (Technical Report No. NREL/TP-510-48087)*. Golden, CO: National Renewable Energy Laboratory, 1–13.

Smith, A. M. (2008). Prospects for increasing starch and sucrose yields for bioethanol production. *Plant J.* 54, 546–558. doi: 10.1111/j.1365-313X.2008.03468.x

Somerville, C., Youngs, H., Taylor, C., Davis, S. C., and Long, S. P. (2010). Feedstocks for lignocellulosic biofuels. *Science* 329, 790–792. doi: 10.1126/science.1189268

Stewart, J. R. (2015). Agave as a model CAM crop system for a warming and drying world. *Front. Plant Sci.* 6:684. doi: 10.3389/fpls.2015.00684

Stock, J. H. (2015). *The Renewable Fuel Standard: A Path Forward*. Columbia: SIPA Center on Global Energy Policy.

Thammasouk, K., Tjo, D., and Penner, M. H. (1997). Influence of extractives on the analysis of herbaceous biomass. *J. Agric. Food Chem.* 45, 437–443. doi: 10.1021/JF960401R

Wang, J., Chio, C., Chen, X., Su, E., Cao, F., Jin, Y., et al. (2019). Efficient saccharification of agave biomass using *Aspergillus niger* produced low-cost enzyme cocktail with hyperactive pectinase activity. *Bioresour. Technol.* 272, 26–33. doi: 10.1016/j.biortech.2018.09.069

Weishaar, J. L., Aiken, G. R., Bergamaschi, B. A., Fram, M. S., Fujii, R., and Mopper, K. (2003). Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of dissolved organic carbon. *Environ. Sci. Technol.* 37, 4702–4708. doi: 10.1021/es030360x

Woerner, A. C., and Martin, C. E. (1999). Mechanistic basis of differences in water-use efficiency between a CAM and a C3 species of *Peperomia* (Piperaceae). *New Phytol.* 144, 307–312. doi: 10.1046/j.1469-8137.1999.00525.x

Ximenes, E., Kim, Y., Mosier, N., Dien, B., and Ladisch, M. (2010). Inhibition of cellulases by phenols. *Enzyme Microb. Technol.* 46, 170–176. doi: 10.1016/j.enzmictec.2009.11.001

Yan, J., Hu, Z., Pu, Y., Charles Brummer, E., and Ragauskas, A. J. (2010). Chemical compositions of four switchgrass populations. *Biomass Bioenergy* 34, 48–53. doi: 10.1016/j.biombioe.2009.09.010

Yan, X., Tan, D. K. Y., Inderwildi, O. R., Smith, J. A. C., and King, D. A. (2011). Life cycle energy and greenhouse gas analysis for agave-derived bioethanol. *Energy Environ. Sci.* 4:3110. doi: 10.1039/c1ee01107c

Yang, L., Lu, M., Carl, S., Mayer, J. A., Cushman, J. C., Tian, E., et al. (2015). Biomass characterization of *Agave* and *Opuntia* as potential biofuel feedstocks. *Biomass Bioenergy* 76, 43–53. doi: 10.1016/j.biombioe.2015.03.004

York, W. S., Darvill, A. G., McNeil, M., Stevenson, T. T., and Albersheim, P. (1986). "Isolation and caracterization of plant cell walls and cell wall components," in *Methods in Enzymology*, Vol. 118, eds J. Abelson, and M. Simon, (Cambridge: Academic Press).

Zacarías-Toledo, R., González-Mendoza, D., Mendiola, M. A. R., Villalobos-Maldonado, J. J., Gutiérrez-Oliva, V. F., Dendooven, L., et al. (2016). Plant Growth and Sugars Content of *Agave americana* L. Cultivated with Vermicompost and Rock Phosphate and Inoculated with *Penicillium* sp. and *Glomus fasciculatum*. *Compost Sci. Util.* 24, 259–265. doi: 10.1080/1065657X.2016.1155512

Zhang, Y. H. P., Himmel, M. E., and Mielenz, J. R. (2006). Outlook for cellulase improvement: screening and selection strategies. *Biotechnol. Adv.* 24, 452–481. doi: 10.1016/j.biotechadv.2006.03.003

Zhao, Y. L., Dolat, A., Steinberger, Y., Wang, X., Osman, A., and Xie, G. H. (2009). Biomass yield and changes in chemical composition of sweet sorghum cultivars grown for biofuel. *Field Crops Res.* 111, 55–64. doi: 10.1016/j.fcr.2008.10.006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 17 June 2020
doi: 10.3389/fpls.2020.00817

Morphological and Genetic Variation in Monocultures, Forestry Systems and Wild Populations of *Agave maximiliana* of Western Mexico: Implications for Its Conservation

Dánae Cabrera-Toledo[1], Ofelia Vargas-Ponce[1], Sabina Ascencio-Ramírez[2], Luis Mario Valadez-Sandoval[1], Jessica Pérez-Alquicira[1,3], Judith Morales-Saavedra[1] and Oassis F. Huerta-Galván[4]*

[1] Laboratorio Nacional de Identificación y Caracterización Vegetal (LaniVeg-CONACYT), Departamento de Botánica y Zoología, Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Mexico, [2] Maestría en Ciencias en Recursos Naturales y Desarrollo Rural, Colegio de la Frontera Sur, Tapachula, Mexico, [3] Cátedras CONACYT-Universidad de Guadalajara, Laboratorio Nacional de Identificación y Caracterización Vegetal (LaniVeg), Departamento de Botánica y Zoología, Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Mexico, [4] Maestría en Biosistemática y Manejo de Recursos Forestales y Agrícolas, Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Mexico

OPEN ACCESS

Edited by:
Luis Enrique Eguiarte, National Autonomous University of Mexico, Mexico

Reviewed by:
Angelica Cibrian-Jaramillo, Instituto Politécnico Nacional de México (CINVESTAV), Mexico
Alejandro Casas, National Autonomous University of Mexico, Mexico
Antonio Gonzalez-Rodriguez, National Autonomous University of Mexico, Mexico

***Correspondence:**
Dánae Cabrera-Toledo
danaetoledo@gmail.com

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *14 December 2019*
Accepted: *20 May 2020*
Published: *17 June 2020*

Citation:
Cabrera-Toledo D, Vargas-Ponce O, Ascencio-Ramírez S, Valadez-Sandoval LM, Pérez-Alquicira J, Morales-Saavedra J and Huerta-Galván OF (2020) Morphological and Genetic Variation in Monocultures, Forestry Systems and Wild Populations of Agave maximiliana of Western Mexico: Implications for Its Conservation. Front. Plant Sci. 11:817. doi: 10.3389/fpls.2020.00817

Forestry systems in Mexico are examples of traditional management of land and biodiversity that integrates the use, conservation and restoration of forest elements. Current *in situ* management practices of *Agave maximiliana* in western Mexico include the tolerance of many forest elements, reintroduction of young *Agave* plants and germination of seeds. More intense forms of management include monocultures, which are agroindustrialized systems developed in more recent times and characterized by the establishment of high densities of *A. maximiliana* plants in deforested areas and abandoned agricultural lands. We compared monocultures, forestry systems and wild populations (i.e., non/slightly-exploited forests) in order to evaluate whether these practices have had an effect on intraspecific morphological and genetic variation and divergence. We also tested whether divergence has a positive relationship with environmental and geographic distance. We analyzed 16 phenotypic traits in 17 populations of *A. maximiliana*, and 14 populations were further examined by amplifying 9 SSR loci. We employed multivariate methods and analyses of variance in phenotypic and genetic traits to test whether clusters and the percentage of variation contained in the managed and wild categories can be identified. Tests of isolation by environment (IBE) and distance (IBD) were performed to detect the magnitude of divergence explained by climatic and geographic variables. We found that forestry systems are effective as reservoirs of morphological and genetic diversity, since they maintain levels similar to those of wild populations. Moreover, the monocultures showed similar levels, reflecting their recent emergence. While the species showed high morphological diversity (IMD = 0.638, SE ± 0.07), it had low to intermediate genetic diversity (A = 2.37, H_E = 0.418). Similar morphological and genetic divergences were found among populations, but these were not correlated with each other in population pairs.

Non-significant morphological differentiation was found among categories. Only IBE was significant in the genetic structure ($\beta = 0.32$, $p = 0.007$), while neither IBE nor IBD was detected in the morphological differentiation. We discuss the implications of these results in the context of the weaknesses and strengths of *A. maximiliana* in the face of the socio-ecological changes predicted for the study area in the short term.

Keywords: forestry systems, population genetics, morphological diversity, distilled beverages, raicilla

INTRODUCTION

Agaves have considerable cultural, economic and ecological importance in North America (Gentry, 1982). Around fifty three species of *Agave* L. are used for mescal production (Colunga-GarcíaMarín and Zizumbo-Villarreal, 2007; Torres et al., 2015b); however, most of the plants utilized are not cultivated but rather extracted from natural populations by the simple gathering of plants (*ca*. 37 spp., Torres et al., 2015b). Of these, 53 species used for mescal production, around 12 species of *Agave* are managed *in situ* (Torres et al., 2015b), i.e., a range of practices other than simple gathering is conducted. These populations managed *in situ* constitute valuable Forestry (FS) and Agroforestry (AFS) Systems maintained in Mexico. The terms FS and AFS refer to a land utilization type with several criteria of classification (Nair, 1985). In Mexico, these criteria are based on a traditional management of land and biodiversity that integrates the use, conservation and restoration of multiple useful species of domesticated and wild plants and animals for different purposes (Moreno-Calles et al., 2013). While AFS represent alternatives to the current problems relating to the management and conservation of biodiversity (Torres-García et al., 2019), *Agave* monocultures have had an impact on traditional ecological knowledge and agrobiodiversity, while also depending heavily on the use of toxic agrochemicals (Bowen and Gerritsen, 2007). These agroindustrial management forms are characterized by the establishment of high densities of a single species for one simple purpose in deforested areas and/or abandoned agricultural lands (Torres-García et al., 2019). Less than ten *Agave* species are maintained in monocultures (Torres et al., 2015b). Monocultures of *Agave tequilana* F.A.C. Weber Azul, used in the spirits industry for the production of tequila, have spread rapidly in western Mexico, particularly from the end of the 20th century up to the present day (Leclert et al., 2010).

Analyses based on genetic and phenotypic markers have revealed that differences between wild and managed plant populations depend on life history traits, geographic distribution and cultivation history (Lindsay et al., 2018), as well as on the forms and intensities of management. These constitute an extraordinary broad spectrum of expressions of management types (Casas et al., 1999, 2017; Otero-Arnaiz et al., 2005; Illsey et al., 2007; Blancas et al., 2009, 2013; Parra et al., 2010, 2012; Guillén et al., 2011; Cruse-Sanders et al., 2013; Contreras-Negrete et al., 2015; Torres et al., 2015a). On one hand, agave species with a dominance of vegetative reproduction, form the primary gene pools from which a high diversity of traditional landraces have been artificially selected (*A. angustifolia* Haw. and *A. rhodacantha* Trel., Vargas-Ponce et al., 2007, 2009; Rivera-Lugo et al., 2018; Trejo et al., 2018). Vegetative reproduction can also become dominant in managed populations, although it is not a common reproductive system in wild populations (*A. parryi* Engelm., Parker et al., 2010) or in their wild relatives (*A. hookeri* Jacobi, Figueredo-Urbina et al., 2017). This can, in turn, produce faster morphological divergences between wild and managed populations (Parker et al., 2010, 2014; Figueredo et al., 2014; Figueredo-Urbina et al., 2017). The constant reintroduction of wild germplasm to traditional landraces and high exchange of germplasm between farmers can maintain high levels of genetic diversity (*A. hookeri*, Figueredo-Urbina et al., 2017; *A. angustifolia*, *A. rhodacantha*, Vargas-Ponce et al., 2009), compared to the wild populations.

On the other hand, species that present sexual reproduction, found in wild populations or managed in FS and monocultures, have been studied to a lesser extent. While morphological divergences can be detected between wild and managed species (*A. inaequidens* K. Koch, *A. cupreata* Trel. & A. Berger, Figueredo et al., 2014; Figueredo-Urbina et al., 2017), the genetic divergences explained by management have been reported as null (*A. inaequidens*, *A. cupreata*, Figueredo-Urbina et al., 2017) or incipient (*A. potatorum* Zucc., Félix-Valdez et al., 2016). This reflects the fact that the human activities practiced in these systems are still governed by natural gene flow.

Wild and domesticated *Agave* spp. used for the production of spirits have paniculate inflorescences, a frequent nocturnal anther dehiscence and, in some cases, occur in geographic areas that overlap with those of the long-nosed bats (Arita, 1991). These aspects suggest the occurrence of chiropterophily (Molina-Freaner and Eguiarte, 2003; Rocha et al., 2006; Estrella-Ruiz, 2008; León, 2013). Indeed, their strong relationship with the bats could have driven the radiation of the genera and their close linages (Flores-Abreu et al., 2019). These bats present a foraging behavior over a wide geographical range (e.g., *Leptonycteris curasoae* Miller, 1900, *L. yerbabuenae*, Martinez and Villa, 1940, of distances of up to 100 km, Horner et al., 1998; Medellín et al., 2018) which is congruent with the low genetic differentiation presented by some *Agave* species used for the production of spirits, and presenting sexual reproduction, (e.g., *A. inaequidens*, *A. palmeri* Engelm., *A. cupreata*, **Table 1**). Since this dynamic maintains highly diverse gene pools, monocultures of sexually reproduced plants lose genetic diversity at a slower rate (see *A. inaquidens*, *A. cupreata*, wild vs. monocultures, **Table 1**) compared to those of vegetative propagation (see *A. angustifolia* vs. *A. tequilana*; *A. inaequidens* vs. *A. hookeri*, **Table 1**).

TABLE 1 | Genetic diversity (SSRs) and population differentiation parameters, pollination vectors and reproductive system of *Agave* species.

Species	Management status	H_E	A	F_{ST}; Φ_{ST}	R.S.	Poll.	VP	References
***Mezcalero Agave* species**								
Agave parryi	Wild Antropogenic	0.62 0.43	7.09 2.57	0.34	–	Bats, passerines	Yes	Parker et al., 2010
A. parryi	Wild Cultivated	0.73 0.62	8.59 4.36	–	–	Bats, passerines	–	Lindsay et al., 2018
A. palmeri	Wild	0.67	7.9	0.013	–	Bats, passerines	–	Lindsay et al., 2018
A. inaequidens	Wild Managed Cultivated	0.70 0.73 0.69	7.6 7.7 7.1	0.11	SI	Bats	No	Figueredo et al., 2015
A. cupreata	Wild Cultivated	0.51 0.48	4 3.2	0.07	–	Bats, bees	No	Figueredo-Urbina et al., 2017
A. potatorum	Wild Managed Nursery	0.87 0.72 0.69	–	0.36	SI	Bats	No	Félix-Valdez et al., 2016
A. angustifolia	Wild Cultivated	0.45 0.59	30 8	0.16 –	– –	Bats –	Yes	Trejo et al., 2018
A. tequilana "Azul"	Cultivated	0	–	–	–	–	Yes	Trejo et al., 2018
A. tequilana "Sigüin"	Cultivated	0.49	–	–	–	–	Yes	Trejo et al., 2018
A. tequilana "Chato"		0.43	–	–	–	–	Yes	Trejo et al., 2018
A. rhodacantha		0.49	–	–	–	–	Yes	Trejo et al., 2018
A.maximiliana	Wild Managed Cultivated	0.44 0.39 0.43	2.5 2.3 2.3	0.41 0.38 0.36	–	Bats	No	This study
Not *mezcalero Agave* species								
Agave utahensis subsp. utahensis	Wild	0.49	5	0.24	–	Bats, birds, bees, and hawkmoths	Yes	Byers et al., 2014
A. hookeri	Cultivated	0.40		0.28	VP	–	Yes	Figueredo et al., 2015

H_E, expected heterozygosity; A, number of alleles per locus; F_{ST}, genetic differentiation; Φ_{ST}, genetic structure; R.S., reproductive system; SI, self-incompatible; Poll., pollinators; VP, vegetative propagation.

In this study, we explored the case of *Agave maximiliana* Baker growing in pine-oak forest in western Mexico. This is one of the *Agave* species used in Jalisco for the production of spirits other than tequila. The spirit known as "raicilla" forms part of the history in this region thanks to mining, an important activity during the colonial period, when raicilla was in high demand but also prohibited by the Spanish crown, along with all agave spirits at that time (Valenzuela-Zapata, 2015). The spirit is known as "raicilla de la sierra" (hereafter, "raicilla") corresponding to a highland region, and distinguishable from "raicilla de la costa," which is produced using other *Agave* species (*A. rhodacantha* and *A. angustifolia*) in the coastal region of Jalisco. The historical impact of this activity on *A. maximiliana* populations has never been evaluated, and no landraces or ecotypes have been documented.

Recently, our research group carried out studies through structured interviews. The study showed that most of the people who produce raicilla have actually learned the tradition transversally; i.e., not as a result of inherited knowledge. It appears that there has been a resurgence of interest in this activity and that the management practices of this species are therefore probably recent (Huerta-Galván, 2018). In addition, the study showed that nearly 90% of people interviewed conduct "promotion," a management practice that conserves many forest elements, such as pines and oaks that are maintained for the protection of the *Agave* plants against frost and pests. This action is a common element of *Agave* management in Mexico, especially on species managed at elevations of 800–2,500 m.a.s.l (e.g., *A. inaequidens*, *A. cupreata*), where frosts, are light and fleeting, but where some species such as *A. maximiliana* are frequently damaged (Gentry, 1982). The management of agave plants and native trees, both elements of the forest, are therefore combined in areas known as "forestry systems (FS)" which differ from AFS by the absence of domesticated crop elements (Nair, 1985; Torres-García et al., 2019). "Promotion" also includes the reintroduction of young plants (3 years of age) to the forest, by collecting seeds from multiple wild plants, the origin of which is located at a linear distance of less than 10 km, or buying plants from other producers. Most people also germinate seeds, the resulting seedlings of which are grown in their backyards and

remove weeds from their FS or monocultures, either manually with tools or through the use of herbicides. Some people conduct simple gathering, where they establish closed seasons to allow population regeneration, and some few (nearly 30%) grow their plants in monocultures, which is a much more recent development over the last 15–30 years since the current boom in mezcal began (Huerta-Galván, 2018). The current FS therefore provide most of the raw material that sustains raicilla production in this temperate region. Monocultures are characterized by a high transformation of the habitat, with no associated native trees within these systems, and greater dependence on the use of agrochemicals. It is notable that there is no clear criteria regarding what traits are selected, except for one producer who selects large plants with the highest sugar content in the leaves. Management practice details and associated native tree species, where applicable, are presented in **Supplementary Material SM1** for monocultures, forestry systems and wild populations. It is also important to mention that, while FS tolerate several tree elements naturally occurring in these habitats, a qualitative observation is that the density of shrub elements may be lower compared to that of wild populations. In contrast, the density of agave plants is generally higher in FS than in wild populations (**Figure 1**). While human actions are more evident in FS than in wild populations, it is important to mention that the latter are not pristine: they are found in accessible sites and present signs of light exploitation. Most of these populations present a small number of plants with the inflorescence removed, as these are occasionally sold as food.

A common perception among raicilla producers is that a drastic decrease (60–80%) has occurred in wild populations in the region in the last 30 years (stated by 60% of the informants in Huerta-Galván, 2018). If these perceptions reflect a true decrease, genetic bottlenecks and demographic stochasticity should be expected (Luikart and Cornuet, 1998; Frankham et al., 2006) in *A. maximiliana* populations. The plants are harvested prior to flowering, thus precluding the incorporation a high proportion of new individuals from one generation to the next (cf. Aguirre-Dugua and Eguiarte, 2013; Torres et al., 2015b). Moreover, the density of flowering plants is an important attractant for pollinators of *Agave* (Estrella-Ruiz, 2008) such that, even when some plants are allowed to flower, these may still be insufficient in number to maintain an effective pollination dynamic. Thus, overexploitation promotes population decline, leading to some cases of local extinction (e.g., *Agave potatorum*, Delgado-Lemus et al., 2014). There are no studies addressing the pollination system in *A. maximiliana* but, given the general aspects described above, i.e., paniculate inflorescences, probable nocturnal anthesis dehiscence and geographic area overlapping with that of *Leptonycterys yerbabuenae*, *L. nivalis* Saussure, 1860 (Arita, 1991) and *Choeronycteris mexicana* Tschudi, 1844 (Arroyo-Cabrales et al., 1982; Ortega-García et al., 2017), we infer a chiropterophilous syndrome, with these bat species as likely pollinators.

Finally, we envision social changes for the management of *A. maximiliana* in the short term, given that raicilla recently obtained its denomination of origin (DOR, Diario Oficial de la Federación, 2019). Considering the case of tequila production in the state of Jalisco, where an increased demand for plants to produce this spirit occurred following its DOR (Bowen and Valenzuela-Zapata, 2009; Leclert et al., 2010), we consider that a similar increase in the demand of plants for raicilla production is highly probable.

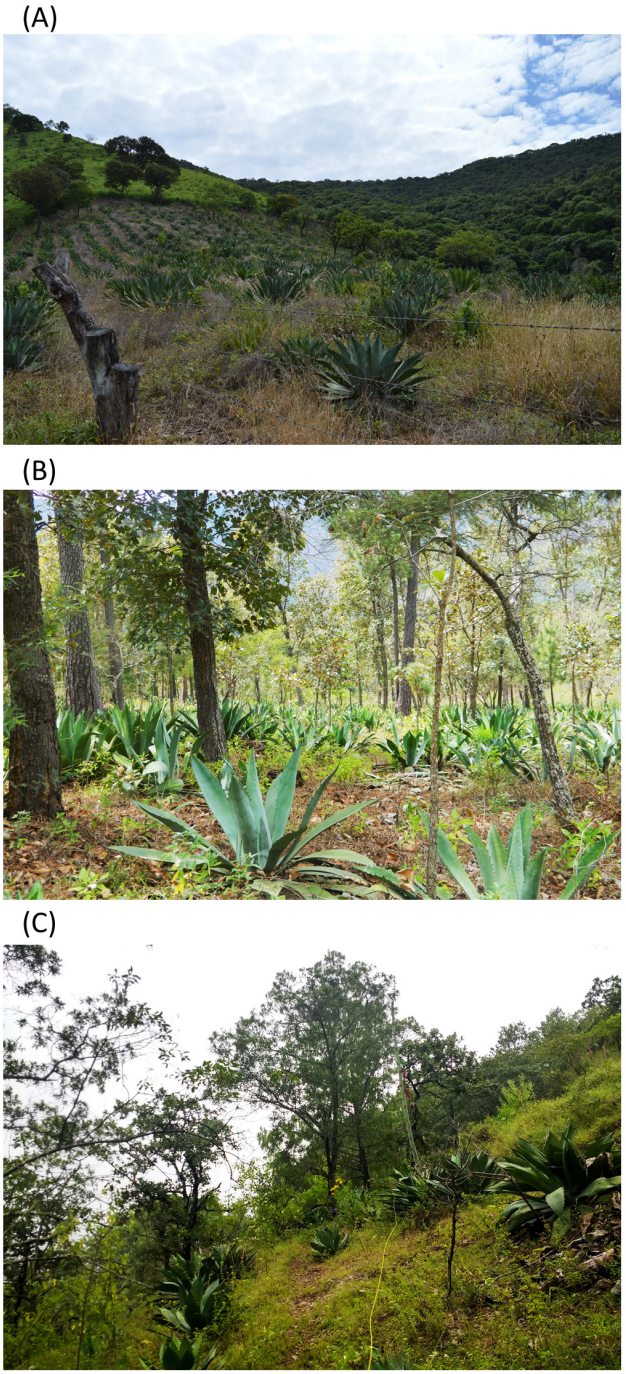


FIGURE 1 | Categories of studied populations of *Agave maximiliana*. **(A)** Monoculture "El Mosco," **(B)** Forestry System "Rincón de Mirandillas," **(C)** Wild population "El Palmito." Photos by: SA-R, DC-T and OH-G, respectively.

In summary, we consider that a recent but dynamic management exists in *A. maximiliana*, where increased population density of this species is a priority, along with the conservation of dominant forest elements, the benefits of which are still recognized by most of the producers; seeds and plants from multiple sites in the region are constantly reintroduced to forestry systems and monocultures and a high exchange of these germplasm also takes place among producers. These actions have acted to maintain, to date, the forestry systems that in turn support most raicilla production. Such systems conserve many elements of the original habitat and they are still governed, to some extent, by the environmental and geographical location effects of the natural landscapes. This is in contrast to the agricultural systems, i.e., monocultures, that present higher levels of habitat modification and more artificially controlled natural processes (e.g., pests, germplasm origin, density and reproduction of plants, etc.). Finally, we assume that the sexual reproductive system and probably chiropterophilous pollination system of *A. maximiliana* are biological strengths that have acted to maintain the connectivity of populations through gene flow.

In the context of these observations and assumptions, we aim to determine: (1) whether the methods and intensity of the management practices realized in *A. maximiliana* populations have had an effect on their intraspecific diversity; and (2) whether isolation by environment (IBE, Lichstein, 2007; Wang, 2013) and by distance (IBD) are significant in this species. We hypothesized that: (1) the recent management context will favor morphological and genetic diversity in this species, where divergence among categories is incipient, and thus we expect high morphological and genetic diversity for the species, with equal or higher levels in forestry systems compared to monocultures and wild populations and low differentiation among management categories; (2) while an assumed wide geographical range of the foraging behavior of the candidate pollinators of *A. maximiliana* will be reflected in a high genetic connectivity, the influence of geographical and environmental factors will have a positive relationship with the morphological and genetic divergences; therefore, we expect low genetic differentiation among populations and significant isolation by distance (IBD) and by environment (IBE). This is to say, the greater the geographical distance and the environmental differentiation, the greater the genetic and morphological divergence. Our purpose was to delimit and define the state of diversity and intraspecific differentiation of *A. maximiliana* growing in western Mexico in order to identify biological, ecological and management weaknesses and strengths in the face of imminent socioecological changes in the region.

MATERIALS AND METHODS

Study Species

Agave maximiliana, known locally as "Lechuguilla," is endemic to Mexico. It commonly inhabits rocky slopes in the states of Jalisco, Sinaloa, Durango, Nayarit, Colima and Zacatecas, at 900–2,000 m.a.s.l, and requires an annual precipitation of 750–1,000 mm (Gentry, 1982). It is frequent in pine-oak forests on calcareous soils or on those derived from igneous rock (Gentry, 1982; González-Elizondo et al., 2009). It is a perennial single rosette, monocarpic and medium size plant, balled or open, that habitually lacks asexual reproduction. Its leaves are broadly lanceolate, curved or straight, soft fleshy, with heteromorphic teeth of greater size toward the middle part of the leaf. The inflorescence is paniculate, of 5–8 m in height, branched, and with 15–30 small umbels. The greenish-yellow flowers are 52–65 mm in length. The fruits are small capsules of 3.5–5 × 1.7–2 cm, and are short oblong, stipitate, with rounded apex. The seeds measure 5.5–6 × 4.5–5 mm, the testa wavy, finely punctate, with marginal wing abruptly raised (Gentry, 1982; Vázquez-García et al., 2007). Gentry (1982) recognized two varieties with an overlapped distribution (**Figure 2**): *A. maximiliana* var. *maximiliana* and *A. maximiliana* var. *katherinae* (A. Berger) Gentry. The latter differs from the typical variety by its larger rosettes, greener leaves with more undulate repand margins with numerous variable interstitial teeth, and a deeper flower tube. However, McVaugh (1989) treated *A. maximiliana* var. *katherinae* as synonymous of *A. maximiliana sensu stricto*. Recently, Scheinvar (2018) defined the current potential distribution maps of both varieties and indicated that *A. maximiliana* var. *katherinae* could occupy the northern windward zone of the Sierra Madre Occidental (SMO), while *A. maximiliana* var. *maximiliana* could be found in the southern leeward zone of the SMO, but also in the Sierra Madre del Sur (SMS) and Transmexican Volcanic Belt (TVB) (Morrone et al., 2017). We perceived a wide morphological variation within populations in all sites sampled and decided to use *A. maximiliana* s.l. as recognized by McVaugh (1989).

Study Area, Populations Sampled and Categorization

In accordance with our hypothesis, core populations of our study area were collected in western Jalisco in municipalities that produce raicilla (**Figure 2**). Populations sampled outside this area, in the states of Zacatecas, Nayarit, Sinaloa, and Durango, were considered wild populations (hereafter, "wild") with little or no evidence of exploitation; these were used in the study as an external reference and for comparison of management gradients. Forestry systems (hereafter, "managed") are those in which one or several of the practices described above are carried out (*sensu* Huerta-Galván, 2018). The monocultures (hereafter, "cultivated") were established very recently, from around 30 or less years ago. The farmers use land in open areas that were previously used for other monocultures, usually maize. *Agave* rows are planted, almost always in lines following the slope. Agrochemicals are added to fertilize the plants or to eliminate weeds. Some farmers also apply chemicals on the leaves to combat pests. Monocultures have a high density of plants compared to plantations managed under the forest, and are characterized by a drastic change in land use. The seeds used for monocultures are collected from plants in the forest, domestic cultivation and *in vitro* propagation. This latter propagation type is currently carried out by a single producer only, who obtains 2–10 seedlings from each seed. This strategy is a response to the scarce availability of seeds and plants in recent years (Huerta-Galván, 2018).

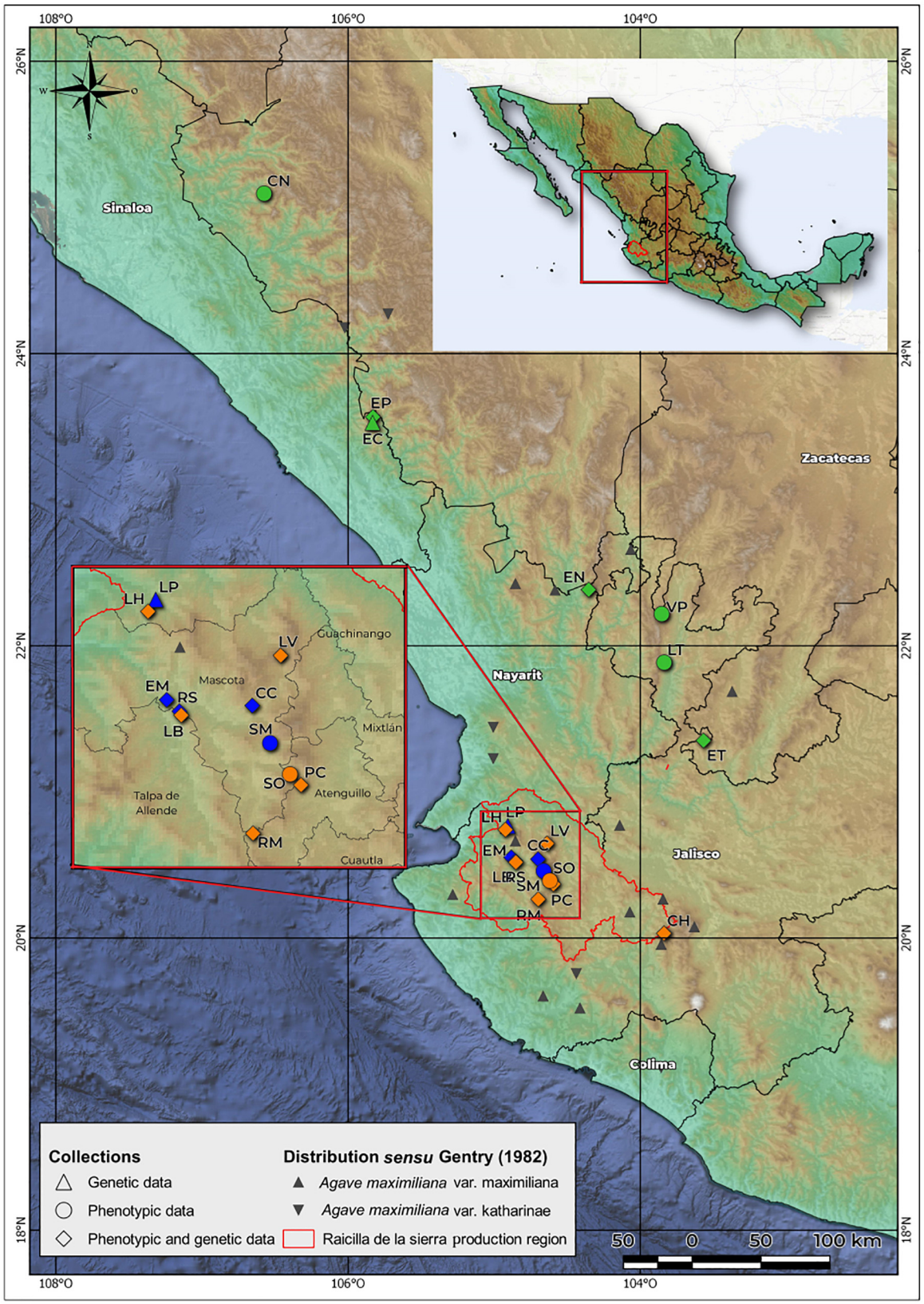


FIGURE 2 | Geographical location of sampled populations. Green: wild populations. Orange: managed populations. Blue: cultivated populations. Raicilla de la sierra production region is the area in which raicilla is distilled and the *A. maximiliana* populations are apparently overexploited.

Morphological data were obtained from 17 populations and the genetic analyses included microsatellite markers from 14 populations. For some populations, it was not possible either to obtain successful amplifications or to collect morphological data or leaf tissue (see **Table 2**). Finally, 12 populations were analyzed for environmental and geographic analyses using both genetic and morphological data.

Data Analyses

To answer our research questions, the following general strategy was adopted: For testing the contribution of management to the diversity and differentiation of *A. maximiliana* populations, we first calculated levels of diversity in the morphological traits and tested for differences among categories. We then employed multivariate methods and permutational analysis of variance to test whether clusters and the percentage of variation contained among cultivated, managed and wild categories can be identified. Analogous methods were conducted for genetic traits where differentiation among populations was also important in terms of detecting the contribution of gene flow. Finally, to identify the contribution of geographical distribution and environmental factors to the morphological and genetic divergences, we determined the IBD and IBE.

Diversity and Differentiation of Morphological Traits

Ten adult individuals were selected per population. Only plants with reproductive structures or visible indications of imminent reproductive maturity were selected, following the criteria of the producers. In total, 16 quantitative morphological traits were evaluated, as proposed by Vargas-Ponce et al. (2007) and Figueredo-Urbina et al. (2017) for agaves (**Table 3**). These traits are vegetative characters, since the presence of inflorescences is rare in cultivated and managed populations, and vegetative rather than floral characters are the targets of artificial selection. In each individual, three leaves of the mid verticil were measured in order to record variation and obtain mean values for these traits. Measurements were taken with a flexometer (resolution ± 0.001 m) and a ruler of 30 cm in length, and the data were used to construct a matrix for subsequent analysis. To discriminate the correlated morphological traits, a Pearson correlation test ($r \geq 0.9$) was used following the criteria of Valdivia-Mares et al. (2016). Four traits (i.e., minimum diameter (MinD), leaf length (LL), leaf width at middle (LWm) and number of teeth in 10 cm^2 (NT10) were excluded from the analysis because they presented multicollinearity (**Table 3** and **Supplementary Material SM2**).

To analyze the morphological diversity per population and predicted management category, the Index of Morphological Diversity (IMD) was estimated. For this, the traits with the highest eigenvalues in the LDA (explained below) were selected in order to utilize those that better explained morphological diversity. Individual plant values of each trait from all populations were assigned to different morphological states according to the Sturges rule (Daniel, 2006). This procedure produced a matrix of discrete values and the IMD was estimated based on the Simpson Diversity Index (SDI). This index was defined as $IMD = 1 - \Sigma_{1-s}(pi)^2$, in which pi is the proportion of the total number of individual plants sampled in a population showing the *i*th state of a morphological trait (categorized with the Sturges rule) and *s* is the number of states of that character (Casas et al., 2006). In this way, a unique estimator of the mean morphological diversity was obtained for all traits per population and category (cf. Figueredo-Urbina et al., 2017). The existence of significant differences among management categories by IMD was evaluated using a Kruskal Wallis test in Primer v6.1.11 (Anderson, 2017).

To estimate the multivariate differentiation among groups and determine which morphological traits have greater influence on the dispersion of the variation, a Linear Discriminant Analysis (LDA) was performed. For this, the database was standardized with the expression $Y_0 = (Y–a)/b)$; where Y_0 is the standardized value, Y is the observed value of the state of a character, *a* is its mean and *b* is its standard deviation (Figueredo-Urbina et al., 2017). From the LDA, the centroids were obtained for each population, paired Euclidean distances were calculated and random resampling conducted to construct the null model (with 10,000 replicates) as a test of significance. To corroborate the identity of the cultivated plants, on the assumption that cultivated plants came mainly from managed rather than wild populations (if either the farmers collected seeds or bought the whole plant), we cross-validated the population assignation of each individual, from which we obtained a confidence value (i.e., the probability of correctly classifying an individual with respect to its population of origin). With these values and the predict function obtained from the LDA, we estimated the probable population of origin for each cultivated individual in monocultures (Legendre and Legendre, 1998; Medina-Villarreal and González-Astorga, 2016). All analyses were performed in R (R Development Core Team, 2012).

In order to determine whether the existing set of phenotypes differs significantly among management categories, a nested two-way permutational multivariate analysis of variance (PERMANOVA) was conducted with a type III (mixed effects) model: a fixed factor (condition) with three levels (wild, managed and cultivated) and a random factor (populations) with 17 levels. A PERMANOVA was performed with the records obtained from twelve non-correlated traits in Primer v6.1.11 /PERMANOVA+ v1.0.1 (Anderson et al., 2008). First, a pretreatment was conducted in order to standardize the data (z-values) of the 12 traits analyzed, in order to transform the variables of different nature and measurement units to the same scale. A matrix of Euclidean distances was then constructed and the PERMANOVA performed. The statistical significance of the analysis was tested with 10,000 permutations based on a type III sum of squares. Likewise, the general patterns of morphological variation of the studied populations in each management category were analyzed using a Principal Coordinates Analysis (PCoA, Legendre and Legendre, 1998). This PCoA was performed based on the mean values per population of the 12 morphological traits used in the

TABLE 2 | Populations of *Agave maximiliana* sampled for this study.

Population	Code	Data	State	Location	Elevation (m.a.s.l)
Cultivated					
Cimarrón Chico	CC	Phe/Gen	Jalisco	Mascota	1,519
El Mosco	EM	Phe/Gen	Jalisco	Mascota	1,181
Rincón Seco	RS	Phe/Gen	Jalisco	Mascota	1,267
San Miguel	SM	Phe	Jalisco	Mascota	1,599
Las Palmas	LP	Gen	Jalisco	San Sebastián del Oeste	1,000
Managed					
Chiquilistlán	CH	Phe/Gen	Jalisco	Chiquilistlán	1,938
La Berenjena	LB	Phe/Gen	Jalisco	Mascota	1,380
La Vieja	LV	Phe/Gen	Jalisco	Mascota	2,440
Los Hornos	LH	Phe/Gen	Jalisco	San Sebastián del Oeste	1,317
Puerto la Campana	PC	Phe/Gen	Jalisco	Atenguillo	1,968
Rincón de Mirandillas	RM	Phe/Gen	Jalisco	Mascota	1,630
Sol de Oros	SO	Phe	Jalisco	Mascota	1,876
Wild					
El Nayar	EN	Phe/Gen	Nayarit	El Nayar	2,269
El Palmito	EP	Phe/Gen	Sinaloa	Concordia	2,033
El Teúl	ET	Phe/Gen	Zacatecas	Teúl de González Ortega	1,699
Canelas	CN	Phe	Durango	Canelas	1,800
La Toma	LT	Phe	Jalisco	Bolaños	2,022
Valparaíso	VP	Phe	Zacatecas	Valparaíso	2,290
El Carrizo	EC	Gen	Sinaloa	Concordia	1,943

Phe/Gen, Phenotypic and genetic data; Phe, phenotypic data; Gen, genetic data. m.a.s.l, meters above sea level. The localities are shown in ***Figure 2****.*

TABLE 3 | Vegetative characters measured for wild, managed and cultivated populations of *Agave maximiliana*.

Variable	Code	P	Units
Total plant height	TPH	1	cm
Minimum diameter[a]	MinD*	1	cm
Maximum diameter	MxD*	1	cm
Leaf length[a]	LL	3	cm
Maximum leaf width	MxLW	3	cm
Leaf width at middle[a]	LWm	3	cm
Terminal thorn length	TTL	3	mm
Terminal thorn width at the base	TTW	3	mm
Number of teeth	NT	3	Amount
Number of teeth in 10 cm^{2}[,a]	NT10	3	Amount
Longest tooth length	LTL	3	mm
Number of teeth/leaf length (thorniness)	NT/LL	3	Ratio
Number of teeth in 10 cm/leaf length (spacing)	NT10/LL	3	Ratio
Leaf width at middle/Maximum leaf width (width index)	LWm/MxLW	3	Ratio
Leaf length/Maximum leaf width (leaf shape)	LL/MxLW	3	Ratio
Terminal thorn width at the base/Terminal thorn length (thorn shape)	TTW/TTL	3	Ratio

*P, Number of parts measured by individual; *two diameters were measured, in order to use the greatest one. [a]Trait excluded from the morphological analysis due to high collinearity (see* ***Supplementary Material SM2****).*

PERMANOVA. Therefore, the same pretreatment and Euclidian distance considered in the PERMANOVA design were used.

Diversity and Differentiation of Genetic Traits

From each population, samples of plant tissue were collected from 15 individuals, transported at 4°C and stored at −45°C. From this tissue, DNA was extracted following Doyle (1991), modified with sodium chloride-tris-EDTA (STE). The DNA was quantified with a spectrophotometer and its quality evaluated by absorbance at 260/280 ng/μL. Dilutions at 40 ng/μL of all samples were prepared. A total of 16 microsatellites were tested; seven described by Lindsay et al. (2012), four by Parker et al. (2010) and five designed by Cabrera-Toledo (unpublished data)

from the genome of *A. tequilana*. Only nine microsatellites were reproducible and polymorphic (**Supplementary Material SM3**). Amplification by PCR was conducted using the Multiplex PCR kit with TaqMan® (Qiagen) in a final reaction volume of 11 μL. Each reaction contained 120 ng of DNA, 4 μL of Multiplex, 0.8 μL of each primer (10 mM) and 2.4 μL of molecular grade water (MiliQ). The amplification conditions were: 15 min at 95°C, followed by 40 cycles of 30 s at 94 °C, 30 s at the annealing temperature of each primer (**Supplementary Material SM3**) and an extension step of 2 min at 72°C, followed by a final extension step of 30 min at 60°C. The amplification was corroborated in 1.5% agarose gels with 1X TBE buffer. The PCR products of each primer per population were separated by vertical electrophoresis in 8% polyacrylamide gels (1–1.5 mm thickness) without urea. Each gel was left to run for 5 h. A 100 bp marker (Invitrogen) was used as a reference, and silver staining was used to visualize the bands (Sanguinetti et al., 1994). The gels were photo-documented with the software Gel Logic 100, Kodak and the fragment size in base pairs of the amplified bands was determined in Phoretix 1D v10 (Total lab Ltd.).

To analyze the genetic diversity, five parameters were evaluated: percentage of polymorphic loci (P), mean number of alleles per locus (A), mean number of effective alleles per locus (A_E), observed heterozygosity (H_O), expected heterozygosity (H_E) under Hardy-Weinberg equilibrium (HWE). Also, we performed HWE analysis based on the X^2 test (GenAlex 6.5, Peakall and Smouse, 2006). A Kruskal-Wallis test was conducted (SigmaPlot, v 13.0) to evaluate the significance of the differences in genetic diversity among management categories. The fixation indices F_{IS} and F_{IT} of Wright (1965) were used to evaluate inbreeding. To evaluate the possible effect of null alleles on the levels of inbreeding, a Bayesian Individual Inbreeding Model (IIM) analysis was performed in INest 2.2 (Chybicki and Burczyk, 2009). This permits non-biased calculation of the coefficient of inbreeding for multilocus data in the presence of null alleles (Fi). This analysis evaluates the effect of null alleles (n), inbreeding (f) and genotyping errors (b) on the values of homozygosity using the complete model (*nfb*) and its comparison with the model excluding the effect of the null alleles (*fb*), through the Deviance Information Criterion (DIC) between the two models. The analysis was run with 500,000 MCMC iterations and 50,000 burn-in cycles. Where the *nfb* model obtains the lowest DIC, this indicates that the null alleles have an important weight in the calculation of inbreeding. The null alleles also have a potential effect on genetic differentiation (F_{ST}). To evaluate this, the FreeNA software was used to perform this estimation by detecting the frequency of null alleles for each locus and population based on the Expectation-Maximization (EM) algorithm (Dempster et al., 1977; Chapuis and Estoup, 2007). The population differentiation (F_{ST}) was evaluated excluding null alleles, i.e., the ENA correction method (Chapuis and Estoup, 2007). A total of 50,000 bootstrap replicates were used to calculate the confidence intervals of the corrected global F_{ST}.

Multivariate analyses were used to identify the genetic structure. First, analysis of molecular variance (AMOVA) was performed (GenAlex 6.5, Peakall and Smouse, 2012) to quantify the proportion of diversity within populations, among populations and among categories of management. This analysis uses the Φ_{ST} index, which is analogous to Wright's index of differentiation F_{ST} (Hedrick, 2005). In addition, a Discriminant Principal Components Analysis (DPCA) (*Adegenet*, R Development Core Team, 2012) was executed to visualize the genetic structure in a graphic form. DPCA retains the virtues of a discriminant analysis, defining a model in which components of the genetic variation are maximized among groups and minimized within groups. The DPCA is supported by a transformation using a Principal Components Analysis (PCA) as a first step, guaranteeing that the subjected variables are not correlated and are lower in number than those of the individuals analyzed (Jombart et al., 2010). Finally, to visualize the degree of differentiation between pairs of populations, a paired matrix of genetic distances of Nei and Chakraborty (1973) was generated in the program GenAlex version 6.5 (Peakall and Smouse, 2012), with which a dendogram was constructed using the UPGMA method.

Isolation by Distance/Isolation by Environment

Genetic/Morphological Distances – Geographic+Environmental Distances Matrix

To identify whether similar processes are molding morphological and genetic distances among populations, a simple regression analysis between these two distance matrices was conducted using the Multiple Regression Method (MRM) (Legendre et al., 1994; Lichstein, 2007) in R platform (ecodist v2.0.1, Goslee and Urban, 2007). The significance of the coefficient of regression was estimated with 1,000 permutations.

Separately, to test for IBD and IBE, we analyzed the relationship of genetic distances with geographic and environmental conditions, respectively, using the MRM method. The significance of the coefficient of regression was estimated with 1,000 permutations. The genetic distances were calculated by using the pairwise comparisons corrected for the presence of null alleles (Chapuis and Estoup, 2007) and linearized according to the equation $F_{ST}/1 - F_{ST}$. The geographic distance matrix was generated using the geosphere package version 1.5.10 (Hijmans, 2015). To construct an environmental distance matrix, the 19 environmental variables of WorldClim Global Climate data version 2[1], with a resolution of 30 arcsec (1 km^2), were obtained. To reduce multicollinearity among environmental variables, those that presented a correlation ≥0.8 (JMP version 9, SAS Institute) were excluded. Separately, for each environmental variable, a simple regression analysis between the environmental distance matrix and the genetic distances was performed using the MRM method. In order to capture the most informative variables, those that exhibited a significant association with genetic distances were selected. Finally, Principal Component Analysis (PCA) was used to reduce the variation of these data, and the two first Principal Components (explaining 97% of the total variation) were used to calculate the Euclidean similarity distance matrix (PAST version 3.24, Hammer et al., 2001).

[1]https://worldclim.org/version2

To test IBD and IBE in morphological distances, we performed the MRM method (1,000 permutations) using the geographical and environmental matrices generated with the procedure explained above. The morphological distance matrix was obtained based on the 12 morphological traits and the same pretreatment and Euclidian distance used in the PCoA and PERMANOVA (Primer v6.1.11.).

RESULTS

Morphological and Genetic Diversity

The Index of Morphological Diversity (IMD) reflected a general mean of 0.638 ± 0.07 (interval 0.48–0.75, **Supplementary Material SM4**). In the cultivated populations, a narrow interval of 0.620–0.673 was observed. In the managed populations, this interval broadened to 0.493–0.747 and in the wild populations it was 0.460–0.700. However, no significant differences were found in mean IMD among categories or populations (H = 16, p = 0.453), suggesting that the plants were morphologically diverse regardless of management condition or the population to which they belong. However, variation tended to be lower among cultivated populations (SD ± 0.03), compared to the managed (SD ± 0.09) and wild populations (SD ± 0.1) (**Supplementary Material SM4**).

The mean percentage of polymorphic loci for all of the populations was 94.75%, the mean number of alleles per locus was 2.37 and the mean number of effective alleles was 1.87. There was an observed heterozygosity of 0.295 and an expected heterozygosity of 0.418 (**Supplementary Material SM4**). Mean values per management category (**Table 4**) did not present significant differences in any of the parameters of genetic diversity (P, A, A_E, H_O and H_E). On average, close to half of the loci (49.2%) were found to be in Hardy-Weinberg equilibrium, with slight variation presented in each category (cultivated 41.67%, managed 51.85% and wild 52.78%). Another 38.09% of the loci, on average, were found to present a deficit of heterozygotes, also with slight variation among categories (cultivated 38.89%, managed 35.18% and wild 41.67%). A total of 7.14% of the loci presented greater heterozygosity than expected in equilibrium: the highest average (13.89%) was found in the cultivated plants, followed by the managed (5.56%) and wild (2.78%) plants. The inbreeding coefficient was F_{IS} = 0.273 ± 0.134, only five of the nine loci were significant (estimated X^2 > critical value, p = 0.001). The loci that did not present local inbreeding corresponded to the primers 12, 1763, 28 and 20. A frequency of null alleles was found in a range of 0–0.392. Inbreeding corrected for null alleles was of Avg (Fi) = 0.1529, with a density interval after 95% of 0.0024–0.3179. The model with the lowest DIC value was *nfb* (DIC: 10216), compared to the model *fb* (DIC: 10424), indicating that the null alleles had an important effect on the calculation of the inbreeding.

Morphological and Genetic Structure

The first two functions of the Linear Discriminant Analysis (LDA) explained 70% of the morphological variation. The maximum leaf width (MxLW) had the highest value in the LDA1 and traits related to plant thorniness (NT/LL; NT) and plant size (MxD) had the highest values in the LDA2. This analysis did not define a congruent grouping with the management categories established *a priori* (**Figure 3** and **Supplementary Material SM5**). The cross validation test showed confidence values of between 30 and 100%, which corresponded to the proportion of individuals correctly identified in their population of origin based on morphological similarity. Prediction functions reclassified monocultures by their morphological similarity within mainly managed populations (**Figure 4**). Moreover, the PERMANOVA evidenced that 36% of the variation was presented among the populations (p < 0.05) with a non-significant 19.8% variation found among *a priori* categories (**Table 5**). Likewise, the PCoA showed a wide pattern of variation similar to that identified by the LDA and PERMANOVA. In PCoA (**Figures 5A,B**) three groups of populations and two isolated entities were distinguished. Group I comprised LH, EP, RS and CN, which are plants with a tendency to present high numbers of teeth (NT ≥ 73), large terminal thorns (TTL ≥ 32 mm) and more elliptic leaves (LWm/MxLW = 0.9) compared to the rest of the populations, which tend to have spathulate leaves (LWm/MxLW = 0.8). Group II comprised CH, CC, EM and LB, which include the largest and robust plants, of greatest height (TPH ≥ 116 cm), maximum diameter (MxD ≥ 176 cm), and with the widest leaves (LWm ≥ 19.5 cm). Group III comprised VP, EN, RM, SM, PC, LV and LT, which present plants of smaller dimensions (compared to group II) and most populations with a high thorniness index (NT10/LL = 0.2), as well as shorter and stouter terminal spines (TTW/TTL = 0.2). Finally, the ET and SO populations are isolated entities: ET (group IV) had the smallest plants (TPH = 58 cm; MxD = 101 cm; LWm = 10.8 cm) with the highest thorniness (NT10/LL = 0.4; NT/LL = 1.5) and, finally, SO (group V), much like group II, had large plants but presented the lowest thorniness index of all (NT10/LL = 0; NT/LL = 0.5) (**Figures 5A,B**). Even though such grouping tendencies can be observed, the values of the variation coefficient are wide in all traits (**Supplementary Material SM6**), constituting another expression of the high morphological variation of the species.

With respect to the genetic data, the AMOVA (**Table 6**) revealed that 61% of the variation was found within populations and 38% was found among populations (Φ_{CT} = 0.38, p = 0.001), which is similar to that found with the morphological data, while 1% corresponded to the variation among management categories (Φ_{PC} = 0.01, p = 0.003). Population differentiation, corrected for null alleles, was F_{ST} = 0.43 with ± 0.34–0.51 confidence intervals. Finally, the DAPC showed no genetic structure congruent with the pre-established categories (**Figure 6**). Most of the populations overlapped in the multidimensional space, apart from three: two wild (EN and EP), which were the most geographically distant, being located in Nayarit and Sinaloa, and one managed (RM). The population EC is geographically very close to EP (**Figure 2**); however, it overlapped genetically with managed and cultivated populations (**Figure 6**). The UPGMA showed a pattern of genetic distances that is not congruent with geography, but which groups together almost all of the cultivated and managed populations (except for EM and LV) (**Figure 7**). The wild EP is the

TABLE 4 | Mean genetic and morphological diversity in populations of *A. maximiliana*.

Categories	N	NP	P	A	A_E	H_O	H_E	IMD
Cultivated	12.44	4 (G,Ph)	94.44	2.30	1.89	0.359	0.427	0.648
	±0.82		±6.41	±0.14	±0.16	±0.08	±0.07	±0.03
Managed	12.7	6 (G), 7 (Ph)	92.59	2.33	1.82	0.248	0.388	0.674
	±1.13		±13.45	±0.29	±0.22	±0.07	±0.07	±0.09
Wild	12.5	4 (G), 6 (Ph)	97.22	2.5	1.92	0.277	0.438	0.591
	±1.07		±5.55	±0.23	±0.18	±0.10	±0.07	±0.10
Total	12.55	14 (G), 17 (Ph)	94.75	2.37	1.87	0.295	0.418	0.638
	±0.19		±2.54	±0.06	±0.05	±0.02	±0.02	±0.07

N, sample size per population; NP, number of populations, G, genotypic data, Ph, phenotypic data; P, percentage of polymorphic loci; A, number of alleles per locus; A_E, effective number of alleles per locus; H_O, observed heterozygosity; H_E, expected heterozygosity in Hardy-Weinberg equilibrium; IMD, index of morphological diversity ± standard error.

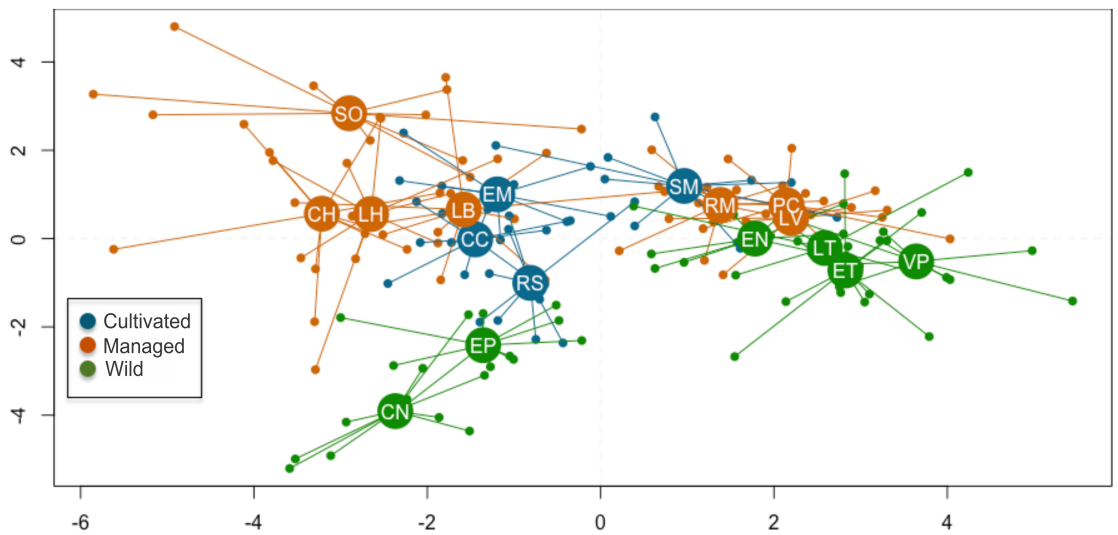


FIGURE 3 | Scatterplot of discriminant analysis, explaining 70% of the total morphological variation. Small circles represent the 170 adult individuals evaluated, while larger circles are the centroids in each population. See **Table 2** for definition of codes.

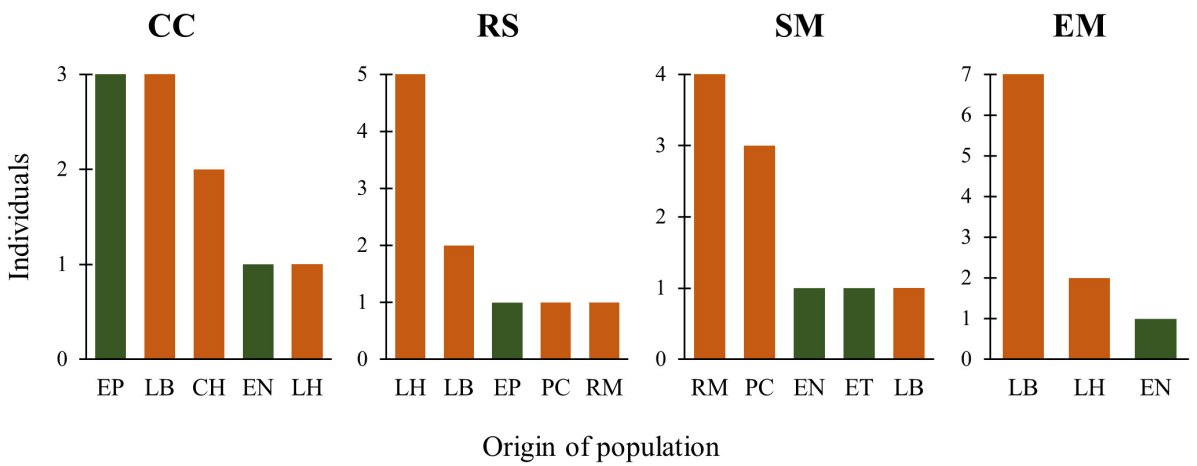


FIGURE 4 | Histograms of probable origin of cultivated plants according to their morphological similarity. Green: wild populations, orange: managed populations. Black letters are monoculture codes. CC = Cimarrón Chico, RS = Rincón Seco, SM = San Miguel, EM = El Mosco. The names of the population of probable origin follow those in **Table 2**.

TABLE 5 | Permutational analysis of variance (PERMANOVA) for twelve morphological traits in wild, cultivated, and managed populations of *A. maximiliana*.

Source	*df*	SS	MS	Pse-F	*p*-value	Unique permutations	% CV
Among categories	2	232.25	116.12	2.2255	0.0551	9916	19.80
Populations (categories)	14	730.49	52.178	7.4941	0.0001	9816	36.40
Residual	153	1065.30	6.962	–	–	–	45.18
Total	169	2028	–	–	–	–	–

df, degrees of freedom; SS, sum of squares; MS, mean squares; Pse-F, Pseudo-F statistic; % CV, percentage of the components of the variation.

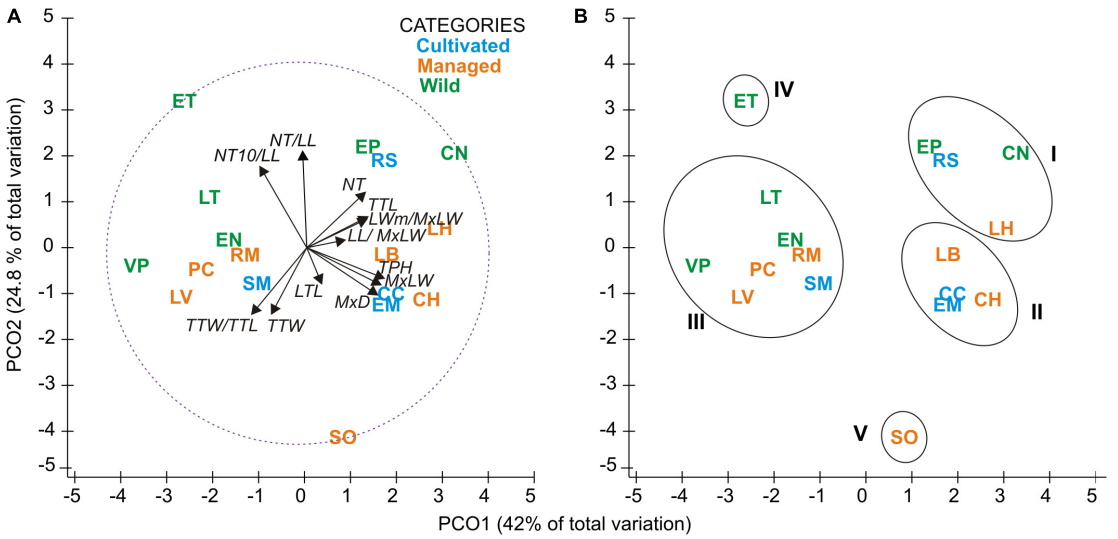


FIGURE 5 | Principal coordinates analysis (PCoA) ordination; **(A)** Relationships among populations and morphological traits. Longitude of vectors corresponds to the magnitude of the contribution (correlation) of each trait, and the circle with dotted line represents the Pearson correlation ranges from -1 to 1. **(B)** Ellipsoids correspond to population groups (I to III) with similar traits, and two isolated entities (IV and V) with divergent traits. See **Tables 2** and **3** for definition of codes.

most distant population of a second cluster group comprising a cultivated (EM) and a managed population (LV) from Jalisco, and two wild populations (EN) from Nayarit and (EC) from Sinaloa.

Isolation by Distance/Isolation by Environment for Genetic and Morphological Distance Matrices

Morphological and genetic distances were not related to each other (R^2 = 0.02, p = 0.22). The general model for IBD and IBE in terms of genetic data was not significant (R^2 = 0.20, p = 0.06); however, we detected a significant effect of environmental variables on the genetic distance matrix (β = 0.32, p = 0.007). These variables were annual precipitation (BIO12; 785–1312 mm), precipitation of wettest month (BIO13; 187–333 mm), precipitation of driest month (BIO14; 2–11 mm), precipitation seasonality (BIO15; 87–115 mm), precipitation of warmest quarter (BIO18; 239–795 mm), precipitation of coldest quarter (BIO19; 45–156 mm). This result supports an IBE process, while the geographical distances matrix did not present a significant association with the genetic distances matrix (β = −0.13, p = 0.28) and therefore does not support an IBD process.

On the other hand, the MRM general model used to explore the influence of geographical and environmental variables on the morphological distance matrix was not statistically significant (R^2 = 0.20, p = 0.06), and neither were the regression coefficients for the geographic and environmental distance matrices (β = 0.03, p = 0.74 and β = 0.13, p = 0.12, respectively).

DISCUSSION

This research showed that forestry systems of *Agave maximiliana* are effective as reservoirs of morphological and genetic diversity, since their levels are similar to those presented in wild populations. This result is congruent with other *Agave* species with similar management practices (*A. inaequidens*, *A. cupreata*, *A. potatorum*, **Table 1**). However, despite having similar reproductive and pollination systems to these species, *A. maximiliana* showed high levels of morphological diversity while, contrary to our hypothesis, the species presented low levels of genetic variation and high genetic differentiation among populations. Morphological divergences are not congruent with management categorization, although a slight genetic differentiation is. Morphological and genetic distances are not correlated, which suggests that different ecological-evolutionary processes are governing these traits. Alternatively, the number of loci evaluated might be not enough to test an association of genetic and morphological traits. Thus, further studies,

TABLE 6 | Molecular analysis of variance based on 9 SSR loci, in wild, cultivated and managed populations of *A. maximiliana*.

Source	*df*	SS	MS	SV	Φ	*p*	% CV
Among categories (CT)	2	178.554	89.277	0.128	0.01	0.003	1
Among populations (PC)	11	886.710	80.610	4.876	0.38	0.001	38
Within populations	195	1521.257	7.801	7.801	–	0.001	61
Total	208	2586.522	–	12.806	–	–	100

Φ_{CT} indicates the structure among categories; Φ_{PC} indicates the structure among populations within categories (F_{ST} analogous), and their respective percentage of variation (%) explained by the model. df, degrees of freedom; SS, sum of squares; MS, mean squares; SV, standard variation; Φ, Phi statistics; p, p value; % CV, percentage of the components of the variation.

including more genetic markers, are needed to analyze this relationship. Precipitation variables influence the environmental heterogeneity that in turn partly governs genetic structure in *A. maximiliana*. At the same time, no isolation by distance or by environment was detected in terms of the morphological differentiation in population pairs.

Morphological and Genetic Diversity

Levels of morphological diversity in *A. maximiliana* are high (IMD = 0.628 on average, with an interval of 0.460–0.747), with values that are comparable to those reported for columnar cacti under *in situ* management categories, with mainly sexual (*Myrtillocactus schenckii* (J.A. Purpuss) Britton & Rose, IMD 0.652–0.703, Blancas et al., 2009) and, to a lesser extent, asexual (*Stenocereus pruinosus* (Otto ex Pfeiff.) Buxb. IMD 0.647–0.677, Parra et al., 2012) reproduction, but higher than *Stenocereus stellatus* (Pfeiff.) Riccob. with asexual reproduction (IMD 0.408–0.479, Casas et al., 2006). With respect to other agaves, the levels of morphological diversity are higher than those reported for species propagated from seeds (IMD = 0.413 *Agave inaequidens*; 0.489 *A. cupreata*) and from shoots (0.481 *A. hookeri*, Figueredo-Urbina et al., 2017). In all of the cases mentioned, managed populations present morphological diversity that is greater than or similar to that of wild populations. This provides evidence that traditional *in situ* management maintains the levels of diversity found in the wild state, and is in contrast to the forms of modern plant improvement responsible for the reduction of genetic and morphological diversity in global agriculture (Vellvé, 1992; Clunies-Ross, 1995). The evidence provided here confers *A. maximiliana* with the highest morphological diversity reported for a species of *Agave* to date.

In contrast, the low genetic diversity found in *A. maximiliana*, with similar values across the management categories (**Table 4**), suggests that there is a general biological and/or ecological process underway that is independent of the management context pre-established in this study. This result is incongruent with its life history characteristics, particularly the fact that it propagates from seed. The genetic diversity of *A. maximiliana* is lower than that of other long-living perennial cross-pollinating species (H_E = 0.61–0.68, Nybom, 2004), as well as that of its congeners *A. inaequidens* and *A. potatorum* (**Table 1**). It is barely comparable with other agaves of vegetative propagation (e.g., cultivated plants of *A. parryi*, *A. angustifolia*) and even equal or inferior to some landraces of domesticated species such as *A. angustifolia*, *A. rhodacantha* and *A. tequilana* (**Table 1**). Moreover, the levels of inbreeding found are significant at population level (F_{IS} = 0.273), although it should be noted that the inbreeding corrected for null alleles was lower (Fi = 0.153). This value must therefore be treated with some caution. Assuming that *A. maximiliana* has an outcrossing reproductive system, this would depend to a large extent on the pollinators. The low genetic diversity and presence of inbreeding in this species could be the consequence of low visit rates or inefficient pollinator behavior.

The Crenatae group, to which *A. maximiliana* (Gentry, 1982) belongs, includes species with strong inbreeding depression and a xenogamic pollination system that are pollinated by bats. In general, agaves with paniculate flowers are often pollinated by bats (Gentry, 1982; Rocha et al., 2006). However, the specific reproductive and pollination system of *A. maximiliana* has not been examined in detail. Arizaga et al. (2000) demonstrated in *A. macroacantha* Zucc. that, without pollination by bats, the plants produce only approximately 50% of the number of seeds. The size of the population of chiropterans that visit the plants influences the proportion of viable seeds (Howell and Roth, 1981; Arizaga et al., 2000). González (2005) found that visits by a low density of bats in *A. garciae-mendozae* Galván and Hernández produce low dispersion of pollen, since these bats visit flowers of a given plant at different stages, thus causing inbreeding.

Seed scarcity is frequently mentioned by various producers in the raicilla de la sierra region, and has led to the implementation of an *in vitro* reproduction technique from seeds in the local technical high school (pers. obs.). In some of the sample populations, there is also a high incidence of plants with poorly developed inflorescences; i.e., of small size and with few flowers, and with evidence of infection by fungi and bacteria, (pers. obs.). Low levels of genetic diversity are generally associated with inbreeding and this, in turn, can trigger a reduction in the average adaptation of the population, in a process known as inbreeding depression (Charlesworth and Charlesworth, 1987; Reed and Frankham, 2003; Hedrick, 2005). This association between genetic diversity and low seed production has been reported in congeners such as *Agave murphyi* Gibson (Gentry, 1982; Parker et al., 2007) and *Agave schottii* Engelm (Trame et al., 1995). Considering these arguments, inbreeding depression in *A. maximiliana* is an issue that merits further exploration in future studies.

It is notable that management has not eroded the genetic diversity of cultivated and managed populations and that these remain similar to that of the wild populations. This could reflect very recent processes of artificial selection, an

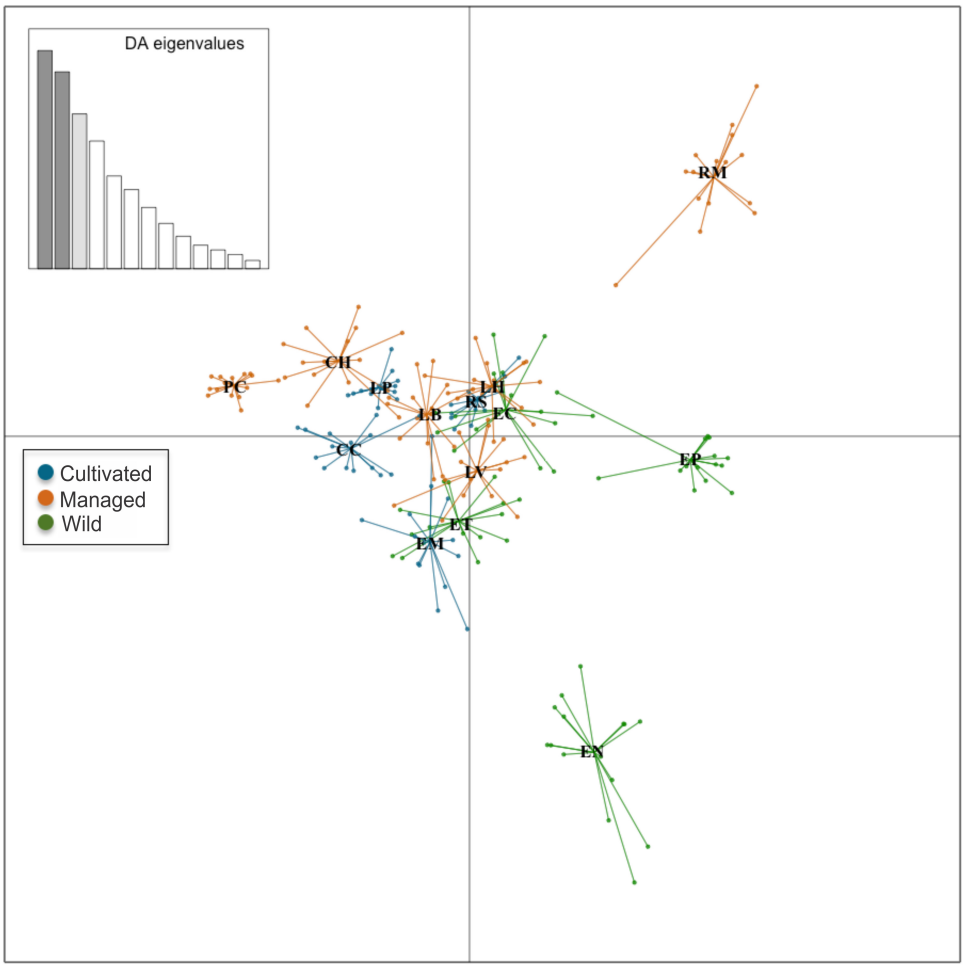


FIGURE 6 | Discriminant analysis of principal components in nine loci and 14 populations of *A. maximiliana*. Small circles represent the 229 genetic samples. Lines represent the genetic variation within populations.

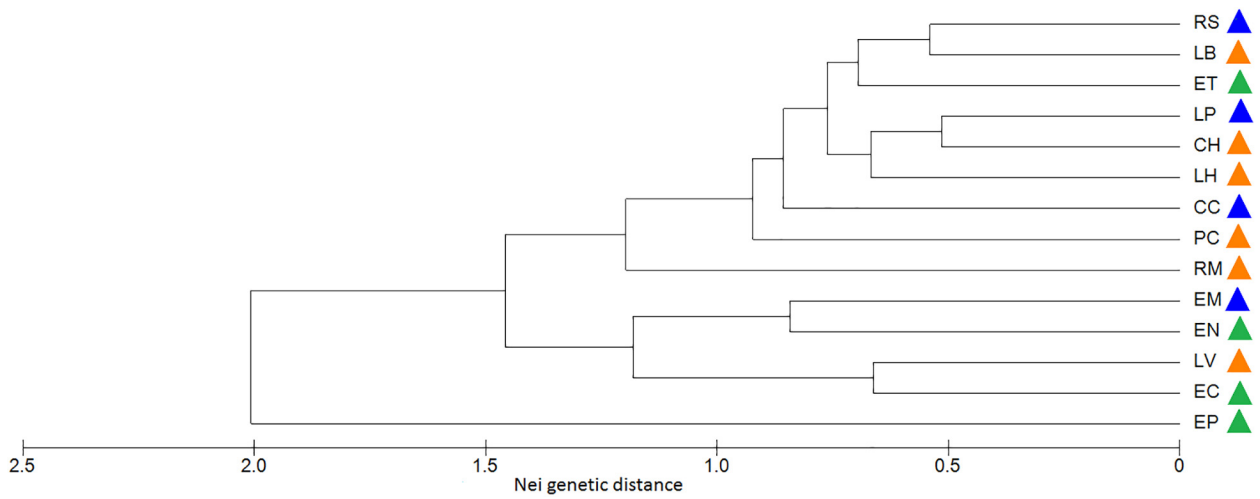


FIGURE 7 | Dendrogram of genetic similarities in *A. maximiliana* populations based on Nei genetic distances. Genetically closer populations show the origin of clusters skewed to the right by reference of the Nei genetic distance scale. Green: wild; orange: managed; blue: cultivated. See **Table 2** for definition of codes.

active exchange of seeds and plants among farmers or a low intensity of exploitation that still does not genetically impact the populations (Casas et al., 2006; Parra et al., 2010). The maintenance and increase in local diversity of *in situ* managed and cultivated populations is relatively common in perennial plants or those with wide generation times in other regions of Mexico and America (*Stenocereus pruinosus*, Parra et al., 2010; *S. stellatus*, Cruse-Sanders et al., 2013; *Myrtillocactus schenckii*, Blancas et al., 2009; *Spondias tuberosa* Arruda, Neto et al., 2013; *Pouteria sapota* (Jacq.) H. E. Moore and Stearn, Arias et al., 2015). In Mexico, this is due to the broad spectrum of activities of *in situ* management that not only include the extraction of plants, but also manual dispersion of seeds, tolerance of some individuals and the protection and promotion of plants with characteristics that are desired by humans, among others. These are practices that have been widely documented for agaves in central (Torres et al., 2015a) and western Mexico (Vargas-Ponce et al., 2007; Huerta-Galván, 2018) and for other useful plants in the Tehuacan Valley region (Blancas et al., 2010).

Morphological and Genetic Structure

Divergences in morphological and genetic characters evaluated for *Agave maximiliana* in the studied area respond mainly to ecological-evolutionary processes related to the habitat and, to a lesser extent, to management practices. This is evidenced by the wide dispersion of variation in morphological and genetic traits that does not correspond to the management categories established *a priori* and contrasts with that reported for congeners such as *A. parryi* (Parker et al., 2014), *A. inaequidens* and *A. cupreata* (Figueredo-Urbina et al., 2017), as well as other species in more advanced stages of domestication, such as *A. angustifolia* and *A. rhodacantha* (Vargas-Ponce et al., 2007, 2009). In these domesticated species, the phenotypic traits are adjusted to the assigned categories or the cultivars differ morphologically among themselves. Our results validate the information recorded by Huerta-Galván (2018), that the interchange of plants or seeds in monocultures of *A. maximiliana* came mainly from the forestry systems within the raicilla de la sierra region and that the monocultures themselves are very recent systems. This can be inferred from their morphological similarity, where 60% (CC) to 90% (EM, RS) of individuals in monocultures can be morphologically identified as belonging to other *in situ* managed populations (**Figure 4**). This can also be inferred by genetic similarity, where most of the monocultures and managed populations (except for EM and LV) tended to form a cluster (**Figure 7**). However, morphological and genetic variation is wider among populations (36 and 38%, respectively) than among categories (19.8 and 1%, respectively). Consequently, our study did not find significant divergence related to management.

Different factors are involved in morphological divergences compared to genetic structure, since these were not related to each other. A correlation between phenotypic variation and neutral genetic markers may occur if the same neutral processes (e.g., gene flow and genetic drift) have historically affected the spatial structure of both types of variation (e.g., Hipp et al., 2018; Rodríguez-Gómez et al., 2018). Another aspect is that populations usually exhibit contrasting spatial patterns for different genetic markers because the mutation rates, combined with the evolutionary forces acting on them, are not the same (Habel et al., 2015). It is therefore necessary to include more genetic markers to support these results. On the other hand, isolation by environment (six variables related to precipitation; $\beta = 0.32$, $p = 0.007$) and not geographic distance is one of the mechanisms that structure the populations of *A. maximiliana*. This means that, to some extent, populations are genetically isolated as a result of different precipitation conditions (e.g., annual precipitation varies from 785 to1312 mm, precipitation of warmest quarter varies from 239 to 795 mm; cf. Trejo et al., 2016). Local precipitation is related to the regeneration and colonization of plant populations (Jiang et al., 2019). Thus, the influence of precipitation on demographic dynamics may explain its effect on the population genetic structure as revealed by neutral markers (Jiang et al., 2019), where demographic parameters mediate ecological–evolutionary interactions via adaptive and non-adaptive mechanisms (Lowe et al., 2017). Gene flow could be reduced by the poor establishment success of immigrant seeds. Moreover, a decreased chance of outcrossing conditions may compound these gene flow reductions and accelerate the genetic fixation rate in populations (Jiang et al., 2019). This phenomenon not only appears in adaptive loci but could also extend to the whole genome via genetic drift caused by selective sweeps (Nosil et al., 2009). Since morphological variation is not explained by any of these isolation processes, this could be a sign to discard phenotypic plasticity (cf. Rodríguez-Gómez et al., 2018) in the evaluated traits. However, other approaches, such as wide genome scans or/and common garden experiments, are necessary to confirm these interpretations (de Villemereuil et al., 2018). Finally, it is important to mention that Scheinvar et al. (2017) modeled environmental niche changes in *A. maximiliana* and other congeners, and found that *A. maximiliana* could have undergone an extension of its environmental niche during the Last Interglacial period (LIG, 130 Kyr), followed by a reduction in the Last Glacial Maximum period (LGM, 31Kyr). Thus, past extension-contraction cycles in the demographic history of *A. maximiliana* could result in decreases in the levels of genetic diversity and increases in genetic structure (e.g., *A. lechuguilla* Torr., Scheinvar et al., 2017). This hypothesis should also be tested using phylogeographic approaches and with other types of molecular markers in order to support the main findings of this study.

The hypothetical presence of two varieties of *A. maximiliana* was not clearly reflected in our data. The two northern populations collected for this study (CN and EP, **Figure 2**) occupy a region in which *A. maximiliana* var. *katherinae* is probably distributed (Gentry, 1982; Scheinvar, 2018). Both populations are morphologically similar or close in the multivariate space (**Figures 5A,B**) to LH and RS, which are not characterized by larger rosettes or greater thorniness, compared to *A. maximiliana* var. *maximiliana* as Gentry (1982) described it. The population EP is differentiated genetically (**Figures 6**, **7**) but we did

not obtain a sufficient number of microsatellite amplifications from CN, and therefore do not know its genetic components and cannot determine its location in the genetic multivariate space. The population EC, which is geographically close to EP and also within the hypothetical distribution area of *A. maximiliana* var. *katherinae*, is grouped as a genetic entity similar to managed and cultivated populations of the raicilla de la sierra region. This suggests that an important gene flow is still perceptible between the two hypothetical varieties. We consider that further research with a sample design focused on this taxonomic approach is necessary to test this hypothesis.

Finally, *A. maximiliana* maintains the highest genetic structure reported to date among species of *Agave* (Φ_{ST} = 0.4, **Table 1**; Eguiarte et al., 2013). This is perhaps a reflection of the type and extension of the distribution areas of their populations; this can be perceived in other congeners, in which a discontinuous distribution is associated with high genetic structure, in contrast to those of continuous distribution (e.g., *A. parryi* vs. *A. palmeri*, Lindsay et al., 2018). Furthermore, together with the results of genetic diversity and inbreeding, this reinforces the need to investigate the reproductive and pollination system of this species. The higher the genetic structure, the greater the probability that populations will reflect changes as independent evolutionary units (Slatkin, 1994), whether through genetic drift or selection.

Perspectives and Recommendations

Based on the evidence presented in this study, we consider that traditional management represented by forestry systems of *A. maximiliana* has been a sustainable option in terms of genetic germplasm conservation. However, the current threatening scenario of the use of *Agave maximiliana* as a raw material for agroindustrial production of raicilla is disheartening. Of particular concern are the low genetic diversity, high genetic structure and presence of inbreeding in this species. This fact should be taken into account in management when translocating seeds or plants from one site to another, due to possible processes of local adaptation that could preclude the optimum development of plants outside of their populations of origin (Habel et al., 2015). We perceive this as a weakness of the species, if the forms of management move to more intensive practices, which is highly probable in the light of the recent denomination of origin of raicilla (Diario Oficial de la Federación, 2019). However, recovery of this situation is possible based on promotion of these forestry systems that still maintain *A. maximiliana*.

Our research approach had restrictions related to the identification of local adaptation processes and phenotypic plasticity. As a result, we cannot make direct conclusions about the extent to which the wild and managed populations (either *in situ* or monocultures) of *A. maximiliana* are locally adapted. In addition, adoption of new approaches regarding developmental plasticity, variation in quantitative trait loci related to natural and artificial selection or niche construction theory (Chen et al., 2017; Piperno, 2017) could complete the understanding of diversification *versus* domestication contexts (Pickersgill, 2018). Despite this, we consider that the general patterns of intraspecific diversity in *in situ* contexts of management practices are still poorly documented, and that this study constitutes a valuable base on which to establish more specific research questions to further the understanding of the origin of domestication processes.

Finally, we recognize in this case study that, beyond the improvement of technical options to multiply the number of plants available per year, it would be more meaningful to consider broader aspects of sustainability and conservation. This would generate benefits not only for the raicilla production chain, but also for the range of environmental goods and services their habitat provides.

DATA AVAILABILITY STATEMENT

The datasets generated and analyzed for this study can be found in the Mendeley Data, http://dx.doi.org/10.17632/3379nnccmy.4.

AUTHOR CONTRIBUTIONS

DC-T and OV-P conceived and designed the study, partially organized fieldwork and carried out some final statistical analysis. SA-R and LV-S carried out the collection and first analysis of phenotypic and genetic data, respectively, as part of their bachelor theses. JP-A gave input on the study design and carried out some final statistical analysis and design of figures. JM-S led laboratory work. OH-G described the management practices which were the basis of the hypothesis of this study, he also partially organized fieldwork. All authors reviewed and approved the final manuscript.

FUNDING

The research was funded by Consejo Nacional de Ciencia y Tecnología (CONACyT) through the Laboratorio Nacional de Identificación y Caracterización Vegetal (LaniVeg) research grants (Nos. 261099 and 272045).

ACKNOWLEDGMENTS

We thank P. Carrillo-Reyes, A. Dávalos-Martínez, D. Figueroa-Martínez, A. Zabalgoitia, J. Rendón, Z. López-Dellamary, K. Rostro, P. Velázquez-Ríos, J. González-Gallegos, and A. Castro-Castro for field assistance, and P. Zamora for laboratory assistance. P. Carrillo-Reyes, A. Castro-Castro, D. Figueroa-Martínez, and A. Zabalgoitia for their contributions in floristic information. We also thank C. J. Figueredo-Urbina and Anwar Medina-Villareal for their help and advice in the analysis of morphological data. Thanks also go to the raicilla de la sierra producers, Gerardo Peña, Juan Lomelí, Aristeo

Macedo, Manuel Rodríguez, Benito Salcedo, Manuel Salcedo, Gelo Aguirre, Don Delfino, the families Arrizon, and Segura for their kindness and permission to work on their properties, and to the communities of El Palmito, Sinaloa, and Canelas, Durango for their kind welcome.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00817/full#supplementary-material

REFERENCES

Aguirre-Dugua, X., and Eguiarte, L. E. (2013). Genetic diversity, conservation and sustainable use of wild *Agave cupreata* and *Agave potatorum* extracted for mezcal production in Mexico. *J. Arid Environ.* 90, 36–44. doi: 10.1016/j.jaridenv.2012.10.018

Anderson, M. J. (ed.). (2017). "Permutational multivariate analysis of variance (PERMANOVA)," in *Wiley StatsRef: Statistics Reference Online*, (Hoboken, NJ: Wiley), 1–15. doi: 10.1002/9781118445112.stat07841

Anderson, M. J., Gorley, R., and Clarke, R. (2008). *PERMANOVA+para PRIMER: Guía Para el Programa y Métodos Estadísticos. PRIMER-E*. Plymouth: Massey University Press.

Arias, R., Martínez-Castillo, J., Sobolev, V., Blancarte-Jasso, N., Simpson, S., Ballard, L., et al. (2015). Development of a large set of microsatellite markers in zapote mamey (*Pouteria sapota* (Jacq.) HE Moore & Stearn) and their potential use in the study of the species. *Molecules* 20, 11400–11417. doi: 10.3390/molecules200611400

Arita, H. (1991). Spatial segregation in long-nosed bats, *Leptonycteris nivalis* and *Leptonycteris curasoe*, in Mexico. *Am. Soc. Mammal.* 72, 706–714. doi: 10.2307/1381831

Arizaga, S., Ezcurra, E., Peters, E., de Arellano, F. R., and Vega, E. (2000). Pollination ecology of *Agave macroacantha* (Agavaceae) in a Mexican tropical desert. II. The role of pollinators. *Am. J. Bot.* 87, 1011–1017. doi: 10.2307/2657001

Arroyo-Cabrales, J., Hollander, R. R., and Jones, J. K. J. (1982). *Choeronycteris mexicana*. *Mammal. Species* 291, 1–5.

Blancas, J., Casas, A., Lira, R., and Caballero, J. (2009). Traditional management and morphological patterns of *Myrtillocactus schenckii* (Cactaceae) in the Tehuacán Valley, Central Mexico. *Econ. Bot.* 63, 375–387. doi: 10.1007/s12231-009-9095-2

Blancas, J., Casas, A., Pérez-Salicrup, D., Caballero, J., and Vega, E. (2013). Ecological and socio-cultural factors influencing plant management in Náhuatl communities of the Tehuacán Valley, Mexico. *J. Ethnobiol. Ethnomed.* 9:39. doi: 10.1186/1746-4269-9-39

Blancas, J., Casas, A., Rangel-Landa, S., Moreno-Calles, A., Torres, I., Pérez-Negrón, E., et al. (2010). Plant management in the Tehuacán-Cuicatlán Valley, Mexico1. *Econ. Bot.* 64, 287–302. doi: 10.1007/s12231-010-9133-0

Bowen, S., and Gerritsen, P. R. W. (2007). Reverse leasing and power dynamics among blue Agave farmers in western Mexico. *Agric. Hum. Values* 24, 473–488. doi: 10.1007/s10460-007-9088-7

Bowen, S., and Valenzuela-Zapata, A. G. (2009). Geographical indications, terroir, and socioeconomic and ecological sustainability: the case of tequila. *J. Rural Stud.* 25, 108–119. doi: 10.1016/j.jrurstud.2008.07.003

Byers, C., Maughan, P. J., Clouse, J., and Stewart, J. R. (2014). Microsatellite primers in *Agave utahensis* (Asparagaceae), a keystone species in the Mojave Desert and Colorado Plateau. *Appl. Plant Sci.* 2:1400047. doi: 10.3732/apps.1400047

Casas, A., Caballero, J., Valiente-Banuet, A., Soriano, J. A., and Dávila, P. (1999). Morphological variation and the process of domestication of *Stenocereus stellatus* (Cactaceae) in Central Mexico. *Am. J. Bot.* 86, 522–533. doi: 10.2307/2656813

Casas, A., Cruse-Sanders, J., Morales, E., Otero-Arnaiz, A., and Valiente-Banuet, A. (2006). Maintenance of phenotypic and genotypic diversity in managed populations of *Stenocereus stellatus* (Cactaceae) by indigenous peoples in Central Mexico. *Biodivers. Conserv.* 15, 879–898. doi: 10.1007/s10531-004-2934-7

Casas, A., Parra-Rondinel, F., Aguirre-Dugua, X., Rangel-Landa, S., Blancas, J., Vallejo, M., et al. (2017). "Manejo y domesticación de plantas en mesoamérica," in *Domesticación en el Continente Americano*, eds A. Casas, J. Torres-Guevara, and F. Parra (México: UNAM), 69–102.

Chapuis, M. P., and Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol.* 24, 621–631. doi: 10.1093/molbev/msl191

Charlesworth, D., and Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annu. Rev. Ecol. Syst.* 18, 237–268. doi: 10.1146/annurev.es.18.110187.001321

Chen, Y. H., Shapiro, L. R., Benrey, B., and Cibrián-jaramillo, A. (2017). Back to the origin: in situ studies are needed to understand selection during crop diversification. *Front. Ecol. Evol.* 5:125. doi: 10.3389/fevo.2017.00125

Chybicki, I. J., and Burczyk, J. (2009). Simultaneous estimation of null alleles and inbreeding coefficients. *J. Hered.* 100, 106–113. doi: 10.1093/jhered/esn088

Clunies-Ross, T. (1995). Mangolds, manure and mixtures. The importance of crop diversity on British farms. *Ecologist* 25, 181–187.

Colunga-GarcíaMarín, P., and Zizumbo-Villarreal, D. (2007). Tequila and other *Agave* spirits from west-central Mexico: current germplasm diversity, conservation and origin. *Biodivers. Conserv.* 16, 1653–1667. doi: 10.1007/s10531-006-9031-z

Contreras-Negrete, G., Ruíz-Durán, M. E., Cabrera-Toledo, D., Casas, A., Vargas, O., and Parra, F. (2015). Genetic diversity and structure of wild and managed populations of *Polaskia chende* (Cactaceae) in the Tehuacán-Cuicatlán Valley, Central Mexico: insights from SSR and allozyme markers. *Genet. Resour. Crop Evol.* 62, 85–101. doi: 10.1007/s10722-014-0137-y

Cruse-Sanders, J. M., Parker, K. C., Friar, E. A., Huang, D. I., Mashayekhi, S., Prince, L. M., et al. (2013). Managing diversity: domestication and gene flow in *Stenocereus stellatus* Riccob.(Cactaceae) in Mexico. *Ecol. Evol.* 3, 1340–1355. doi: 10.1002/ece3.524

Daniel, W. (2006). *Bioestadística*, 4th Edn. México: Editorial Limusa.

de Villemereuil, P., Mouterde, M., Gaggiotti, O. E., and Till-bottraud, I. (2018). Patterns of phenotypic plasticity and local adaptation in the wide elevation range of the alpine plant *Arabis alpina*. *J. Ecol.* 106, 1952–1971. doi: 10.1111/1365-2745.12955

Delgado-Lemus, A., Casas, A., and Téllez, O. (2014). Distribution, abundance and traditional management of *Agave potatorum* in the Tehuacán Valley, Mexico: bases for sustainable use of non-timber forest products. *J. Ethnobiol. Ethnomed.* 10, 1–12. doi: 10.1186/1746-4269-10-63

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* 39, 1–22.

Diario Oficial de la Federación (2019). *Declaración General de Protección de la Denominación de Origen Raicilla*. Available online at: https://dof.gob.mx/nota_detalle.php?codigo=5564454&fecha=28/06/2019&print=true (accessed June 1, 2020).

Doyle, J. (1991). "DNA protocols for plants," in *Molecular Techniques in Taxonomy*, eds G. Hewitt, A. Jhonston, and J. Young (Brussels: NATO), 283–293. doi: 10.1007/978-3-642-83962-7_18

Eguiarte, L. E., Aguirre-Planter, E., Aguirre, X., Colíni, R., González, A., Rocha, M., et al. (2013). From isozymes to genomics: population genetics and conservation of *Agave* in México. *Bot. Rev.* 79, 483–506. doi: 10.1007/s12229-013-9123-x

Estrella-Ruiz, P. (2008). *Efecto de la Explotación Humana en la Biología de la Polinización de Agave salmiana y Agave potatorum en el Valle de Tehuacán-Cuicatlán*. Bachelor thesis, Universidad Nacional Autónoma de México, Mexico.

Félix-Valdez, L. I., Vargas-Ponce, O., Cabrera-Toledo, D., Casas, A., Cibrian-Jaramillo, A., and de la Cruz-Larios, L. (2016). Effects of traditional management for mescal production on the diversity and genetic structure of *Agave potatorum* (Asparagaceae) in central Mexico. *Genet. Resour. Crop Evol.* 63, 1255–1271. doi: 10.1007/s10722-015-0315-6

Figueredo, C. J., Casas, A., Colunga-GarcíaMarín, P., Nassar, J. M., and González-Rodríguez, A. (2014). Morphological variation, management and domestication of "maguey alto" (*Agave inaequidens*) and "maguey manso" (*A. hookeri*) in

Michoacán, México. *J. Ethnobiol. Ethnomed.* 10, 1–12. doi: 10.1186/1746-4269-10-66

Figueredo, C. J., Casas, A., González-Rodríguez, A., Nassar, J. M., Colunga-GarcíaMarín, P., and Rocha-Ramírez, V. (2015). Genetic structure of coexisting wild and managed *Agave* populations: implications for the evolution of plants under domestication. *AoB Plants* 7:lv114. doi: 10.1093/aobpla/plv114

Figueredo-Urbina, C. J., Casas, A., and Torres-García, I. (2017). Morphological and genetic divergence between *Agave inaequidens*, *A. cupreata* and the domesticated *A. hookeri*. Analysis of their evolutionary relationships. *PLoS One* 12:e0187260. doi: 10.1371/journal.pone.0187260

Flores-Abreu, I. N., Trejo-Salazar, R. E., Sánchez-Reyes, L. L., Good, S. V., Magallón, S., García-Mendoza, A., et al. (2019). Tempo and mode in coevolution of *Agave* sensu lato (Agavoideae, Asparagaceae) and its bat pollinators, Glossophaginae (Phyllostomidae). *Mol. Phylogenet. Evol.* 133, 176–188. doi: 10.1016/j.ympev.2019.01.004

Frankham, R., Ballou, J., and Briscoe, D. (2006). *Introduction to Conservation Genetics*, 6th Edn. Cambridge: Cambridge University press.

Gentry, H. S. (1982). *Agaves of Continental North America*, 1st Edn. Tucson, AZ: University of Arizona Press.

González, A. (2005). *Biología Reproductiva y Genética de Poblaciones de Agave garciaemendozae*. Bachelor thesis, Universidad Nacional Auntónoma de México, México.

González-Elizondo, M., Galván-Villanueva, R., López-Enríquez, I., Reséndiz-Rojas, L., and Gonzalez-Elizondo, M. S. (2009). *Agaves -Magueyes, Lechuguillas y Noas- del Estado de Durango y sus Alrededores*. México: CONABIO-Instituto Politécnico Nacional.

Goslee, S. C., and Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *J. Stat. Softw.* 22, 1–19.

Guillén, S., Terrazas, T., De la Barrera, E., and Casas, A. (2011). Germination differentiation patterns of wild and domesticated columnar cacti in a gradient of artificial selection intensity. *Genet. Resour. Crop Evol.* 58, 409–423. doi: 10.1007/s10722-010-9586-0

Habel, J. A. N. C., Zachos, F. E., Dapporto, L., Rödder, D., Radespiel, U. T. E., Tellier, A., et al. (2015). Population genetics revisited – towards a multidisciplinary research field. *Biol. J. Linn. Soc.* 115, 1–12. doi: 10.1111/bij.12481

Hammer, Ø., Harper, D., and Ryan, P. (2001). PAST: paleontological statistics software package for education and data analysis. *Palaeontol. Electronica* 4, 1–9.

Hedrick, P. (2005). *Genetics of Populations*, 3rd Edn. Sudbury, MA: Jones and Bartlett publishers.

Hijmans, R. (2015). *Geosphere: Spherical Trigonometry. R Package*.

Hipp, A. L., Manos, P. S., Gonz, A., Hahn, M., Kaproth, M., Mcvay, J. D., et al. (2018). Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. *New Phytol.* 217, 439–452. doi: 10.1111/nph.14773

Horner, M. A., Fleming, T. H., and Sahley, C. T. (1998). Foraging behaviour and energetics of a nectar-feeding bat, *Leptonycteris curasoae* (Chiroptera: Phyllostomidae). *Zool. Soc. Lond.* 244, 575–586. doi: 10.1111/j.1469-7998.1998.tb00062.x

Howell, D. J., and Roth, B. S. (1981). Sexual reproduction in agaves: the benefits of bats; the cost of semelparous advertising. *Ecology* 62, 1–7. doi: 10.2307/1936660

Huerta-Galván, O. (2018). *Uso y Manejo de Agave maximiliana: Producción de Raicilla en el Municipio de Mascota, Jalisco*. Bachelor thesis, Universidad de Guadalajara, Mexico.

Illsey, C., Vega, E., Pisanty, I., Tlacotempa, A., García, P., Morales, P., et al. (2007). "Maguey papalote: hacia el manejo campesino sustentable de un recurso colectivo en el trópico seco de Guerrero, México," in *En lo Ancestral hay Futuro: Del Tequila, Los Mezcales y Otros Agaves*, eds P. Colunga-GarcíaMarín, A. Larqué, L. Eguiarte, and D. Zizumbo-Villareal (México: CICY, CONACYT, CONABIO, SEMARNAT).

Jiang, S., Luo, M., Gao, R., Zhang, W., Yang, Y., and Li, Y. (2019). Isolation-by-environment as a driver of genetic differentiation among populations of the only broad-leaved evergreen shrub *Ammopiptanthus mongolicus* in Asian temperate deserts. *Sci. Rep.* 9:12008. doi: 10.1038/s41598-019-48472-y

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11:94. doi: 10.1186/1471-2156-11-94

Leclert, L., Gerritsen, P. R. W., and van der Meulen, H. (2010). "*Agave azul*: crisis cíclicas y las posibilidades para la planeación del cultivo en el estado de Jalisco," in *Agave azul, Sociedad y Medio Ambiente*, eds P. R. W. Gerritsen and L. M. Martínez-Rivera (México: DERN-IMECBIO, CUCSUR-UDG), 19–35.

Legendre, P., Lapointe, F., and Casgrain, P. (1994). Modeling brain evolution from behavior: a permutational regression approach. *Evolution* 48, 1487–1499. doi: 10.1111/j.1558-5646.1994.tb02191.x

Legendre, P., and Legendre, L. (1998). *Numerical Ecology*, 2nd Edn. Amsterdam: Elsevier.

León, A. (2013). *Aspectos de la Fenología, Visitantes Florales y Polinización de Agave inaequidens Koch ssp. inaequidens (Agavaceae) en el Estado de Michoacán*. Bachelor Thesis, Universidad Michoacana de San Nicolás de Hidalgo, Morelia.

Lichstein, J. (2007). Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecol.* 188, 117–131. doi: 10.1007/s11258-006-9126-3

Lindsay, D., Edwards, C., Jung, M., Bailey, P., and Lance, R. (2012). Novel microsatellite loci for *Agave parryi* and cross-amplification in *Agave palmeri* (Agavaceae). *J. Bot.* 99, 295–297. doi: 10.3732/ajb.1200033

Lindsay, D. L., Swift, J. F., Lance, R. F., and Edwards, C. E. (2018). A comparison of patterns of genetic structure in two co-occurring *Agave* species (Asparagaceae) that differ in the patchiness of their geographical distributions and cultivation histories. *Bot. J. Linn. Soc.* 186, 361–373. doi: 10.1093/botlinnean/box099

Lowe, W., Kovach, R., and Allendorf, F. W. (2017). Population genetics and demography unite ecology and evolution. *Trends Ecol. Evol.* 32, 141–152. doi: 10.1016/j.tree.2016.12.002

Luikart, G., and Cornuet, J.-M. (1998). Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conserv. Biol.* 12, 228–237. doi: 10.1111/j.1523-1739.1998.96388.x

McVaugh, R. (1989). *Flora Novo-Galiciana: Bromeliaceae to Dioscoreaceae*. Ann Arbor, MI: University of Michigan Herbarium.

Medellín, R., Rivero, M., Ibarra, A., De la Torre, J., Gonzalez-Terrazas, T. P., Torres-Knoop, L., et al. (2018). Follow me: foraging distances of *Leptonycteris yerbabuenae* (Chiroptera: Phyllostomidae) in Sonora determined by fluorescent powder. *J. Mammal.* 99, 306–311. doi: 10.1093/jmammal/gyy016

Medina-Villarreal, A., and González-Astorga, J. (2016). Morphometric and geographical variation in the *Ceratozamia mexicana* Brongn. (Zamiaceae) complex: evolutionary and taxonomic implications. *Biol. J. Linn. Soc.* 119, 213–233. doi: 10.1111/bij.12806

Molina-Freaner, F., and Eguiarte, L. E. (2003). The pollination biology of two paniculate agaves (Agavaceae) from northwestern Mexico: contrasting roles of bats as pollinators. *Am. J. Bot.* 90, 1016–1024. doi: 10.3732/ajb.90.7.1016

Moreno-Calles, A. I., Toledo, V. M., and Casas, A. (2013). Los sistemas agroforestales tradicionales de México: una aproximación biocultural. *Bot. Sci.* 91, 375–398. doi: 10.17129/botsci.419

Morrone, J. J., Escalante, T., and Rodríguez-Tapia, G. (2017). Mexican biogeographic provinces: map and shapefiles. *Zootaxa* 4277, 8–11. doi: 10.11646/zootaxa.4277.2.8

Nair, P. K. R. (1985). Classification of agroforestry systems. *Agroforestry Syst.* 3, 97–128. doi: 10.1007/bf00122638

Nei, M., and Chakraborty, R. (1973). Genetic distance and electrophoretic identity of proteins between taxa. *J. Mol. Evol.* 2, 323–328. doi: 10.1007/bf01654100

Neto, E. M. F. L., de Oliveira, I. F., Britto, F. B., and de Albuquerque, U. P. (2013). Traditional knowledge, genetic and morphological diversity in populations of *Spondias tuberosa* Arruda (Anacardiaceae). *Genet. Resour. Crop Evol.* 60, 1389–1406. doi: 10.1007/s10722-012-9928-1

Nosil, P., Funk, D., and Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* 18, 375–402. doi: 10.1111/j.1365-294X.2008.03946.x

Nybom, H. (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol. Ecol.* 13, 1143–1155. doi: 10.1111/j.1365-294X.2004.02141.x

Ortega-García, S., Arroyo-Cabrales, J., Guevara, L., Lindig-Cisneros, R., Martínez-Meyer, E., Vega, E., et al. (2017). The thermal niche of Neotropical nectar feeding bats: its evolution and application to predict responses to global warming. *Ecol. Evol.* 7, 6691–6701. doi: 10.1002/ece3.3171

Otero-Arnaiz, A., Casas, A., Hamrick, J., and Cruse-Sanders, J. (2005). Genetic variation and evolution of *Polaskia chichipe* (Cactaceae) under domestication in the Tehuacán Valley, central Mexico. *Mol. Ecol.* 14, 1603–1611. doi: 10.1111/j.1365-294X.2005.02494.x

Parker, K. C., Hamrick, J. L., Hodgson, W. C., Trapnell, D. W., Parker, A. J., and Kuzoff, R. K. (2007). Genetic consequences of pre-Columbian cultivation for *Agave murpheyi* and *A. delamateri* (Agavaceae). *Am. J. Bot.* 94, 1479–1490. doi: 10.3732/ajb.94.9.1479

Parker, K. C., Trapnell, D. W., Hamrick, J. L., and Hodgson, W. C. (2014). Genetic and morphological contrasts between wild and anthropogenic populations of *Agave parryi* var. *huachucensis* in south-eastern Arizona. *Ann. Bot.* 113, 939–952. doi: 10.1093/aob/mcu016

Parker, K. C., Trapnell, D. W., Hamrick, J. L., Hodgson, W. C., and Parker, A. J. (2010). Inferring ancient *Agave* cultivation practices from contemporary genetic patterns. *Mol. Ecol.* 19, 1622–1637. doi: 10.1111/j.1365-294X.2010.04593.x

Parra, F., Blancas, J. J., and Casas, A. (2012). Landscape management and domestication of *Stenocereus pruinosus* (Cactaceae) in the Tehuacán Valley: human guided selection and gene flow. *J. Ethnobiol. Ethnomed.* 8:32. doi: 10.1186/1746-4269-8-32

Parra, F., Casas, A., Peñaloza-Ramírez, J. M., Cortés-Palomec, A. C., Rocha-Ramírez, V., and González-Rodríguez, A. (2010). Evolution under domestication: ongoing artificial selection and divergence of wild and managed *Stenocereus pruinosus* (Cactaceae) populations in the Tehuacán Valley, Mexico. *Ann. Bot.* 106, 483–496. doi: 10.1093/aob/mcq143

Peakall, R., and Smouse, P. E. (2012). GenAlEx tutorials-part 2: genetic distance and analysis of molecular variance (AMOVA). *Bioinformatics* 28, 2537–2539.

Peakall, R. O. D., and Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6, 288–295.

Pickersgill, B. (2018). Parallel vs. convergent evolution in domestication and diversification of crops in the Americas. *Front. Ecol. Evol.* 6:56. doi: 10.3389/fevo.2018.00056

Piperno, D. R. (2017). Assessing elements of an extended evolutionary synthesis for plant domestication and agricultural origin research. *Proc. Natl. Acad. Sci. U.S.A.* 114, 6429–6437. doi: 10.1073/pnas.1703658114

R Development Core Team (2012). *R: a Language and Environment for Statistical Computing*. Vienna: R foundation for Statistical Computing.

Reed, D. H., and Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conserv. Biol.* 17, 230–237. doi: 10.1046/j.1523-1739.2003.01236.x

Rivera-Lugo, M., García-Mendoza, A., Simpson, J., Solano, E., and Gil-Vega, K. (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 Syst. Evol.* 304, 969–979. doi: 10.1007/s00606-018-1525-0

Rocha, M., Good-Ávila, S., Molina-Freaner, F., Arita, H., Castillo, A., García-Mendoza, A., et al. (2006). Pollination biology and adaptive radiation of Agavaceae, with special emphasis on the genus *Agave*. *Aliso* 22, 329–344. doi: 10.5642/aliso.20062201.27

Rodríguez-Gómez, F., Oyama, K., Ochoa-Orozco, M., Mendoza-Cuenca, L., Gaytán-Legaria, R., and González-Rodríguez, A. (2018). Phylogeography and climate-associated morphological variation in the endemic white oak *Quercus deserticola* (Fagaceae) along the Trans-Mexican Volcanic Belt. *Botany* 133, 121–133. doi: 10.1139/cjb-2017-0116

Sanguinetti, C. J., Neto, E. D., and Simpson, A. J. G. (1994). Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. *Biotechniques* 17, 915–918.

Scheinvar, E. (2018). *Filogeografía de Agave lechuguilla y Patrones de Distribución de Agave en México*. Ph.D. thesis Universidad Nacional Autónoma de México, México.

Scheinvar, E., Gámez, N., Castellanos-Morales, G., Aguirre-Planter, E., and Eguiarte, L. E. (2017). Neogene and pleistocene history of *Agave lechuguilla* in the Chihuahuan Desert. *J. Biogeogr.* 44, 322–334. doi: 10.1111/jbi.12851

Slatkin, M. (1994). "Gene flow and population structure," in *Ecological Genetics*, ed. L. Real (Princeton, NJ: Princeton University Press), 3–17.

Torres, I., Blancas, J., León, A., and Casas, A. (2015a). TEK, local perceptions of risk, and diversity of management practices of *Agave inaequidens* in Michoacán, Mexico. *J. Ethnobiol. Ethnomed.* 11:61. doi: 10.1186/s13002-015-0043-1

Torres, I., Casas, A., Vega, E., Martínez-Ramos, M., and Delgado-Lemus, A. (2015b). Population dynamics and sustainable management of mescal agaves in central Mexico: *Agave potatorum* in the Tehuacán-Cuicatlán Valley. *Econ. Bot.* 69, 26–41. doi: 10.1007/s12231-014-9295-2

Torres-García, I., Rendón-Sandoval, F. J., Blancas, J., Casas, A., and Moreno-Calles, I. (2019). The genus *Agave* in agroforestry systems of Mexico. El género Agave en los sistemas agroforestales de México. *Bot. Sci.* 97, 261–288. doi: 10.17129/botsci.2202

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

Trejo, L., Alvarado-Cárdenas, L. O., Scheinvar, E., and Eguiarte, L. E. (2016). Population genetic analysis and bioclimatic modeling in *Agave striata* in the Chihuahuan Desert indicate higher genetic variation and lower differentiation in drier and more variable environments. *Am. J. Bot.* 103, 1020–1029.

Trejo, L., Limones, V., Peña, G., Scheinvar, E., Vargas-Ponce, O., Zizumbo-Villarreal, D., et al. (2018). Genetic variation and relationships among agaves related to the production of Tequila and Mezcal in Jalisco. *Ind. Crops Prod.* 125, 140–149. doi: 10.1016/j.indcrop.2018.08.072

Valdivia-Mares, L., Rodríguez Zaragoza, F. A., Sánchez-González, J., and Vargas-Ponce, O. (2016). Phenology, agronomic and nutritional potential of three wild husk tomato species (Physalis, Solanaceae) from Mexico. *Sci. Hortic.* 200, 83–94. doi: 10.1016/j.scienta.2016.01.005

Valenzuela-Zapata, A. G. (2015). "Raicillas, mezcales artesanales de Jalisco," in *Agua de las Verdes Matas, Tequila y Mezcal*, eds J. L. Vera-Cortés and R. Fernández (México: CONACULTA, INA, ARTES DE MÉXICO), 131–146.

Vargas-Ponce, O., Zizumbo-Villarreal, D., and Colunga-GarcíaMarin, P. (2007). In situ diversity and maintenance of traditional *Agave* landraces used in spirits production in west-central Mexico. *Econ. Bot.* 61, 362–375.

Vargas-Ponce, O., Zizumbo-Villarreal, D., Martínez-Castillo, J., Coello-Coello, J., and Colunga-GarcíaMarín, P. (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*. *Am. J. Bot.* 96, 448–457. doi: 10.3732/ajb.0800176

Vázquez-García, J., Cházaro-Basañez, M., Hernández-Vera, G., Flores, E., and Vargas-Rodriguez, Y. (2007). *Agaves del Occidente de México*, 1st Edn. México: Universidad de Guadalajara.

Vellvé, R. (1992). *Saving the Seed: Genetic Diversity and European Agriculture*. London: Earthscan. doi: 10.4324/9781315066837

Wang, I. (2013). Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution* 67, 3403–3411. doi: 10.1111/evo.12134

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19, 395–420.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 23 June 2020
doi: 10.3389/fpls.2020.00863

Phylogeography and Genetic Diversity in a Southern North American Desert: *Agave kerchovei* From the Tehuacán-Cuicatlán Valley, Mexico

Erika Aguirre-Planter[1], J. Gilberto Parra-Leyva[1], Santiago Ramírez-Barahona[2], Enrique Scheinvar[1], Rafael Lira-Saade[3] and Luis E. Eguiarte[1]**

[1] Laboratorio de Evolución Molecular y Experimental, Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico, [2] Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City, Mexico, [3] Unidad de Biotecnología y Prototipos, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Mexico City, Mexico

OPEN ACCESS

Edited by:
Angelica Cibrian-Jaramillo, Centro de Investigación y Estudios Avanzados, Instituto Politécnico Nacional de México (CINVESTAV), Mexico

Reviewed by:
Sara V. Good, University of Winnipeg, Canada
Marie-Stéphanie Samain, Instituto de Ecología (INECOL), Mexico
Tania Hernandez-Hernandez, National Laboratory of Genomics for Biodiversity, Center for Research and Advanced Studies of the National Polytechnic Institute, Mexico

***Correspondence:**
Rafael Lira-Saade
rlira@unam.mx
Luis E. Eguiarte
fruns@unam.mx;
fruns@servidor.unam.mx

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *14 January 2020*
Accepted: *27 May 2020*
Published: *23 June 2020*

Citation:
Aguirre-Planter E, Parra-Leyva JG, Ramírez-Barahona S, Scheinvar E, Lira-Saade R and Eguiarte LE (2020) Phylogeography and Genetic Diversity in a Southern North American Desert: Agave kerchovei From the Tehuacán-Cuicatlán Valley, Mexico. Front. Plant Sci. 11:863. doi: 10.3389/fpls.2020.00863

The Tehuacán-Cuicatlán Valley, located at the southeast of the state of Puebla and the northeast of the state of Oaxaca in Central Mexico, south of the Trans-Mexican Volcanic Belt (TMVB), is of particular interest for understanding the evolutionary dynamics of arid and semi-arid environments, being one of the main reservoirs of biological diversity for the arid zones of North America, including the highest diversity of Agavaceae worldwide and high levels of endemism. Studying in detail the phylogeography, environmental history and population genetics of representative species will hopefully shed light on the evolutionary and ecological dynamics that generated the tremendous biodiversity and endemism of this important region in Mexico. We sequenced three non-coding regions of chloroplast genome of *Agave kerchovei*, a representative species of the Tehuacán Valley, generating 2,188 bp from 128 individuals sampled from eight populations throughout the species range. We used this data set to (i) characterize the levels of genetic diversity and genetic structure in *A. kerchovei*; (ii) predict the distribution of *A. kerchovei* for the present day, and to reconstruct the past geographical history of the species by constructing ecological niche models (ENM); and (iii) compare the levels of diversity in this species with those estimated for the widely distributed *Agave lechuguilla*. *Agave kerchovei* has high levels of total chloroplast genetic variation (*Hd* = 0.718), especially considering that it is a species with a very restricted distribution. However, intrapopulation diversity is low (zero in some populations), and genetic structure is high (F_{ST} = 0.928, G_{ST} = 0.824), which can be expected for endemic species with isolated populations. Our data suggest that Pleistocene glacial cycles have played an important role in the distribution of *A. kerchovei,* where the climatic variability of the region – likely associated with its topographic complexity – had a significant effect on the levels of genetic diversity and population dynamics, while the potential distribution of the species

seems to be stable since the middle Holocene (6 kya). We conclude that in *A. kerchovei* there is a core group of populations in the Tehuacán Valley, and peripheric populations that appear to be evolving independently and thus the species is fundamentally an endemic species from the Tehuacán Valley while the populations outside the Valley appear to be in the process of incipient speciation.

Keywords: Tehuacán-Cuicatlán, *Agave*, species distribution models, incipient speciation, population genetics, Pleistocene, endemic

INTRODUCTION

The Pleistocene glacial periods have been major factors influencing the geographical distribution, demographic dynamics and patterns of genetic diversity of many species (Haffer and Prance, 2001; Hewitt, 2004; Stewart et al., 2010; Ramírez-Barahona and Eguiarte, 2013; Castellanos-Morales et al., 2016). During the Pleistocene pronounced glacial cycles, temperature fluctuations have been very marked, alternating between a cooler climate than the present (6 to 8°C lower) in the glacial periods and a warmer climate than the present (2 to 3°C higher) in the interglacials (Caballero et al., 2010; Ramírez-Barahona and Eguiarte, 2013). Throughout the repeated Pleistocene cycles, populations of warm desert biota retreated from northern to southern latitudes during adverse environmental conditions, and then recolonized or expanded towards the north when conditions improved (Hewitt, 2004; Ramírez-Barahona and Eguiarte, 2013; Loera et al., 2017; Scheinvar et al., 2017). This occurred several times, causing cycles of range contraction and expansion. After the Last Glacial Maximum (LGM ∼ 23–18 kya), different areas of Central and North America became more arid (Ramírez-Barahona and Eguiarte, 2013) and this aridification probably led to important changes in the distribution and composition of species in arid zones.

Of particular interest for understanding the evolutionary dynamics of arid and semiarid environments in North America is the Valley of Tehuacán-Cuicatlán Biosphere Reserve (called Tehuacán Valley hereafter) (Dávila et al., 2002; Valiente-Banuet et al., 2009). The Tehuacán Valley is located at the southeast of the state of Puebla and the northeast of the state of Oaxaca in Central Mexico, south of the Trans-Mexican Volcanic Belt (TMVB), representing the southernmost semiarid area in Mexico (Rzedowski and Huerta, 1978; Hafner and Riddle, 2011; El-Ghani et al., 2017) (**Figure 1A**). It has a distinctive biotic megadiversity and it is one of the main reservoirs of biological diversity for the arid zones of North America. Despite being geographically isolated from the rest of the arid and semiarid regions of North America (e.g., Chihuahuan and Sonoran deserts), the Tehuacán Valley has been characterized as a hot-spot of plant diversity and endemism (UNESCO, 2012), with more than 3,000 species of seed plants reported (Dávila et al., 2002). According to Valiente-Banuet et al. (2009) the late Pleistocene climate changes (10,000 to 1,000 kya) were very important for the current geomorphic and biotic composition of the Tehuacán Valley, suggesting that local plant communities were recently assembled.

Considering that the total area of the Tehuacán Valley is relatively small (10,000 km^2) and that the current vegetation of the area is of recent origin (Valiente-Banuet et al., 2009), it is interesting that the Valley harbors the highest diversity of succulent plant families such as Cactaceae and Agavaceae (Valiente-Banuet et al., 2000; Tambutti, 2002; García-Mendoza, 2011; Delgado-Lemus et al., 2014). In addition, this region comprises an heterogeneous mosaic of environments determined by climates, soils, geomorphology and elevation, which is reflected in nearly 36 types of plant associations (i.e., groups of plant species that occur across the landscape; Valiente-Banuet et al., 2000, 2009), and in a high degree of endemism for several different groups (Dávila et al., 2002). In particular, the Tehuacán Valley contains the highest richness of *Agave* species in the world (Tambutti, 2002; García-Mendoza, 2011). A total of 34 *Agave* species have been recorded, 25 of them native to the region, while seven of them are endemic to this area (García-Mendoza, 2011; Delgado-Lemus et al., 2014). In comparison, in the Chihuahuan Desert, the largest and warmest desert in North America (Hernández et al., 2001; Olson and Dinerstein, 2002; Scheinvar et al., 2017), with an area of approximately 507,000 km^2 (Granados-Sánchez et al., 2011) only 19 *Agave* species have been reported (Chihuahuan Desert Homepage, Centennial Museum, The University of Texas[1]).

We propose that by studying in detail the phylogeography of representative species in the Tehuacán Valley we will be able to better understand the evolutionary dynamics that generate the high biodiversity and endemism found in this semiarid regions. Accordingly, we investigated the population structure and phylogeography of *Agave kerchovei* Lem., examining polymorphisms observed in cpDNA sequences. This species is representative of the Tehuacán Valley, given that it is mostly restricted to this region (García-Mendoza, 2011; Brena-Bustamante, 2012; Brena-Bustamante et al., 2013) (**Figure 1A**). Our study aimed to (i) characterize the levels of genetic diversity and genetic structure in *A. kerchovei*; (ii) predict the distribution of *A. kerchovei* for the present day and reconstruct the past geographical history of the species, by constructing ecological niche models (ENM); and (iii) compare the levels of diversity in this species with those estimated for the widely distributed *A. lechuguilla*, a related species from the Chihuahuan desert. We expected to detect low levels of genetic diversity given the small size and restricted distribution of *A. kerchovei* populations, as well as a high genetic structure, given population isolation. Due to the high topographic complexity

[1]http://museum2.utep.edu/chih/chihdes.htm

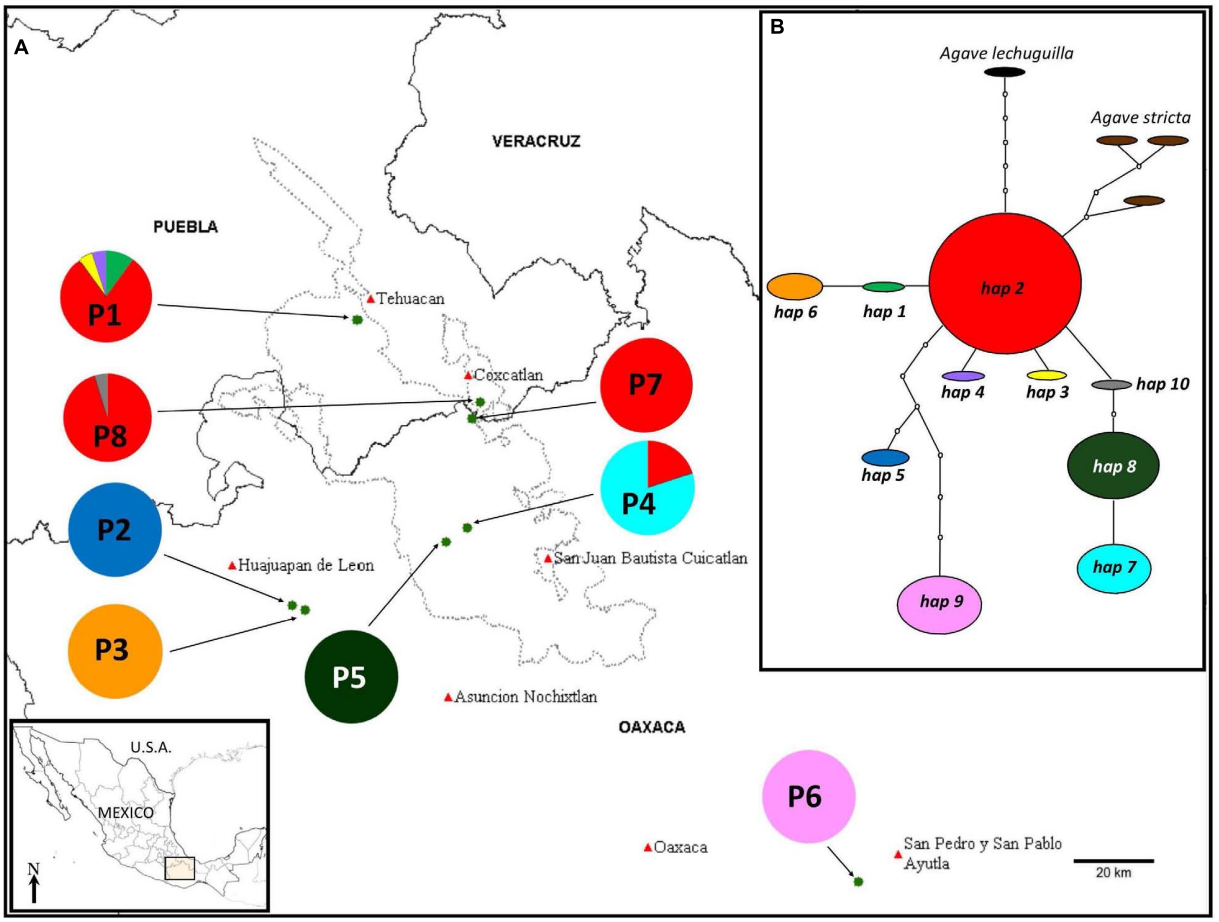


FIGURE 1 | (A) Population locations (names as in **Table 1**) and distribution of haplotypes in the different populations. **(B)** Minimum-spanning network of the chloroplast haplotypes found in the eight populations of *A. kerchovei* studied. The network represents the most parsimonious connections. Each haplotype is represented by a circle whose surface is proportional to the number of individuals bearing it. Lines between haplotypes indicate one mutational change. Missing haplotypes in the sample are represented by dotted circles. Populations 1, 7, and 8 correspond to one SAMOVA group, while each of the rest of the populations represent one different group, with a total of six groups, F_{CT} = 0.915.

of the Tehuacán Valley (Valiente-Banuet et al., 2000, 2009), which translates into high climatic variability, we also predicted that the high environmental variance among populations will translate into high levels of genetic differentiation in *A. kerchovei* compared to the populations of the closely related but widespread *A. lechuguilla*, endemic to the Chihuahuan Desert (Scheinvar et al., 2017).

MATERIALS AND METHODS

Study Species

In this study, we will follow Gentry's classification (1982), who does not recognize subspecies or varieties of *A. kerchovei* [although The Standard Cyclopedia of Horticulture (1914) mentions varieties of *A. kerchovei*, including *A. kerchovei* var. *beaucarnei*, also included in the Tropicos Database[2], these varieties are now regarded as synonyms of *A. kerchovei* (The Plant List[3]; and Global Biodiversity Information Facility, GBIF[4])].

Agave kerchovei is a species with a narrow distribution. Little is known about its natural history or ecology (Brena-Bustamante, 2012; Brena-Bustamante et al., 2013). Gentry (1982, p. 153; 190), reports the species distribution in the states of Puebla and Oaxaca and in one locality in the state of Hidalgo (Barranca de Metztitlán). More recently, García-Mendoza (2011) reported populations only in the states of Oaxaca and Puebla and extensive field work with *Agave* in the Barranca de Metztitlán (Hidalgo) did not detect a single plant of the species (Rocha et al., 2005; Eguiarte and Scheinvar, 2008). *Agave kerchovei's* present distribution is mostly restricted to the Tehuacán Valley (García-Mendoza, 2011; Brena-Bustamante, 2012) (**Figure 1A**), although more

[2]http://legacy.tropicos.org/

[3]http://www.theplantlist.org

[4]https://www.gbif.org

FIGURE 2 | *Agave kerchovei* from Villa Tamazulapám del Progreso, Oaxaca (P2 in **Figure 1**).

populations have been reported in herbarium records (see below). It is currently considered as a vulnerable species in the IUCN Red List of Threatened Species 2019 (García-Mendoza et al., 2019). Furthermore, the known populations of *A. kerchovei* are small and isolated, sometimes consisting of fewer than 10 individuals (Brena-Bustamante, 2012), inhabiting rough terrain within the mountainous regions of the Valley. *Agave kerchovei* is a medium-sized *Agave* with a brilliant green short stem and wide rosettes (Gentry, 1982) (**Figure 2**), and evidence of both sexual and asexual reproduction has been observed in some of its populations (Brena-Bustamante, 2012). Local people sometimes promote *in situ* local propagation of propagules of the species (Delgado-Lemus et al., 2014), collecting the flowers for human consumption and their leaves for making fences (Brena-Bustamante, 2012; Brena-Bustamante et al., 2013).

Population Sampling

Individuals from eight wild populations were sampled across the species' geographical range throughout the Valley and nearby areas (**Figure 1** and **Table 1**). For each population, leaf tissue from 13–23 adult individuals, separated by at least 5 m to avoid sampling the same genetic clone, were collected, with the exception of populations 2 and 3, where only four and eight adult individuals were found and collected, respectively (**Table 1**). Leaf tissue of each collected individual was stored at −80°C in the tissue collection of the Instituto de Ecología, UNAM, and is available upon request. A voucher specimen of population 8 (San Rafael Coxcatlán) (collection number, 117) is deposited at the Facultad de Estudios Superiores, Iztacala, UNAM herbarium, IZTA (acronym fide Thiers, 2016). Additionally, herbarium samples for different localities are deposited at the MEXU, UNAM herbarium.

DNA Extraction, Amplification, and Sequencing

Total DNA was extracted by grinding approximately 0.25 g of fresh leaf tissue in liquid nitrogen using a CTAB (2X) extraction protocol (Doyle and Doyle, 1987) and resuspended in 60 μl of ultrapure water (Molecular Biology Reagent; SIGMA).

Three non-coding chloroplast (cpDNA) regions [*psbJ-petA*, *rpl32-trnL* (Shaw et al., 2007) and *trnL-trnF* (Taberlet et al., 1991)] were amplified by polymerase chain reaction (PCR) and sequenced for 136 individuals. The PCR amplifications were carried out in a GeneAmp® PCR system 2700 (Applied Biosystems) in total reaction volumes of 30 μl, containing 25–40 ng of total DNA, 1X of PCR buffer (100 mM Tris-HCL, 500 mM KCl, 10μg/ml gelatin, 1% Triton, 1.5 mg/ml BSA), 1.5 mM $MgCl_2$ (for primers *rpl32-trnL* and *trnL-trnF*) and 2 mM $MgCl_2$ (for primer *psbJ-petA*), 0.2 mM of each dNTP, 0.3 μM of each primer, and 1 unit of *Taq* DNA polymerase. The cycling conditions consisted of an initial denaturation at 94°C for 5 min, 35 cycles of 94°C for 30 s, 55°C for 30 s (for primers *rpl32-trnL* and *trnL-trnF*) or 55°C for 50 s (for primer *psbJ-petA*), and 72°C

TABLE 1 | Measures of genetic variation for the analyzed populations of *Agave kerchovei*.

Population	Locality	Latitude (N)	Longitude (W)	Altitude (m)	*N*	*S*	*h*	*Hd*	π	*D'* Tajima
1	SW from Santa María Coapan, Puebla	18° 24′	97° 26′	1817	20	3	4	0.36316	0.00018	−1.4407
2	NW from Villa de Tamazulapam del Progreso, Oaxaca	17° 42′	97° 37′	1927	4	0	1	0	0	0
3	NW from Villa de Tamazulapam del Progreso, Oaxaca	17° 41′	97° 35′	1900	8	0	1	0	0	0
4	NE from Santa María Ixcatlán, Oaxaca	17° 52′	97° 10′	2091	20	2	2	0.33684	0.00031	0.4572
5	W from Santa María Ixcatlán, Oaxaca	17° 50′	97°13′	2113	23	0	1	0	0	0
6	W from Santa María Albarradas, Oaxaca	16° 57′	96° 11′	1682	20	0	1	0	0	0
7	"Tecolotiopa", San Gabriel Casa Blanca, Oaxaca	18° 09′	97° 08′	946	20	0	1	0	0	0
8	"Cerro de las Compuertas", San Rafael, Coxcatlán, Puebla	18° 11′	97° 08′	989	13	1	2	0.15385	0.00007	−1.1491
Total					**128**	**8**	**10***	**0.71801**	**0.00078**	**0.3732**

*N, sample size; S, number of polymorphic sites; h, number of haplotypes; Hd, haplotype diversity; π, nucleotic diversity. All D' Tajima test's values were non-significant. *Total number of haplotypes found in A. kerchovei.*

for 1 min followed by a final extension at 72°C for 8 min. PCR products were purified and sequenced in the High Throughput Genomic Unit, University of Washington, USA.

The quality of the sequences and the forward and reverse assembly was assessed by direct inspection using the Phrap-Phred, Consed V 19.0 software (Ewing et al., 1998; Gordon et al., 1998). Sequences alignments were made with CLUSTALW (Thompson, 1994) as implemented in BIOEDIT 7.1.3.0 (Hall, 1999). Indels (insertion/deletion) were coded as single base characters to treat them as single events, rather than multiple independent events, and the chloroplast regions were concatenated with DnaSP v5.10.1 (Librado and Rozas, 2009). Sequences were deposited at NCBI GenBank.

Relationship Among Haplotypes

To assess genetic relationships among haplotypes, we constructed a haplotype network, as implemented in the program TCS v1.21 (Clement et al., 2000) using 95% connection probability limit, treating gaps as single evolutionary events and indels as a fifth state of character. For this analysis, we included sequence data obtained in this study along with sequences from *Agave stricta* and *A. lechuguilla*, from Martínez-Ainsworth (2013) and Scheinvar et al. (2017), respectively.

Genetic Diversity and Structure

The observed number of haplotypes with (h) and without indels (h^*), haplotype diversity (Hd), nucleotide diversity (π), and the Watterson estimator of theta (θ) for each population were obtained using the program DnaSP v5.10.1 (Librado and Rozas, 2009). These summary statistics were re-estimated for the populations of *A. lechuguilla*, reported by Scheinvar et al. (2017).

We used the program Arlequin version 3.5 (Excoffier and Lischer, 2010) to estimate pairwise F_{ST} (Weir and Cockerham, 1984) between populations to test for isolation by distance with a Mantel test (Mantel, 1967) and to conduct a molecular analysis of variance (AMOVA) (Excoffier et al., 1992). Finally, with the program PERMUT (Pons and Petit, 1996), we evaluated with 1000 permutations whether there was significant phylogeographic structure by estimating and comparing the differentiation parameters N_{ST} and G_{ST}.

We conducted a spatial analysis of molecular variance (SAMOVA) using Samova version 1.0 to explore the population groupings that maximized the proportion of genetic variance at the total population level (F_{CT}), without *a priori* assignment of individuals to population groupings (Dupanloup et al., 2002). SAMOVA identifies groups of populations (K) that are geographically homogeneous and genetically differentiated from each other while maximizing the proportion of total genetic variance due to differences among groups of locations (F_{CT}). We explored K-values with 100 permutations for each group of populations. The most likely number of groups (K) was determined by running the program with different groups of populations which ranged from 2 to 14 groups, choosing those partitions with a maximum F_{CT} value, as suggested by Dupanloup et al. (2002). Levels of genetic differentiation among geographic regions and groups identified by SAMOVA were estimated using pairwise F_{ST} and 10,000 permutations were used to calculate the corresponding probabilities in Arlequin version 3.5 (Excoffier and Lischer, 2010).

We calculated Tajima's D with DnaSP v5.10.1 (Librado and Rozas, 2009) to infer basic aspects of demographic histories. Tajima's D (Tajima, 1989) statistic is based on the differences between the number of segregating sites and the average number of nucleotide differences. Significant negative D ($P < 0.05$) statistic can indicate no neutrality, or population expansion.

Ecological Niche Modeling

We used Maximum Entropy Modeling (MaxEnt; Phillips et al., 2006) to predict the distribution of *A. kerchovei* for the present day (PRE). MaxEnt is a presence-background algorithm that estimates a species' potential ecological niche by finding a probability distribution of environmental variables that best describes the occurrence localities, while being able to differentiate between occurrence and background sites (Phillips et al., 2006; Elith et al., 2011; Peterson et al., 2011; Vasconcelos et al., 2012). The predicted distribution was projected into three time periods in the past: the mid Holocene (MH, ∼6 kya); the Last Glacial Maximum (LGM, ∼21 kya), and the Last Inter-Glacial period (LIG, ∼130 kya). These correspond to periods for which global paleo-climate layers are available (Hijmans et al., 2005).

Present day occurrence data for *A. kerchovei* were retrieved from herbarium records and sampled populations, representing 37 unique occurrence localities at a 30 arc-second resolution. The models were built from climate layers obtained from the WorldClim database version 1.4 (Hijmans et al., 2005). We selected five climate layers based on pairwise Pearson correlations. For this, we identified pairs of climate layers with a correlation coefficient greater than 0.8 and then excluded the layer with the highest variance inflation factor (VIF). The five layers correspond to: Annual Mean Temperature (bio 1), Mean Diurnal Range (bio 2), Isothermality (bio 3), Annual Precipitation (bio 12), and Precipitation Seasonality (bio 15). The same five climate layers for the MH and the LGM were obtained from the Community Climate System Model (CCSM; Kiehl and Gent, 2004). The climate layers for the LIG were also obtained from the CCSM (Otto-Bliesner et al., 2006). All climate layers had a 30 arc-second resolution.

We approximated the accessible area (M) for *A. kerchovei* by masking the climate layers with a spatial polygon defined from the terrestrial eco-regions (Olson et al., 2001) inhabited by the species, which was further delimited using a 2° buffer around occurrence localities. The M was extended to include adjacent terrestrial eco-regions, and a 3° buffer for the projection of the models into past climate conditions. Distribution models were predicted using this area to ensure coverage of the temporal range dynamics of the species and to decrease the over-prediction associated to the use of large modeling areas. We used the 'ENMeval' package in R (Muscarella et al., 2014; R Core Team, 2019) to identify the best settings for the regularization multiplier (RM) and 'features' in MaxEnt; we used the 'randomkfolds' method with $k = 20$ over five RM values (0.5, 0.7, 0.9, 1.1, 1.3).

Distribution models were built with 20 replicates using 10,000 random background points, a RM of 0.7, with hinge features only, without extrapolation and no clamping. Models were validated using 25% of the occurrence data using the receiver operating curve (ROC) statistic. We evaluated the replicate models using the area under the receiver operating curve (AUC), where models with values below 0.8 were dismissed. The remaining replicates models were combined to construct the present-day model and then projected into the past climate layers.

We followed the same modeling procedure to generate present-day distribution models for *A. lechuguilla* using 97 unique occurrence localities at a 30 arc-second resolution. These models were projected into the LIG, LGM and MH. In this case, the models were built after selecting eight layers corresponding to Annual Mean Temperature (bio 1), Mean Diurnal Range (bio 2), Mean Temperature of the Driest Quarter (bio 9), Mean Temperature of Warmest Quarter (bio 10), Mean Temperature of Coldest Quarter (bio 11), Annual Precipitation (bio 12), Precipitation of the Driest Month (bio 14), and Precipitation Seasonality (bio 15).

We used the resulting distribution models for *A. kerchovei* and *A. lechuguilla* to visualize the changes in suitable climatic conditions (climate suitability) to which populations may have been subjected since the Last Interglacial period. For this, we extracted the climatic suitability values through time (i.e., LIG, LGM, MH, PRE) for those grid-cells associated with the occurrence localities of sampled populations, being 0 no suitability and 1 the maximum suitability. To account for the possible bias of using a single grid cell to characterize the climate suitability of populations' localities, we generated 100 replicates of sample localities by adding random noise to the populations' geographic coordinates within a buffer of ∼10 km^2 (0.08°) centered on the sampling locality. For each population, we estimated the half sample mode (HSM) of the suitability values across replicates to visualize the changes in climatic conditions through time.

Genetic Diversity and Environmental Variance

We estimated the variance of environmental variables (i.e., climate and altitude) for sampled populations of *A. kerchovei* and *A. lechuguilla* (Scheinvar et al., 2017). We used the 19 climate layers obtained from the WorldClim database (Hijmans et al., 2005) and the GTOPO30 global digital elevation model (DEM) from the USGS-EROS Data Center. For each of the environmental variables, we extracted the data for those grid-cells associated with the occurrence localities of sampled populations. As mentioned above, we accounted for the possible bias of using a single grid cell to characterize the environment of populations by generating 100 replicates of sampled localities.

Populations of *A. lechuguilla* were assigned into four different groups according to their geographic location and genetic composition (Scheinvar et al., 2017), whereas populations of *A. kerchovei* were treated as a single group (**Table 1**). We approximated the environmental variance by performing a Principal Component Analyses on populations' environmental conditions and estimating the variance for the first two principal components. We tested the correlation between environmental variance and genetic diversity (i.e., nucleotide and haplotype diversity) among the five groups of populations using simple linear regression. For this, we used the components of the environmental PCA as predictors and the indices of genetic diversity as response variables.

RESULTS

Genetic Diversity and Structure

The combination of the three non-coding chloroplast (cpDNA) regions [*psbJ-petA*, *rpl32-trnL* (Shaw et al., 2007)] and *trnL-trnF* (Taberlet et al., 1991) resulted in sequences 2,188 bp long. In total, we found eight segregating sites and eight indels, resulting in ten haplotypes *for A. kerchovei* (nine when not considering indels). Total haplotype diversity (*Hd*) was 0.718, but the average per population diversity was far lower (*Hd* = 0.107), indicating strong genetic differentiation among populations (**Table 1**). Total nucleotide diversity (π) was low (0.00078), with most populations having null nucleotide diversity.

The haplotype network was well resolved with ten haplotypes for *A. kerchovei* (**Figure 1B**). The only haplotype shared between more than two populations was haplotype 2, shared among four and was also the haplotype with the highest frequency. The remaining nine haplotypes were found to be exclusive to single populations, with five populations being fixed for a particular haplotype (**Figure 1A**). The network has a star-like shape, with the most common haplotype (h2) at the center of the network; from this haplotype, two singe haplotypes are derived (3 and 4), that belong to population 1 (P1) and three haplogroups, one consisting of haplotypes 1 and 6, another one with haplotypes 10, 8 and 7, where there is one missing mutation, and a third one that comprises haplotypes 2 and 6 with seven missing mutations, which belong to populations located outside the Tehuacán Valley (P2 and P6), being population 6 the most isolated. From the central and most common haplotype (2), the haplotypes found in *A. stricta* and *A. lechuguilla* are also derived (**Figure 1B**).

The prevalence of private haplotypes within populations resulted in a high genetic differentiation (F_{ST} = 0.928). According to the PERMUT analysis, there are no significant differences comparing N_{ST} (0.943) vs. G_{ST} (0.880), suggesting lack of phylogeographic structure. Therefore, a Mantel test was not significant (data not shown), which suggests that there is no evidence of isolation by distance in *A. kerchovei*. Tajima's *D*, although positive, 0.37320, was not significant (p = 0.704).

In the AMOVA, the highest percentage of variation (92.80%) was explained by differences among populations, whereas only 7.20% was found within populations (**Table 2**). Total genetic differentiation (0.928) indicates a high genetic structure. Therefore, most of the variation within the species can be explained by differences between populations, rather than the difference within populations. In accordance, a SAMOVA analysis suggested six groups with a value of F_{CT} = 0.915. One cluster is composed of three populations (1, 7 and 8), in the

TABLE 2 | Results of analysis of molecular variance (AMOVA) of the analyzed populations of *A. kerchovei* for cpDNA.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among populations	7	243.430	2.214*	92.80
Within populations	120	20.623	0.172	7.20
Total	127	264.062	2.385	
*FST** = 0.928				

**P < 0.01, d.f., degrees of freedom.*

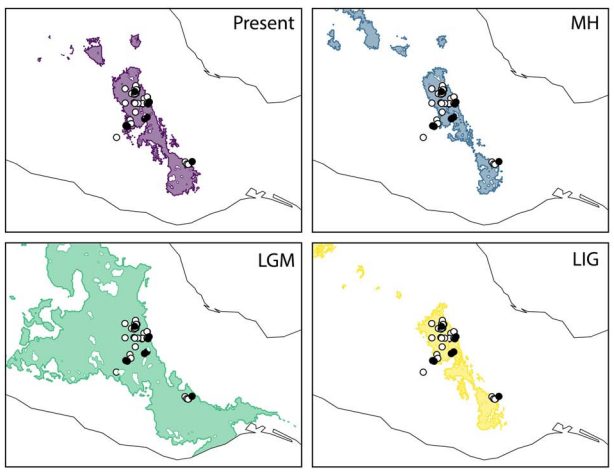


FIGURE 3 | Geographical projections of the ecological niche model for *Agave kerchovei* for the present-day, the middle Holocene (MH, ∼6 kya), the Last Glacial Maximum (LGM, ∼21 kya), and the Last Interglacial (LIG ∼110 kya). Shaded areas in past models represent the projected areas potentially inhabited by the species. Black circles represent the populations sampled for the present study (eight populations). White circles represent the occurrence records used to construct the ecological niche models (37 records).

North, while the rest of the groups are conformed only by one population each (**Figure 1A**).

Ecological Niche Analysis

The predictive performance of the bioclimatic models was adequate with an AUC > 0.82 across replicates. The projected distribution for *A. kerchovei* during the LIG shows the most restricted distribution across all the time periods analyzed, where ideal climate conditions for the species were geographically restricted to an area equivalent to 75% of the present-day distribution (**Figure 3**). Subsequently, according to our models, the LGM witnessed a significant geographical expansion in the ideal climate conditions for *A. kerchovei*, which were broadened by 522% relative to the present-day distribution. The projected models then predicted a geographical contraction for *A. kerchovei* during the MH (93% relative to the present-day), with this geographical extent remaining stable ever since. Our models revealed two main areas that have remained more or less stable and isolated from each other: (1) the Tehuacán Valley in the North, and (2) the Central Valleys of Oaxaca in the South.

Accordingly, we characterized the changes in suitable climatic conditions (climate suitability) to which the geographic regions corresponding to present-day sampled populations may have been subjected since the LIG. We found that the climatic suitability varies significantly through time (i.e., LIG, LGM, MH, PRE) among the occurrence localities of each sampled population (**Figure 4B**). Interestingly, most occurrence localities show an improvement in the suitability values during the LIG–LGM transition, with this improvement trend continuing into the MH and present-day, but only for those populations within the core areas of the Tehuacán Valley (populations P1, P4, P5, P7 and P8; **Figure 4B**). Also, our models show that the three most isolated populations – which lie outside the Tehuacán Valley (P2, P3 and P6, **Figure 1A**) and harbor the most divergent haplotypes – experienced a dramatic decline in climatic suitability toward the MH (**Figure 4B**).

To explore whether the detected patterns are similar in other *Agave* species with wider distribution we characterized the temporal variation in climatic suitability in the widespread *A. lechuguilla* from the Chihuahuan Desert, finding that the temporal changes in climate suitability are different from those observed for *A. kerchovei* (**Figure 4A**). In *A. lechuguilla*, we observed a drastic increase in the climate suitability in occurrence localities during the LGM–MH transition and into the present-day for northern populations of *A. lechuguilla*, whereas the southern populations showed a climate suitability trend resembling *A. kerchovei* (**Figure 4A**). In this context, we suggest that the Tehuacán Valley, along with the southernmost portion of the Chihuahuan Desert, have been regions with relatively stable climatic conditions suitable for the survival of *Agave* species through the last 110,000 years, explaining, at least in part, the high *Agave* species diversity in these areas (Scheinvar et al., 2017).

Genetic Diversity and Environmental Variance

The first two components of the PCA on environmental variables for *A. kerchovei* explained 46.18% and 21.07% of the total climatic variance, respectively (data not shown). The first PCA was most strongly positively correlated with annual precipitation, summer precipitation and altitude, whereas it showed the most negative correlation with summer temperature. On the other hand, the second PCA was most positively correlated with winter precipitation, and negatively with winter temperature and precipitation seasonality (data not shown).

We also compared the environmental variance within groups of sampled populations of *A. kerchovei* and *A. lechuguilla* (Scheinvar et al., 2017; Scheinvar, 2018) with a principal component analyses on populations' environmental (PCA_{ENV}) conditions, and estimating the variance for the first two principal components (**Figures 5A,B**), which jointly accounts for 69.6% of the variance (47.4% and 22.2%, respectively; **Figure 6**). Populations of *A. kerchovei* and *A. lechuguilla* are broadly ordered in a South–North direction along the first two principal components (**Figure 6A**). The $PC1_{ENV}$ was most strongly associated with temperature and precipitation

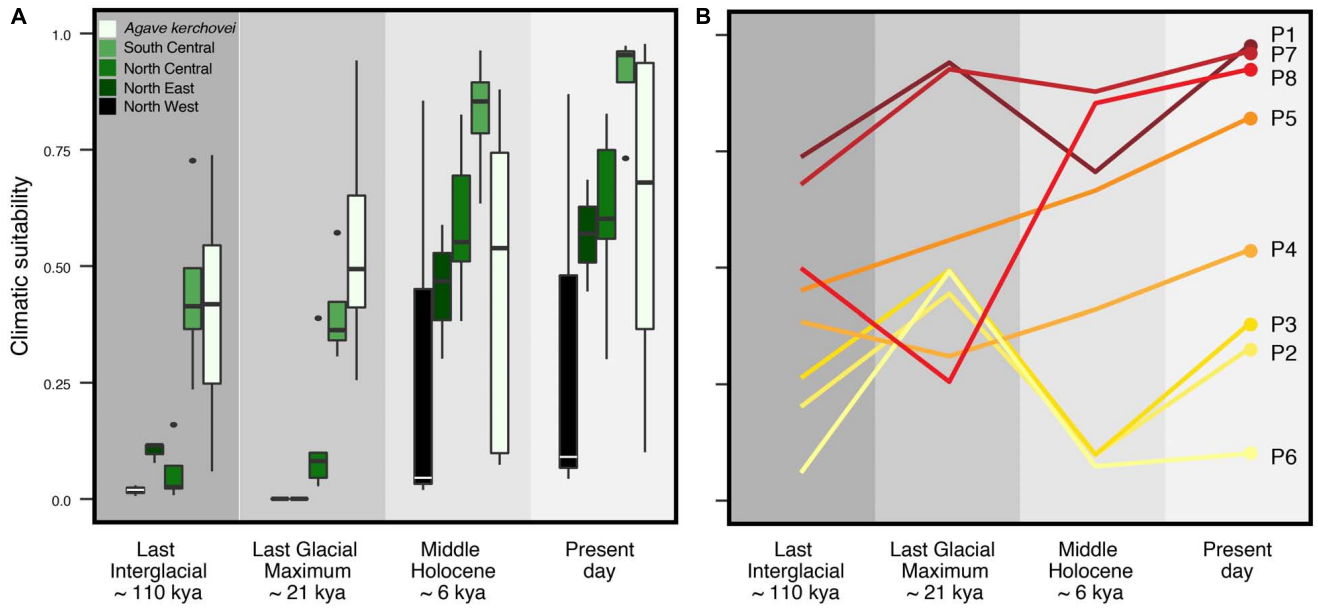


FIGURE 4 | Climatic suitability through time estimated across sampled populations of *Agave kerchovei* and *A. lechuguilla* using the projections of the species' ecological niche models. **(A)** Distribution of climatic suitability values through time within the five groups of populations of *Agave kerchovei* and *A. lechuguilla*. **(B)** Climatic suitability values through time for the eight sampled populations of *A. kerchovei*. Population names as in **Table 1**.

during the summer months (i.e., the Wettest and Warmest Quarters), whereas the $PC2_{ENV}$ was most strongly associated with temperature and precipitation during the winter months (i.e., the Driest and Coldest Quarters) (**Figure 6B**). For these two principal components ($PC1_{ENV}$ and $PC2_{ENV}$), we obtained each population's scores and estimated the variance within geographically defined groups of populations (**Figure 6A**). The range of values within groups of populations for $PC1_{ENV}$ and $PC2_{ENV}$ shows a pattern where more environmental variance exists among populations of *A. kerchovei* than within geographically defined groups of populations of *A. lechuguilla* (**Figures 5A,B**).

The linear regression models showed no association ($b = 0.11$, $p = 0.7$) between levels of haplotype diversity within groups of populations and their variance along $PC1_{ENV}$ (**Figure 5C**). Nonetheless, we found a positive association ($b = 0.39$, $p = 0.08$, $R^2 = 0.59$) between the levels haplotype diversity within groups of populations and their variance along PC2ENV (**Figure 5D**). Sample size was not correlated with haplotype diversity ($b = 0.001$, $p = 0.47$) nor with estimates of environmental variance along $PC2_{ENV}$ ($b = 0.005$, $p = 0.17$), but it was correlated with environmental variance along $PC1_{ENV}$ ($b = 0.006$, $p = 0.02$). The high environmental variance along $PC2_{ENV}$ observed among populations of *A. kerchovei* – which is associated with variance in winter temperatures and precipitation (**Figure 6B**) – is reflected in higher levels of total genetic diversity than those observed for most of the groups of populations of *A. lechuguilla*. This is also probably facilitated by the geographical isolation of populations within the Tehuacán Valley (**Figure 1A**). However, nucleotide diversity did not show a significant correlation with environmental variance.

DISCUSSION

Agave kerchovei is a species with a restricted distribution, with populations mainly in the Tehuacán Valley. As occurs in many rare or endemic species with few and usually small populations, some of its populations are completely depleted of genetic variation and genetic structure is high, which may be a consequence of demographic phenomena such as inbreeding, genetic bottlenecks or drift (Hamrick and Godt, 1990; Nybom and Bartish, 2000; Gibson et al., 2008). Pleistocene glacial cycles had an important role in the distribution of *A. kerchovei,* and climatic variability appears to have had a significant effect on the levels of genetic diversity and differentiation among populations. Overall, in terms of climate, the species' core distribution area within the Tehuacán Valley appears to have remained relatively stable over the last ∼6 kya, whereas peripheric populations outside the Valley, which apparently are genetically isolated, have been subjected to more climatic instability.

Genetic Diversity and Structure

There have been several studies in the population genetics of different *Agave* species. Earlier studies were reviewed in Eguiarte et al. (2013). In general, agaves are rich in genetic variation, and usually genetic differentiation is low among populations, but there is a wide variation among species (e.g., Byers et al., 2014; Parker et al., 2014; Félix-Valdez et al., 2016; Trejo et al., 2016; Figueredo-Urbina et al., 2017). For *A. kerchovei,* considering its restricted distribution, the species shows higher levels of chloroplast genetic variation ($Hd = 0.718$) than those estimated for *A. stricta* ($Hd = 0.524$), which is an endemic of the Tehuacán Valley (Martínez-Ainsworth, 2013). This might

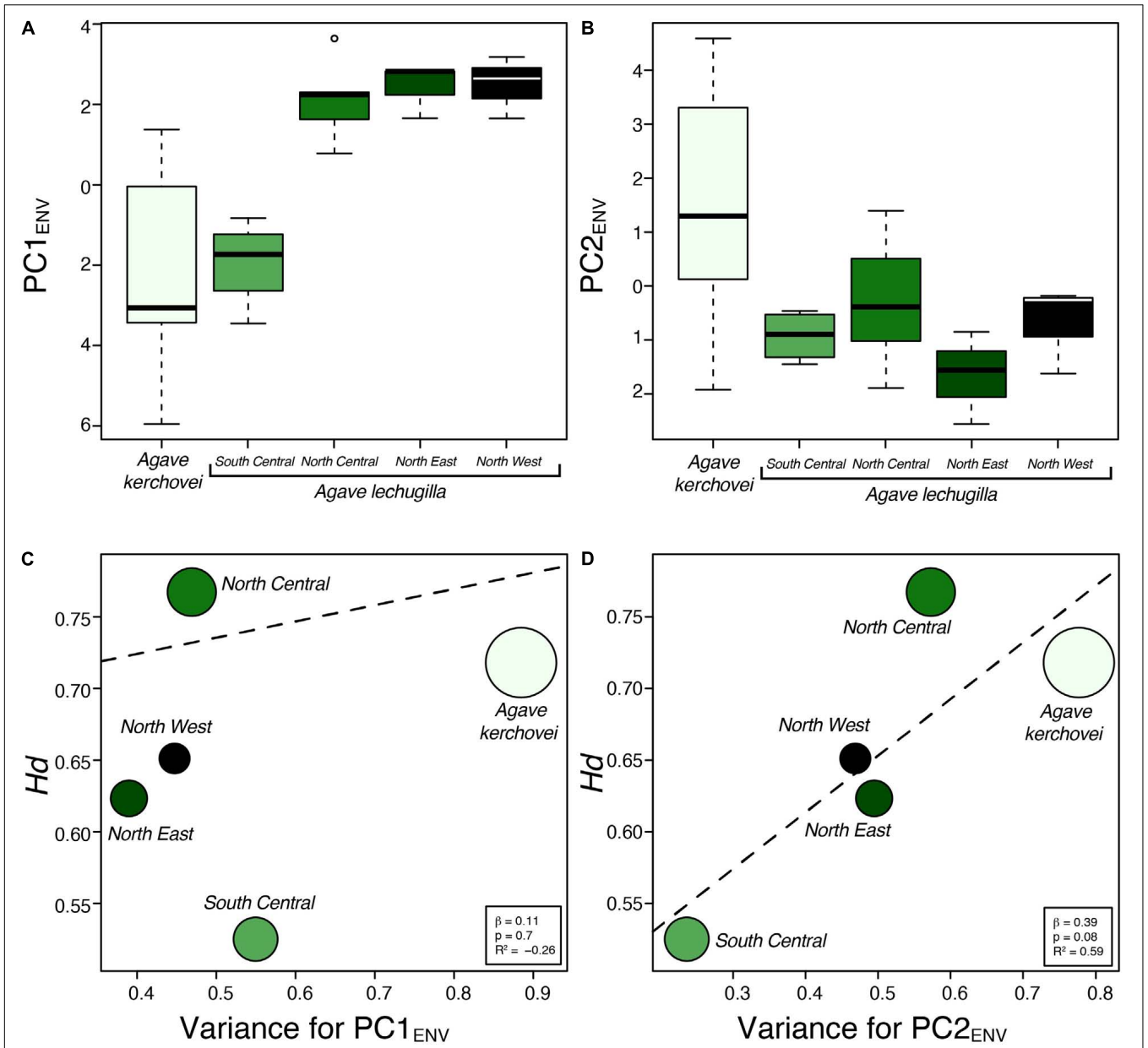


FIGURE 5 | Environmental variance observed across sampled populations of *Agave kerchovei* and *A. lechuguilla*. **(A,B)** Environmental values in the two species of *Agave* obtained from the scores for the first two principal components of the environmental principal component analysis ($PC1_{ENV}$, $PC2_{ENV}$). **(C,D)** Relationship between environmental variance ($PC1_{ENV}$ and $PC2_{ENV}$) and haplotype diversity estimated for the five groups of populations. Dashed line depicts linear regression models. Circle size is proportional to sample size (*n*).

be related to significant biological and ecological differences between the species, with *A. kerchovei* inhabiting regions of more topographic complexity (Gentry, 1982), and hence climate variability, than those inhabited by *A. stricta,* but a detailed study on the differences of the distribution and ecology between this two species would be necessary to deepen in the causes of this observation. Nevertheless, low levels of genetic diversity were found within populations of *A. kerchovei*, with five of the eight populations analyzed having no genetic variation. Although these two *Agave* species show different levels of genetic diversity, it is worth noting that the two show lower levels of chloroplast diversity than one wide ranging *Agave* species from the Chihuahuan Desert, *A. lechuguilla* (*Hd* = 0.931) (Scheinvar et al., 2017; Scheinvar, 2018), while *A. kerchovei* has similar values to another wide ranging *Agave* species from the Chihuahuan Desert, *A. striata (Hd* = 0.713), the sister species of *A. stricta* (Martínez-Ainsworth, 2013).

Genetic differentiation was high in *A. kerchovei* (F_{ST} = 0.928, G_{ST} = 0.824). High genetic structure using similar chloroplast sequences was also found in *A. striata* (F_{ST} = 0.929,

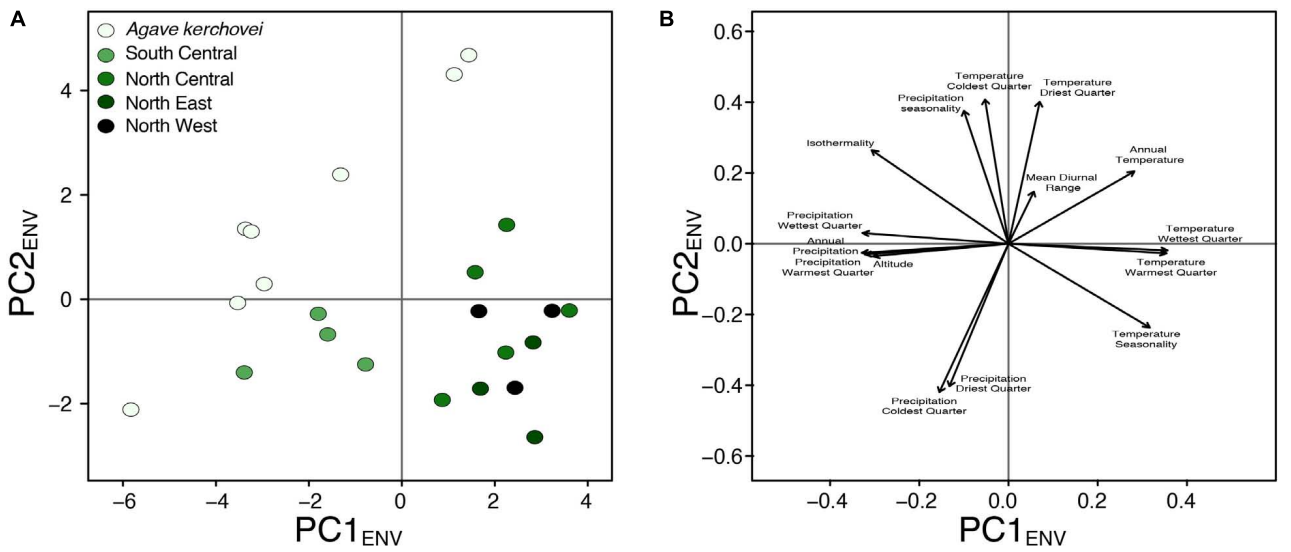


FIGURE 6 | Principal component analyses of environmental preferences for *Agave kerchovei* and *A. lechuguilla*. **(A)** Scores plot, and **(B)** loading plot for the first two principal components ($PC1_{ENV}$, $PC2_{ENV}$).

G_{ST} = 0.697; Martínez-Ainsworth, 2013) and *A. stricta* (F_{ST} = 0.944, G_{ST} = 0.898; Martínez-Ainsworth, 2013), but it was lower in *A. lechuguilla* (G_{ST} = 0.780; Scheinvar, 2018; Scheinvar et al., 2017).

Throughout most of the distribution of *A. kerchovei* differentiation among populations appears to be very high, with the exception of the northernmost populations (P1, P7, P8) (**Figure 1A**), usually harboring a single, unique (private), haplotype. This could be the result of a combination of ecological and historical factors, including changes in climate and the rough and complex topography of the region, generating different conditions that restrict the types of vegetation within the altitudinal ranges (Valiente-Banuet et al., 2000, 2009). In this case, genetic differentiation among populations would result due to restricted gene flow, small effective population sizes and strong adaptation to local abiotic and biotic conditions. Thus, populations in most of the cases seem to be evolving locally, which could suggest a process of local adaptation. For instance, Brena-Bustamante (2012) found that percentage of germination in the field is different for each of the populations studied (P7, San Gabriel, 41.4% under shadow and 6.6 under light conditions versus 19% and 5.4% respectively in P8, San Rafael), suggesting population differentiation. In addition, vegetation cover and species composition are very different for each population, despite being very near (∼5 km, Brena-Bustamante et al., 2013) and belonging to the same broad type of vegetation (tropical deciduous forest), suggesting that the environment is different, and in consequence local adaptation could also differ.

Strong genetic differentiation among populations of *A. kerchovei* is further supported by the absence of a clear phylogeographic signal, the absence of isolation by distance among populations and the fact that some populations consist of very few individuals. Another important biological factor that may be relevant to explain the genetic patterns observed in *A. kerchovei* is clonal reproduction, possibly leading to patches of genetically identical individuals. Accordingly, Brena-Bustamante (2012) found evidence of sexual and asexual reproduction in *A. kerchovei*. Clonal propagation would reduce the number of genetic individuals in a population (Honnay and Jacquemyn, 2008; Vallejo-Marín et al., 2010) and would explain at least in part why most sampled populations only harbor one single chloroplast haplotype. In order to better asses the importance of asexual reproduction, it will be relevant to use in the future nuclear DNA to better disentangle the roles of seed and clonal reproduction.

Evolutionary History

It has been suggested that the Pleistocene was an important period for the diversification of *Agave* (Scheinvar et al., 2017; Scheinvar, 2018). Accordingly, the divergence of *A. kerchovei* has been dated to the Pliocene-Pleistocene transition (∼2.6 Mya) (Scheinvar et al., 2017). The main inferred climate changes during this period in North America, were in mean temperatures, which are believed to be one of the main factors promoting range fragmentations/expansions and population isolation in plant species (Ramírez-Barahona and Eguiarte, 2013; Scheinvar et al., 2017). Accordingly, we inferred a historical range dynamic in *A. kerchovei* over the last 110,000 years, involving range expansion and contraction. More specifically, our results suggest that during the LIG, which in Central Mexico has been characterized as a more humid and warmer period (Metcalfe, 2006), *A. kerchovei* appears to have suffered its maximum range contraction (**Figure 3**), possibly surviving within isolated areas throughout the Tehuacán Valley. Over this period, arid adapted species decreased their distribution area and became isolated into smaller populations or refugia, increasing their population

differentiation (Scheinvar et al., 2017). It has been shown that physiological and morphological strategies allow Agaves to survive under extreme temperature conditions (Nobel and Smith, 1983; Martinez-Rodriguez et al., 2019). Afterward, the species appears to have experienced a significant range expansion during the LGM (∼21 kya), when the climate became drier and colder (Caballero et al., 2010), with a contraction in the MH and remaining stable until the present.

However, changes in climatic suitability appear to have been heterogeneous across the distribution of the species, showing that localities from the area of the Tehuacán Valley have exhibited a continuous improvement in the suitability values in the last 110,000 years, while in contrast, the three most isolated populations (P2, P3, and P6, **Figure 1A**), which also happen to harbor the most divergent haplotypes, experienced a dramatic decline in climatic suitability toward the MH (**Figure 4B**). In other words, populations in the core area appear to have experienced less changes than populations in the periphery, resulting in two main areas that have remained more or less isolated for *A. kerchovei*: (1) the Tehuacán Valley in the North and (2) the Central Valleys of Oaxaca in the South. This is in accordance with our genetic analysis of the distribution of variation in *A. kerchovei*, as the northern populations share a common haplotype, while the rest of the populations have unique haplotypes, resulting in high genetic structure, as shown by the analysis of AMOVA and SAMOVA. In the latter, almost every location is designated as a separate group. This could be attributed to genetic drift and local differentiation and adaptation, as for instance different microenvironmental conditions are found within each area in this region of Mexico (Valiente-Banuet et al., 2000, 2009), further promoting differentiation.

This differentiation could represent a process of incipient speciation, particularly regarding population P6, which is the most isolated at the eastern and southernmost extreme of its distribution. In the northern area, conformed of populations P1, P7, and P8, haplotype 2 is the most common and according to the network it can be considered as the ancestral haplotype, suggesting this area could be the ancestral area of the distribution.

When we compare temporal variation in climatic suitability in *A. lechuguilla* from the Chihuahuan Desert and in *A. kerchovei*, we can appreciate temporal changes appear to be very different in northern populations of *A. lechuguilla*, but not in the southernmost populations of the species, which show a climate suitability trend resembling the one observed for *A. kerchovei*. We suggest that the Tehuacán Valley, together with the southernmost portion of the Chihuahuan Desert, have been regions with climatic conditions suitable for the survival of *Agave* species through the last 110,000 years.

CONCLUSION

Our results suggest that Pleistocene climate fluctuations and the resulting contraction and expansion of *A. kerchovei* populations have led to changes in population sizes in which both genetic drift and the subsequent expansion of the distribution were probably important factors in generating the currently observed genetic structure. The high environmental variance observed among populations of *A. kerchovei* is reflected in high levels of among population differentiation, probably due to the geographical isolation of populations within and outside the Tehuacán Valley. Thus, topographic complexity through its effects on climatic variability appears to have a significant impact on the levels of genetic diversity and structure among populations.

We suggest that in *A. kerchovei* we have a core group of populations in the Tehuacán Valley, and peripheric populations that seem to be evolving independently, as shown also by the haplotype network. In this sense, we think that *A. kerchovei* is basically an endemic species from the Tehuacán Valley, and that populations outside the Valley are in the process of incipient speciation.

DATA AVAILABILITY STATEMENT

Sequences were deposited at NCBI GenBank (KX444126–KX444129, KX444111–KX444115, MT513760–MT513762, and MT511772–MT511781).

AUTHOR CONTRIBUTIONS

EA-P contributed to laboratory work, genetic analysis, and drafting the manuscript. JP-L contributed to laboratory work, genetic analysis and the design of some figures. SR-B contributed to fieldwork, data analysis, and helped in drafting and correcting sections of the manuscript, and designing some figures. ES contributed with genetic analysis. RL-S contributed with logistics and ideas for the design of the project. LE project leader, designed and coordinated the project, logistics, drafted and corrected the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by CONACYT Investigación Científica Básica 2011.167826 (clave de identificación oficial CB2011/167826), Genómica de poblaciones: estudios en el maíz silvestre, el teosinte (*Zea mays* ssp. *parviglumis* y *Zea mays* ssp. *mexicana*), by the Project "Conservación de semillas de plantas útiles de San Rafael, Municipio de Coxcatlán, Puebla, MGU-Useful Plants Project México," with the support of the Kew Botanical Gardens and by funding (operating budget) from the Instituto de Ecología, UNAM.

ACKNOWLEDGMENTS

We thank the Laboratorio de Evolución Molecular y Experimental of the Instituto de Ecología. We specially thank, Dra. Laura Espinosa Asuar and Silvia Barrientos for help in the laboratory, and Paulina Brena-Bustamante and Natalia Martínez for assisting in fieldwork, sample collection and providing the photograph (**Figure 2**).

REFERENCES

Brena-Bustamante, P. (2012). *El Aprovechamiento y la Estructura Poblacional de Agave kerchovei LEM. en Tehuacán-Cuicatlán, México*. Masters dissertation, Colegio de Postgraduados, Campus Montecillo, México City.

Brena-Bustamante, P., Lira-Saade, R., García-Moya, E., Romero-Manzanares, A., Cervantes-Maya, H., López-Carrera, M., et al. (2013). Aprovechamiento del escapo y los botones florales de *Agave kerchovei* en el Valle de Tehuacán-Cuicatlán, México. *Bot. Sci.* 91, 181–186.

Byers, C., Maughan, P. J., Clouse, J., and Stewart, J. R. (2014). Microsatellite primers in *Agave utahensis* (Asparagaceae), a keystone species in the mojave desert and colorado plateau. *Appl. Plant Sci.* 2:1400047. doi: 10.3732/apps.1400047

Caballero, M., Lozano-García, S., Vázquez-Selem, L., and Ortega, B. (2010). Evidencias de cambio climático y ambiental en registros glaciales y en cuencas lacustres del centro de México durante el último máximo glacial. *Bol. Soc. Geol. Mex.* 62, 359–377. doi: 10.18268/bsgm2010v62n3a4

Castellanos-Morales, G., Gámez, N., Castillo-Gámez, R. A., and Eguiarte, L. E. (2016). Peripatric speciation of an endemic species driven by Pleistocene climate change: the case of the Mexican prairie dog (*Cynomys mexicanus*). *Mol. Phylogenet. Evol.* 94, 171–181. doi: 10.1016/j.ympev.2015.08.027

Clement, M., Posada, D. C. K. A., and Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9, 1657–1659. doi: 10.1046/j.1365-294x.2000.01020.x

Dávila, P., Arizmendi, M. D. C., Valiente-Banuet, A., Villaseñor, J. L., Casas, A., and Lira, R. (2002). Biological diversity in the Tehuacán-Cuicatlán valley, Mexico. *Biodivers. Conserv.* 11, 421–442.

Delgado-Lemus, A., Torres, I., Blancas, J., and Casas, A. (2014). Vulnerability and risk management of *Agave* species in the Tehuacán Valley, México. *J. Ethnobiol. Ethnomed.* 10:53. doi: 10.1186/1746-4269-10-53

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.

Dupanloup, I., Schneider, S., and Excoffier, L. (2002). A simulated annealing approach to define the genetic structure of populations. *Mol. Ecol.* 11, 2571–2581. doi: 10.1046/j.1365-294x.2002.01650.x

Eguiarte, L., and Scheinvar, E, eds (2008). *Agaves y Cactáceas de Metztitlán: Ecología, Evolución y Conservación*. México City: UNAM, 128.

Eguiarte, L. E., Aguirre-Planter, E., Aguirre, X., Colín, R., González, A., Rocha, M., et al. (2013). From isozymes to genomics: population genetics and conservation of Agave in México. *Bot. Rev.* 79, 483–506. doi: 10.1007/s12229-013-9123-x

El-Ghani, M. M. A., Huerta-Martínez, F. M., Hongyan, L., and Qureshi, R. (2017). *The Deserts of Mexico in Plant Responses to Hyperarid Desert Environments*. Cham: Springer International Publishing, 473–501.

Elith, J., Phillips, S. J., Hastie, T., Dudik, M., Chee, Y. E., and Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* 17, 43–57. doi: 10.1111/j.1472-4642.2010.00725.x

Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998). Base-calling of automated sequencer traces usingPhred. I. Accuracy assessment. *Genome Res.* 8, 175–185. doi: 10.1101/gr.8.3.175

Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5:a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491.

Félix-Valdez, L. I., Vargas-Ponce, O., Cabrera-Toledo, D., Casas, A., Cibrian-Jaramillo, A., and de la Cruz-Larios, L. (2016). Effects of traditional management for mescal production on the diversity and genetic structure of *Agave potatorum* (Asparagaceae) in central Mexico. *Genet. Resour. Crop Evol.* 63, 1255–1271. doi: 10.1007/s10722-015-0315-6

Figueredo-Urbina, C. J., Casas, A., and Torres-García, I. (2017). Morphological and genetic divergence between *Agave inaequidens*, *A. cupreata* and the domesticated *A. hookeri*. Analysis of their evolutionary relationships. *PLoS One* 12:e0187260. doi: 10.1371/journal.pone.0187260

García-Mendoza, A. J. (2011). *Flora del Valle de Tehuacán-Cuicatlán. Fascículo 88. Agavaceae. Universidad Nacional Autónoma de México*. México City: Instituto de Biología, Departamento de Botánica, 95.

García-Mendoza, A. J., Sandoval-Gutiérrez, D., Casas, A., and Torres-García, I. (2019). Agave kerchovei. *IUCN Red List Threat. Spec.* 2019:e.T115644715A116354028.

Gentry, H. S. (1982). *Agaves of Continental North America*. Tucson, AZ: The University of Arizona Press, 670.

Gibson, J. P., Rice, S. A., and Stucke, C. M. (2008). Comparison of population genetic diversity between a rare, narrowly distributed species and a common, widespread species of *Alnus* (Betulaceae). *Am. J. Bot.* 95, 588–596. doi: 10.3732/ajb.2007316

Gordon, D., Abajian, C., and Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res.* 8, 195–202. doi: 10.1101/gr.8.3.195

Granados-Sánchez, D., Sánchez-González, A., Victorino, G., Linnx, R., and Borja de la Rosa, A. (2011). Ecología de la vegetación del desierto chihuahuense. *Rev. Chapingo Ser. Ciencias Forestales y del Ambiente* 17, 111–130. doi: 10.5154/r.rchscfa.2010.10.102

Haffer, J., and Prance, G. T. (2001). Climatic forcing of evolution in Amazonia during the Cenozoic: on the refuge theory of biotic differentiation. *Amazoniana* 16, 579–607.

Hafner, D. J., and Riddle, B. R. (2011). "Boundaries and barriers of North American warm deserts: an evolutionary perspective," in *Palaeogeography and Palaeobiogeography: Biodiversity in Space and Time*, eds P. Upchurch, A. McGowan, and C. Slater (Boca Raton, FL: CRC Press), 75–114. doi: 10.1201/b11176-5

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.

Hamrick, J. L., and Godt, M. W. (1990). "Allozyme diversity in plant species," in *Plant Population Genetics, Breeding, and Genetic Resources*, eds A. D. H. Brown, M. T. Clegg, A. L. Kahler, and B. S. Weir (Sunderland, MA: Sinauer Associates Inc), 43–63.

Hernández, H. M., Gómez-Hinostrosa, C., and Bárcenas, R. T. (2001). Diversity, spatial arrangement, and endemism of Cactaceae in the Huizache area, a hot-spot in the Chihuahuan Desert. *Biodivers. Conserv.* 10, 1097–1112.

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 359, 183–195. doi: 10.1098/rstb.2003.1388

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. doi: 10.1002/joc.1276

Honnay, O., and Jacquemyn, H. (2008). A meta-analysis of the relation between mating system, growth form and genotypic diversity in clonal plant species. *Evol. Ecol.* 22, 299–312. doi: 10.1007/s10682-007-9202-8

Kiehl, J. T., and Gent, P. R. (2004). The community climate system model, version 2. *J. Clim.* 17, 3666–3682. doi: 10.1175/1520-0442(2004)017<3666:tccsmv>2.0.co;2

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187

Loera, I., Ickert−Bond, S. M., and Sosa, V. (2017). Pleistocene refugia in the Chihuahuan Desert: the phylogeographic and demographic history of the gymnosperm Ephedra compacta. *J. Biogeogr.* 44, 2706–2716. doi: 10.1111/jbi.13064

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27(2 Pt 1), 209–220.

Martínez-Ainsworth, N. E. (2013). *Genética de Poblaciones de Agave Stricta* Salm-Dyck, Especie Endémica al Valle de Tehuacán-Cuicatlán, México. Undergraduate dissertation, Universidad Nacional Autónoma de Mexico (UNAM), Mexico City.

Martinez-Rodriguez, A., Macedo-Raygoza, G., Huerta-Robles, A. X., Reyes-Sepulveda, I., Lozano-Lopez, J., García-Ochoa, E. Y., et al. (2019). "Agave seed endophytes: ecology and impacts on root architecture, nutrient acquisition, and cold stress tolerance," in *Seed Endophytes*, eds S. Verma and J. White Jr. (Cham: Springer), 139–170. doi: 10.1007/978-3-030-10504-4_8

Metcalfe, S. E. (2006). Late Quaternary environments of the northern deserts an central trans-volcanic belt of Mexico. *Ann. Mo. Bot. Gard.* 93, 258–274.

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., et al. (2014). ENM eval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological

niche models. *Methods Ecol. Evol.* 5, 1198–1205. doi: 10.1111/2041-210X.12261

Nobel, P. S., and Smith, S. D. (1983). High and low temperature tolerances and their relationships to distribution of agaves. *Plant Cell Environ.* 6, 711–719. doi: 10.1111/1365-3040.ep11589339

Nybom, H., and Bartish, I. V. (2000). Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. *Perspect. Plant Ecol. Evol. Syst.* 3, 93–114. doi: 10.1078/1433-8319-00006

Olson, D., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., et al. (2001). Terrestrial ecoregions of the world: a new map of life on earth: a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience* 51, 933–938.

Olson, D. M., and Dinerstein, E. (2002). The Global 200: priority ecoregions for global conservation. *Ann. Mo. Bot. Gard.* 89, 199–224.

Otto-Bliesner, B. L., Marshall, S. J., Overpeck, J. T., Miller, G. H., and Hu, A. (2006). Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *Science* 311, 1751–1753. doi: 10.1126/science.1120808

Parker, K. C., Trapnell, D. W., Hamrick, J. L., and Hodgson, W. C. (2014). Genetic and morphological contrasts between wild and anthropogenic populations of Agave parryi var. huachucensis in south-eastern Arizona. *Ann. Bot.* 113, 939–952. doi: 10.1093/aob/mcu016

Peterson, T. C., Willett, K. M., and Thorne, P. W. (2011). Observed changes in surface atmospheric energy over land. *Geophys. Res. Lett.* 38:L16707.

Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259. doi: 10.1016/j.ecolmodel.2005.03.026

Pons, O., and Petit, R. J. (1996). Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* 144, 1237–1245.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing,Vienna, Austria.

Ramírez−Barahona, S., and Eguiarte, L. E. (2013). The role of glacial cycles in promoting genetic diversity in the Neotropics: the case of cloud forests during the Last Glacial Maximum. *Ecol. Evol.* 3, 725–738. doi: 10.1002/ece3.483

Rocha, M., Valera, A., and Eguiarte Luis, E. (2005). Reproductive ecology of five sympatric Agave littaea (Agavaceae) species in central Mexico. *Am. J. Bot.* 92, 1330–1341. doi: 10.3732/ajb.92.8.1330

Rzedowski, J., and Huerta, L. (1978). *Vegetación de México*. México City: Editorial Limusa.

Scheinvar, E. (2018). *Filogeografía de Agave Lechuguilla* y *Patrones de Distribución de Agave en México*. Ph.D. dissertation, Universidad Nacional Autónoma de Mexico (UNAM), Mexico City.

Scheinvar, E., Gámez, N., Castellanos−Morales, G., Aguirre−Planter, E., and Eguiarte, L. E. (2017). Neogene and Pleistocene history of *Agave lechuguilla* in the Chihuahuan Desert. *J. Biogeogr.* 44, 322–334. doi: 10.1111/jbi.12851

Shaw, J., Lickey, E. B., Schilling, E. E., and Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am. J. Bot.* 94, 275–288. doi: 10.3732/ajb.94.3.275

Stewart, J. R., Lister, A. M., Barnes, I., and Dalén, L. (2010). Refugia revisited: individualistic responses of species in space and time. *Proc. R. Soc. B Biol. Sci.* 277, 661–671. doi: 10.1098/rspb.2009.1272

Taberlet, P., Gielly, L., Pautou, G., and Bouvet, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* 17, 1105–1109. doi: 10.1007/bf00037152

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.

Tambutti, M. (2002). *Diversidad del Género Agave* en *México*: *Una Síntesis Para su Conservación*. Undergraduate dissertation, Universidad Nacional Autónoma de Mexico (UNAM), Mexico City.

Thiers, B. (2016). *Index Herbariorum: A Global Directory of Public Herbaria and Associated Staff*. Bronx, NY: New York Botanical Garden's Virtual Herbarium.

Thompson, J. D. (1994). CLUSTAL W: improving the sensitivity of progressive sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4553–4559.

Trejo, L., Alvarado−Cárdenas, L. O., Scheinvar, E., and Eguiarte, L. E. (2016). Population genetic analysis and bioclimatic modeling in *Agave striata* in the Chihuahuan Desert indicate higher genetic variation and lower differentiation in drier and more variable environments. *Am. J. Bot.* 103, 1020–1029. doi: 10.3732/ajb.1500446

UNESCO (2012). *World Water Development Report 4–Managing Water Under Uncertainty and Risk*. Paris: UNESCO.

Valiente-Banuet, A., Casas, A., Alcántara, A., Dávila, P., Flores-Hernández, N., del Coro Arizmendi, M., et al. (2000). The vegetation of the Valley of Tehuacan-Cuicatlan. *Bot. Sci.* 67, 25–74.

Valiente-Banuet, A., Solis-Rojas, L., Dávila, P., Arizmendi, M., Pereyra, C., Ramírez, J., et al. (2009). *Guía de la Vegetación del Valle de Tehuacán-Cuicatlán*. México City: Universidad Nacional Autónoma de México.

Vallejo-Marín, M., Dorken, M. E., and Barrett, S. C. (2010). The ecological and evolutionary consequences of clonality for plant mating. *Annu. Rev. Ecol. Evol. Syst.* 41, 193–213. doi: 10.1146/annurev.ecolsys.110308.120258

Vasconcelos, T. S., Rodríguez, M. Á, and Hawkins, B. A. (2012). Species distribution modelling as a macroecological tool: a case study using New World amphibians. *Ecography* 35, 539–548. doi: 10.1111/j.1600-0587.2011.07050.x

Weir, B. S., and Cockerham, C. C. (1984). Estimating F−statistics for the analysis of population structure. *Evolution* 38, 1358–1370. doi: 10.1111/j.1558-5646.1984.tb05657.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 16 July 2020
doi: 10.3389/fpls.2020.01028

Bacterial Diversity and Interaction Networks of *Agave lechuguilla* Rhizosphere Differ Significantly From Bulk Soil in the Oligotrophic Basin of Cuatro Cienegas

Nguyen E. López-Lozano[1*], *Andrea Echeverría Molinar*[1], *Elizabeth Alejandra Ortiz Durán*[2], *Maribel Hernández Rosales*[3†] *and Valeria Souza*[4]

[1] CONACyT-División de Ciencias Ambientales, Instituto Potosino de Investigación Científica y Tecnológica (IPICyT), San Luis Potosí, Mexico, [2] Centro de Física Aplicada y Tecnología Avanzada, Universidad Nacional Autónoma de México, Juriquilla, Mexico, [3] CONACyT-Instituto de Matemáticas, Universidad Nacional Autónoma de México, Juriquilla, Mexico, [4] Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

Due to the environmental conditions presented in arid zones, it is expected to have a high influence of deterministic processes over the community assemblages. Symbiotic interactions with microorganisms could increase colonization and survival of plants in difficult conditions, independent of the plants physiological and morphological characteristics. In this context, the microbial communities associated to plants that inhabit these types of areas can be a good model to understand the community assembly processes. We investigated the influence of stochastic and deterministic processes in the assemblage of rhizosphere microbial communities of *Agave lechuguilla* and bulk soil on the Cuatro Cienegas Basin, a site known for its oligotrophic conditions. We hypothesize that rhizospheric microbial communities of *A. lechuguilla* differ from those of bulk soil as they differ in physicochemical properties of soil and biotic interactions, including not only the plant, but also their microbial co-occurrence networks, it is expected that microbial species usually critical for plant growth and health are more common in the rhizosphere, whereas in the bulk soil microbial species related to the resistance to abiotic stress are more abundant. In order to confirm this hypothesis, 16S rRNA gene was sequenced by Illumina from rhizospheric and bulk soil samples in two seasons, also the physicochemical properties of the soil were determined. Our results showed differences in bacterial diversity, community composition, potential functions, and interaction networks between the rhizosphere samples and the ones from bulk soil. Although community structure arises from a complex interplay between deterministic and stochastic forces, our results suggest that *A. lechuguilla* recruits specific rhizospheric microbes with functional traits that benefits the plant through growth promotion and nutrition. This selection follows principally a deterministic process that shapes the rhizospheric microbial communities, directed by the plant modifications around the roots but also subjected to the influence of

OPEN ACCESS

Edited by:
Jim Leebens-Mack, University of Georgia, United States

Reviewed by:
Ramona Marasco, King Abdullah University of Science and Technology, Saudi Arabia
Eva Stricker, University of New Mexico, United States

***Correspondence:**
Nguyen E. López-Lozano
nguyen.lopez@ipicyt.edu.mx

†Present address:
Maribel Hernández Rosales, Departamento de Ingeniería Genética, Centro de Investigación y de Estudios Avanzados del IPN (CINVESTAV), Unidad Irapuato, Irapuato, Mexico

Specialty section:
This article was submitted to Functional Plant Ecology, a section of the journal Frontiers in Plant Science

Received: *28 November 2019*
Accepted: *23 June 2020*
Published: *16 July 2020*

Citation:
López-Lozano NE, Echeverría Molinar A, Ortiz Durán EA, Hernández Rosales M and Souza V (2020) Bacterial Diversity and Interaction Networks of Agave lechuguilla *Rhizosphere Differ Significantly From Bulk Soil in the Oligotrophic Basin of Cuatro Cienegas. Front. Plant Sci. 11:1028. doi: 10.3389/fpls.2020.01028*

other environmental variables, such as seasonality and soil properties. Interestingly, keystone taxa in the interactions networks, not necessarily belong to the most abundant taxonomic groups, but they have an important role by their functional traits and keeping the connections on the community network.

Keywords: agave microbiome, microbial co-occurrence, keystone species, community assembly, functional traits

INTRODUCTION

Soil microbes represent most of the biodiversity in terrestrial ecosystems and are main contributors to the preservation of soil quality and functioning (Philippot et al., 2013). Within the soil system, the rhizosphere, defined as "the immediate surroundings of the plant root", is a highly dynamic interface, rich in microbial diversity and biomass (Philippot et al., 2013). The rhizosphere harbors complex microbial communities, whose dynamic associations are critical for plant growth and health, since many microorganisms increase nutrient availability (Richardson et al., 2009), contribute to the plant hormonal balance (Spaepen et al., 2007), prevent the attack of plant pathogens by antibiotic and antifungal compounds (Saharan and Nehra, 2011; Doornbos et al., 2012), to give some examples. Interestingly, these relationships are context-dependent, that under some conditions usually mutualistic taxa can become parasitic (Philippot et al., 2013). These complex interactions are mediated by plant-released nutrients, which act as one of the main regulators of microbial diversity, abundance, and activity in the rhizosphere (Sasse et al., 2018; Zhalnina et al., 2018). Despite part of the plant microbiome is determined by vertically transmitted endophytes (Shahzad et al., 2018), the bulk soil, adjacent root-free soil, is the principal source of the species richness in the rhizosphere. By comparing differences in taxonomic and functional traits between rhizosphere and bulk soil communities, we can get insights into how the selection processes operating in the rhizosphere are based in functionality (Mendes et al., 2014; Yan et al., 2017); or there are other factors involved. Although the composition of rhizosphere bacterial assemblages has been extensively studied (Compant et al., 2019), identifying and defining the interactions that occur among soil microorganisms has been little explored, despite its importance to understand microbial diversity and function (Shi et al., 2016).

Several hypotheses have been raised regarding microbial community assembly, it has been proposed that a combination of deterministic and stochastic processes shape microbial communities (Nemergut et al., 2013). Some examples of stochastic processes that produce random patterns in species co-occurrence are dispersal limitation, mass effects, and random demographics (Martiny et al., 2006; Vellend, 2010). In contrast, deterministic processes, are driven by niche partitioning and species interactions, producing segregation or even aggregation of the species (for example in the rhizosphere). The effect of deterministic processes is more notorious when extreme changes in crucial abiotic variables occur, this is the case of such as seasonal variation or changes in soil physicochemical properties (Chave, 2004; Dornelas et al., 2006; Chase, 2007). Due to the environmental conditions presented in arid zones, deterministic processes are expected to be very influential given their natural fluctuations and constant stresses. Arid zones are characterized by a frequent hydric stress, low organic matter, and nutrient content, particularly of nitrogen and phosphorus, as well as high salinity, temperatures, and exposure to UV radiation. Nevertheless, plants living in these habitats are well adapted to survive difficult conditions but independently due to their physiologic and morphological adaptations. It has been suggested that, symbiotic interactions with microorganisms could increase colonization and survival of plants in difficult conditions, independent of the plants physiological and morphological characteristics (Bashan et al., 1995; Puente et al., 2009). In this context, the microbial communities associated to plants that inhabit these arid soils can be a good model to understand assembly processes. *Agave lechuguilla*, is a very common plant in CCB as it is part of the desert scrub in arid and semiarid zones, having the widest natural distribution comparing it with the other Agave species (Gentry, 1982). Usually *A. lechuguilla* is found in rocky soils of limestone origin, it is considered ecologically important because is associated with soil formation and stabilization, and it has been reported as nurse plant for some cactus species (Silva-Montellano and Eguiarte, 2003). In arid rural areas of Mexico, it has economic importance (Castillo Quiroz et al., 2013) as well as potential biotechnological applications (Morales-Luckie et al., 2016). However, its microbial diversity has not been explored yet.

It has been demonstrated that plants in arid ecosystems, including some *Agave* species, are able to reconditioning, their surrounding soil, driving the selection and recruitment of associated microbes in the rhizosphere. This process, generally results in a significantly difference in the diversity, as well as in the microbial interaction networks, among rhizosphere and bulk soil (Desgarennes et al., 2014; Coleman-Derr et al., 2016; Marasco et al., 2018; Mosqueira et al., 2019). One of the factors that can drastically affect microbial communities in arid ecosystems are the climatic variations derived from seasonality. Interestingly, for some *Agave* species only the endophytic microbial communities have shown differences between seasons, but not the rhizospheric microbial communities (Coleman-Derr et al., 2016). However, the *Agave* species analyzed in the literature, generally grow as independent individuals exposed to the sun, and therefore, to evaporation. In contrast, *A. lechuguilla* forms patches of various individuals, which could allow a higher concentration of exudates and humidity, we believe, that, as a consequence, their rhizospheric effect could be enhanced. In this scenery, rhizospheric conditions could be more favorable for microbial

communities having a greater amount of C sources and humidity that ameliorate the hard conditions during dry season, in contrast bulk soil is more susceptible to desiccation and high temperatures.

The Cuatro Ciénegas Basin (CCB) is located in the Chihuahuan Desert of Mexico. This place has an arid climate with two outstanding seasons: dry season from November to April and rainy season from May to October. It is geographically isolated and presents highly spatial heterogeneity because of its topology and irregular availability of water (Meyer, 1973). In addition to this, it is characterized by haline and oligotrophic soils (López-Lozano et al., 2012; Tapia-Torres et al., 2015; Pajares et al., 2016). These features caused an enormous diversification and a considerable amount of endemism, being reported as the place with the greatest number of endemism in North America, including macroorganisms and microorganisms (Souza et al., 2006; Alcaraz et al., 2008; Cerritos et al., 2008; Desnues et al., 2008; Escalante et al., 2009). *A. lechuguilla*, is a very common plant in CCB as it is part of the desert scrub in arid and semiarid zones, having the widest natural distribution comparing it with the other *Agave* species (Gentry, 1982). Usually, *A. lechuguilla* is found in rocky soils of limestone origin, it is considered ecologically important because it is associated with soil formation and stabilization, and it has been reported as nurse plant for some cactus species (Silva-Montellano and Eguiarte, 2003). In arid rural areas of Mexico, it has economic importance (Castillo Quiroz et al., 2013) as well as potential biotechnological applications (Morales-Luckie et al., 2016). However, *A. lechuguilla* associated microbial diversity has not been explored yet.

According with the aforementioned ideas, we investigated the influence of stochastic and deterministic processes in the assemblage of rhizospheric microbial communities of *A. lechuguilla* and bulk soil on the Cuatro Cienegas Basin. We hypothesized that rhizospheric microbial communities of *A. lechuguilla* differ from those of bulk soil as they differ in physicochemical properties of soil and biotic interactions. These effects will include not only the plant, but also their microbial co-occurrence networks. Under this scenario, it is expected that microbial species usually critical for plant growth and health are more common in the rhizosphere, whereas in the bulk soil microbial species related to the resistance to abiotic stress are more abundant.

MATERIALS AND METHODS

Sample Collection

Four sites at the Cuatro Cienegas Basin (CCB) were selected by the presence of *A. lechuguilla* populations: "Becerra" — 26° 52.758′ N, 102° 08.19′W; "Carranza" — 26°59.519′ N, 102° 02.741′ W, "Orozco" — 26° 54.478′ N, 102° 07.169′ W, and "Madera" — 26° 57.609′ N, 102° 10.523′ W (**Figure S1**). The vegetation in the four sites was xerophytic scrub dominated by individuals of *A. lechuguilla, Larrea tridentata,* and some cacti species (**Figures S4A, B**). In each site, a sampling unit of 8 m × 8 m was placed, inside three independent adult individuals of *A. lechuguilla* with at least 30 leaves and without signs of flowering (Silva-Montellano and Eguiarte, 2003), and three vegetation-free interspaces (bulk soil) were selected randomly (**Figures S4C, D**). Two samplings from the rhizosphere of *A. lechuguilla* and bulk soil at a depth of 10 cm were carried out, one in the dry season (March) and the other one in the rainy season (October) of 2016. The rhizosphere is very difficult to delimit by definition (Philippot et al., 2013), for this reason we considered the soil attached to the roots for molecular analyzes and also used the soil surrounding the roots for physicochemical analysis. For this sampling, four equidistant points were located around the root system for a better representation of the rhizosphere, it was dug to a depth of 10 cm looking for the roots, then sub-samples were taken from the fraction that remains attached to the roots with a sterile spatula. The sub-samples were homogenized, and approximately 500 mg of soil were taken and stored in 2-ml Eppendorf tubes, 1 ml of DNA/RNA Shield™ was added for the nucleic acid preservation. The samples were kept on ice until their transportation to the laboratory and then stored at −80°C until its analysis. At the same time, we took as much soil as possible that was adhered to the roots, as well as the soil surrounding them was collected (approximately 500 g), to characterize the soil physicochemical properties, this sample was homogenized and was considered also as "rhizospheric soil." These soil samples were kept in dark plastic bags, sealed, and stored at 4°C until its posterior analysis. As a result, in total 48 samples were collected, both for the microbial community analysis and soil physicochemical properties characterization.

Soil Physicochemical Properties Analyses

Soil samples were oven-dried at 60°C, after that water content was determined using the gravimetric method. Soil pH and electric conductivity (EC) were measured in deionized water (soil/solution ratio, 1:2 w/v). Organic matter (OM) was determined by the calcination method (Schulte and Hopkins, 1996). Total carbon (TC) and total nitrogen (TN) were quantified with an elemental combustion system (4010, Costech Analytical Technologies, Valencia, CA). To measure soluble Phosphorous (P) and cations (Na+, K+, Ca2+, Mg2+) an inductively coupled plasma optical emission spectroscopy was used (ICP-OES) (730-ES, Varian, Palo Alto, CA). Finally, Ammonium (NH4+) and nitrate (NO3-) concentrations were determined using colorimetric methods (Forster, 1995; Miranda et al., 2001; Doane and Horwáth, 2003).

Characterization of Microbial Communities

DNA was extracted using ZR Soil Microbe Miniprep™ kit (Zymo Research, Irvine, CA) following manufacturer's protocol, using 350 mg of soil. For the microbe characterization, the V3-V4 regions of the 16S rRNA gen were amplified using the primers 357 forward (5′-CTCCTACGGGAGGCAGCAG-3′) and CD reverse (5′-CTTGTGCGGGCCCCCGTCAATTC-3′). Illumina Miseq

(Illumina, San Diego, CA) method was used for sequencing in pair-end 2 bp × 300 bp format at the LANGEBIO (Laboratorio Nacional de Genómica para la Biodiversidad).

The reads were assembled and filtered according to their quality (value of Q = 25, length = 300 bp, elimination of primers, barcodes, chimeras, and homopolymers≥ 8) using Mothur 1.36.1 (Schloss et al., 2009). Chimeras were identified using VSEARCH algorithm (Rognes et al., 2016) and removed. Sequences were classified using SILVA 123 (www.arb-silva.de) reference data base. The sequences that were classified as chloroplast, mitochondria, archaea or unassigned were removed as well as the singletons. In addition, to determine the diversity of the microbial communities, OTUs with a 97% of similarity were generated to calculate Chao1 estimator and Shannon index. A total of 6,837,394 reads were analyzed, after quality filtering, an average of 142,445 sequences were obtained for each library (**Table S1**). To avoid the bias of different depths of sequencing in the alpha and beta diversity measurements, the libraries were normalized to the same size (60,618 sequences) based on the library with the lowest number of sequences.

Based on literature review we identify possible changes in the relative abundance of functional guilds between soil types and seasons, using only the genera that were classified taxonomically with a degree of certainty greater than or equal to 80% of bootstrap according to the Naïve Bayesian Classifier (Wang et al., 2007). The genera classified as candidates or for which there is no information available in the literature constituted ~5% of the total community. We grouped the taxa into categories according to the presence of metabolic pathways that participate in the cycling of nutrients (considering the possession of genes involved in the trait or detected activity under laboratory conditions), its ability to tolerate stress or adapt to the environment (**Table S7**). The relative abundance of each functional category was calculated by adding the number of sequences of the genera within that category. A comparison between the treatments was made using ANOVA.

Statistical Analyses

Shapiro-Wilk test was used to test the normality of the data. In case it does not have a normal distribution, data transformation was applied. Auto-correlation analysis of the physicochemical properties of the soil, was performed using Pearson method. However, none of the properties presented a significant auto-correlation. Afterwards, to look for differences between soil type (rhizosphere or bulk soil) and season, as well as the interaction between both variables, a two-way ANOVA was used. To verify the model fit, the normality assumptions of the residuals were assessed. As post-hoc evaluation for those properties presenting significant differences, Tukey test was applied. These analyses were carried out with using R software v 3.4.0., which was used in all posterior analyses. Ordination analysis was carried out using Non-Metric Multidimensional Scaling (NMDS) method based on Bray Curtis dissimilarity index, for this all analysis, only genera with relative abundance greater than 1% were considered. An environmental fit analysis of abiotic variables was applied (physicochemical properties of soil). These calculations were carried out with vegan package and the graphics with ggplot2 package. Heatmaps were constructed using gplots package, the clustering analyses on them were based in Bray Curtis dissimilarity index. PERMANOVA analyses was performed using the function adonis2 in vegan package.

Interaction Networks

The inference of interaction networks was done using the software MetaMIS (Shaw et al., 2016) that uses a time-series Lotka-Volterra approach. Using the raw abundance data, we obtain the underlying interactions among the OTUs found in Agave rhizosphere and Bulk soil. The consensus networks inferred by MetaMIS are directed networks that include positive and negative interactions.

In order to analyze the interactions among the OTUs, we made some improvements to the software NetAn (https://github.com/valdeanda/NetAn) first used in (De Anda et al., 2018) and used it to obtain the global features of the interaction networks. NetAn is a network analysis tool programmed in Python, which extracts the following properties: order, size, density, diameter, radius, clustering coefficient, mean degree, centrality, hubs, vertex, and edge min-cut sets, connected components, cycles, maximal independent sets, degree distribution, modularity, communities. Some of these properties, such as diameter and radius, are calculated over the underlying undirected network.

We identified key features that show to be significant in comparison with random networks. NetAn generates hundred random networks with the same size and order of the ecological network, then extracts all the properties mentioned above and averages them. The values of each property were the compared to that of the ecological network, and we focus on the analysis of such features that are significant.

We also used Mfinder (Milo et al., 2002) to identify network motifs that can play an important role as building blocks of these networks, however, no network motif showed to be significant, telling us about the independency of the taxa in the network.

RESULTS

Physicochemical Properties of Soils

ANOVA results indicated that water content and pH presented significant differences ($P < 0.05$) between Agave rhizosphere and bulk soil samples; Ca2+, NH4+, TN, and C/N ratio were different between seasons; NO3-, EC, OM, K+, Mg2+ and P were different in both conditions (season and soil type). Na+ and TC did not present significant differences in any condition (**Table 1**).

Bacterial Composition and Diversity

Diversity Shannon-Wiener index presented significant differences ($P < 0.05$) between rhizosphere and bulk soil conditions, diversity was higher on the rhizosphere compared to bulk soil (6.88 ± 0.2 vs. 6.74 ± 0.2). There was not any significant difference between seasons. **Table S1** provides the data of observed OTUs, Chao1

TABLE 1 | Average and standard error values obtained from physicochemical properties of *Agave lechuguilla* rhizosphere and bulk soil in the interspaces.

Physicochemical property	Dry season		Rainy season		*F* value		
	Rhizosphere	Bulk soil	Rhizosphere	Bulk soil	Season	Soil type	Interaction
Moisture (%)	14.85 ± 3.12^{a}	7.65 ± 2.41^{b}	12.67 ± 2.82^{ac}	11.04 ± 2.77^{c}	0.643	30.643***	12.270**
pH	7.72 ± 0.31^{a}	7.95 ± 0.34^{ab}	7.87 ± 0.19^{ab}	8.03 ± 0.19^{b}	2.218	6.597*	0.205
Ca^{2+} (mg/kg)	1037.11 ± 78.90^{a}	1004.68 ± 34.21^{a}	1987.62 ± 148.62^{b}	1917.81 ± 120.85^{b}	945.317***	2.845	0.38
NH_4^+ (mg/kg)	3.04 ± 0.83^{a}	2.06 ± 0.89^{a}	13.55 ± 1.81^{b}	14.10 ± 2.95^{b}	454.684***	0.162	2.093
TN (mg/kg)	1.70 ± 0.49^{a}	1.07 ± 0.43^{a}	1.76 ± 1.02^{a}	4.49 ± 2.67^{b}	12.64***	1.644	18.491***
C/N	17.74 ± 7.72^{a}	30.76 ± 2.61^{a}	25.80 ± 6.50^{a}	7.92 ± 2.30^{b}	14.191***	2.316	15.935***
NO_3^- (mg/kg)	25.95 ± 6.49^{a}	18.21 ± 10.76^{b}	49.48 ± 13.73^{c}	28.81 ± 7.20^{a}	33.075***	22.616***	1.408
EC (μS/cm)	322.31 ± 88.15^{a}	224.34 ± 75.50^{bc}	232.97 ± 47.57^{b}	179.02 ± 20.07^{c}	11.010**	18.968***	0.827
OM (%)	3.61 ± 1.69^{ab}	1.98 ± 0.62^{a}	3.98 ± 1.11^{b}	2.94 ± 0.73^{ab}	10.18**	21.09***	2.43
K^+ (mg/kg)	186.22 ± 56.60^{ac}	125.15 ± 42.49^{a}	305.90 ± 81.11^{b}	228.87 ± 49.43^{c}	42.682***	16.309***	0.218
Mg^{2+} (mg/kg)	102.34 ± 34.14^{ac}	70.09 ± 12.05^{a}	164.83 ± 32.35^{b}	134.34 ± 26.09^{c}	63.449***	15.554***	0.012
P (mg/kg)	0.17 ± 0.09^{a}	0.11 ± 0.05^{a}	0.80 ± 0.12^{b}	0.61 ± 0.09^{c}	445.656***	20.865***	5.534*
Na^+ (mg/kg)	6.79 ± 2.34^{a}	6.56 ± 1.48^{a}	8.44 ± 2.61^{a}	7.13 ± 1.34^{a}	5.71	0.731	0.699
TC (mg/kg)	27.48 ± 7.43^{a}	29.48 ± 6.21^{a}	26.35 ± 11.85^{a}	21.54 ± 9.85^{a}	2.975	0.286	1.677

Letters beside standard deviation indicate significant differences.
$^{***}p \leq 0.001$, $^{**}p \leq 0.01$, $^{*}p \leq 0.05$.

estimator, Shannon-Wiener, and Simpson index values for every sample, and **Table S2** provides the results of the two-way ANOVA.

The most abundant phyla in all samples were Acidobacteria, Actinobacteria, Chloroflexi, and Proteobacteria. At this level, there was not so much variation on the relative abundance between seasons, except for Firmicutes which increase its relative abundance in rainy season (**Table S5**). Between rhizosphere and bulk soil Actinobacteria were more abundant in bulk soil, unlike the Proteobacteria which were more abundant in the rhizosphere (**Figure 1**, **Table S5**). The analysis at genus level of the most abundant phyla (considering only the genera with a relative abundance greater or equal to 1% in at least one sample, **Figure S2** and **Table S5**) revealed that *Rhizobiales FFCH5858,* and *Kallotenuales AKIW781* were significantly ($P < 0.05$) more abundant in the dry season, and *Actinobacteria TakashiACB11* and *Acidobacteria S6* increased its relative abundance in rainy season. According to the soil type, *Sphingomonas, Microvirga, Rhizobiales JG34KF361, Thermomicrobia JG30KFCM45* and *Bryobacter* increased their relative abundance in the rhizosphere; and *Chloroflexi TK10*, *Kallotenuales AKIW781*, *Actinobacteria TakashiACB11*, *Rubrobacter,* and *Actinobacteria MBA2108* had a preference for bulk soil (**Table S6**).

As a result of the analysis based on literature review about the functional capacities of the taxa identified, significant changes between conditions in the relative abundance of some categories were observed (**Figure 2**, **Table S4**). We found a higher relative abundance of N fixers in the rhizosphere of *A. lechuguilla* than in the bulk soil, represented principally by *Sphingomonadales* and *Rhizobiales*. In contrast, denitrifiers and P solubilizers were both more abundant in bulk soil. Particularly, denitrifiers were primarily represented by the genera *Bacillus, Streptomyces, Euzebya,* and *Bosea*, whereas P solubilizers were represented by members of the family *Gaiellaceae*. In general, the groups whose stress tolerance has been reported were more abundant in the bulk soil. Regarding traits of environmental adaptation, genera with the reported capacity of antibiotic production and filamentous formation were more abundant in the bulk soil, while genera that produce exopolysaccharides and the capacity of nutrients storage (formation of polyhydroxyalkanoates and polyphosphates inclusions) were more abundant in the rizosphere.

The results of the NMDS analysis using the data from the bacterial composition at genus level (**Figure 3**) show that the most remarkable difference corresponds to the soil type and that the bulk soil community is more sensitive to variation caused by the season. Additionally, making the adjustment of physicochemical properties data is observed that pH, EC, soil moisture, OM, and C/N ratio are the properties that show a significant relationship with the composition of the microbial community.

Interaction Between Bacteria

To analyze the changes in the ecological interaction patterns between conditions, interactions networks were generated based in the relative abundance of the identified genera. In general, density and clustering coefficients of the networks were low, indicating that these communities have few strong interactions (**Table 2**, **Figure 4**), having a tendency to form few modules, even less than those formed in random networks. The interaction networks are not dense and show that the degree distribution follows a power law distribution, common to complex networks. Moreover, in these networks, there are just couple of nodes (and sometimes only one) that plays an important role, being the genera with more positive and negative interactions, which is very unlikely in random networks. The number of communities (modules) was between 5 and 8, and the number of nodes in each module varied between 3 and 15 (**Table S3**). The clustering coefficient was slightly higher in the rhizosphere than in the bulk soil networks when the data was separated by season.

Interestingly, there was a high number of independent sets of genera that interact with the rest of the community through a single hub. The existence of these single hubs that connect a

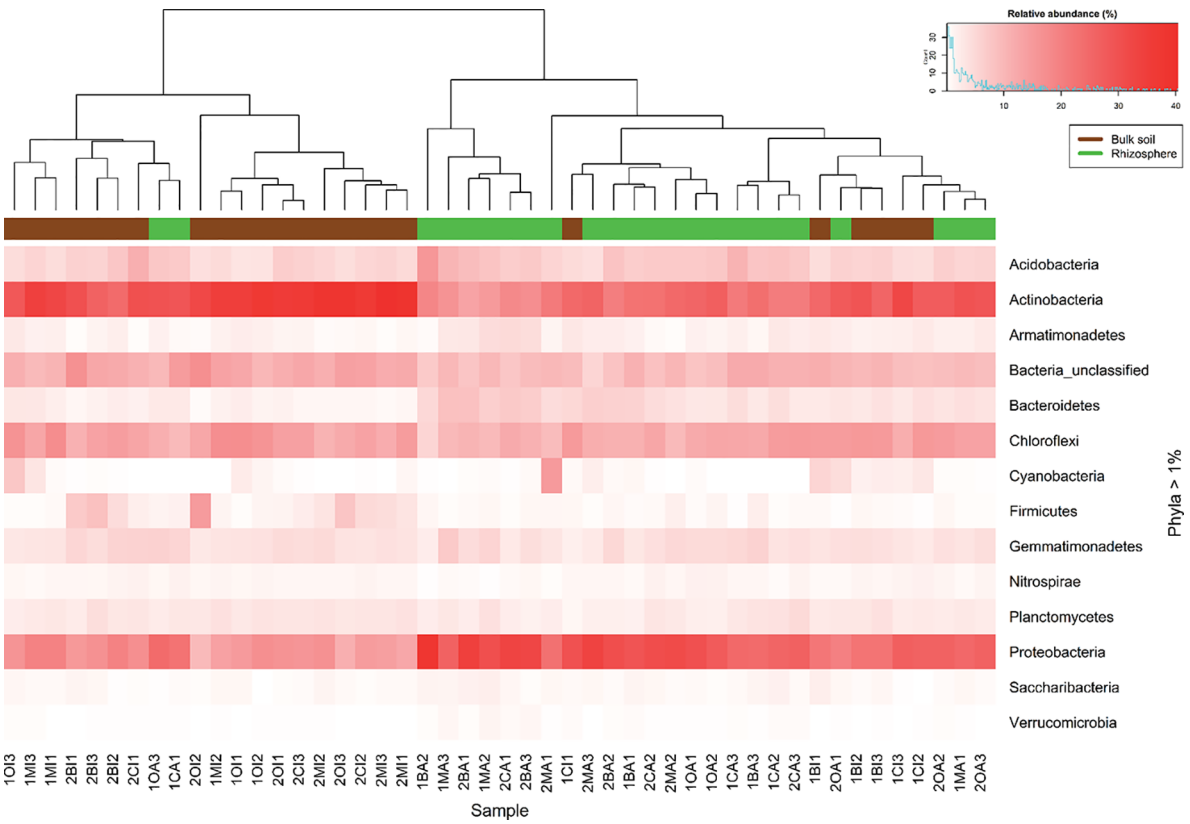


FIGURE 1 | Heatmap constructed using the relative abundance of the identified phyla in all samples. The intensity of red color represents relative abundance. Superior dendogram is the result of the clustering analysis using Bray Curtis dissimilarity index, each sample is colored according to their origin, rhizosphere (green) or bulk soil (brown).

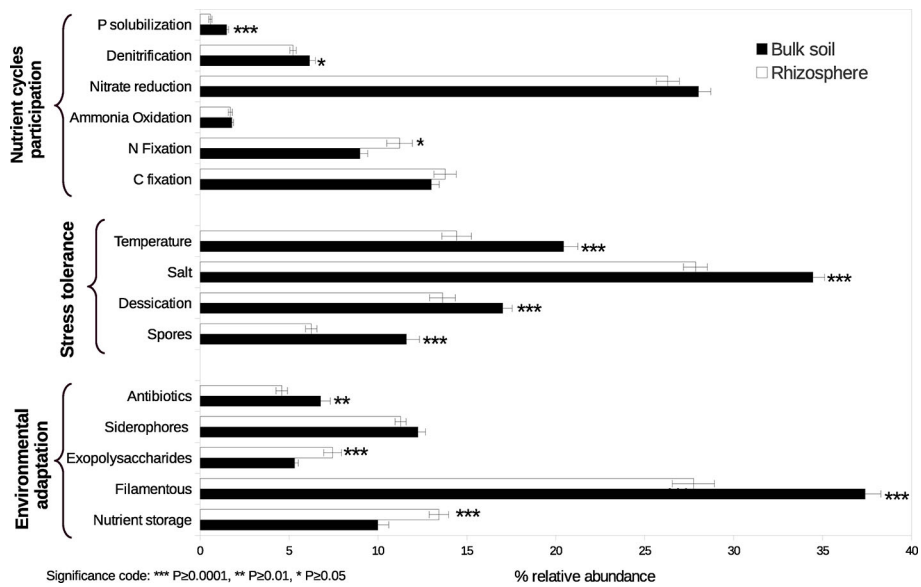


FIGURE 2 | Relative abundance of functional guilds participating in nutrient cycles, traits conferring stress tolerance, and traits conferring environmental adaptation in rhizosphere and bulk soil. Asterisks indicate significant differences. Bars represent standard errors.

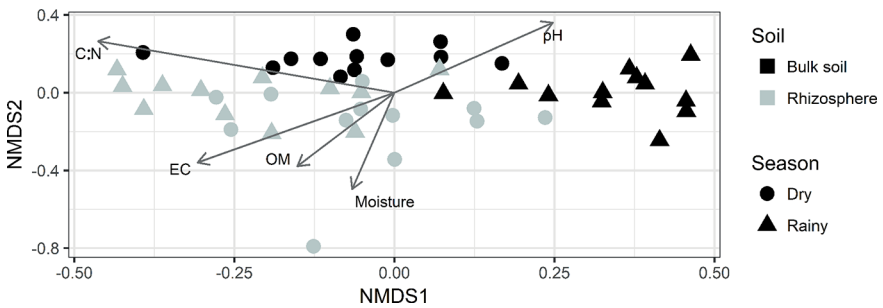


FIGURE 3 | NMDS analysis using relative abundance at a genus level. Arrows indicate the physicochemical properties that resulted in a significant relationship with the microbial community.

TABLE 2 | Global network measures obtained from real (consensus) and random networks.

	Both soils together		Rhizosphere			Bulk soil		
	Dry	Rainy	Both seasons together	Dry	Rainy	Both seasons together	Dry	Rainy
Total nodes	61	61	62	53	54	59	74	71
Density	0.017	0.022	0.020	0.023	0.030	0.021	0.021	0.029
Clustering Coefficient	0.000	0.016	0.023	0.039	0.037	0.028	0.012	0.018
Mean degree	1.017	1.151	1.017	1.104	1.256	1.093	1.233	1.420
Modularity	0.711	0.607	0.665	0.466	0.456	0.644	0.584	0.522
No. of modules	8	7	8	5	8	8	7	6

module with the rest of the network, could be due to the fact that OTUs within a genus were merged together. This is very interesting, since it indicates the importance of these genera in the community. In this sense, it would be very easy to split apart these networks by removing only one node or one connection (one genus), which tells us about the vulnerability of the network. If environment perturbations affect the network, it will be very likely that the interactions that maintain the dynamics of the network would be lost. It is important to note, that there are some networks which are disconnected, i.e. a network formed by a set of subnetworks, where only one of them contains most of the genera, and the others have only few members (from 2 to 7). In this study, we focused mainly in such subnetworks with 5 or more nodes. In each subnetwork is identified a connected component. Even though, most of the genera do not share interactions with other members in the community, as we can see by studying the independent sets, it is surprising that they are still able to form communities. Members of a community have more interactions among them, that with members outside the community. These communities are formed by a keystone hub and a set of genera which have no interactions among them. Since these important hubs could represent "key species" for the community structure, these strongly interconnected taxa or hubs in the networks were identified and the directionality of the interactions that they establish with the other members in the community was determined (**Figure 4**, **Table 3**). The hubs were different between type of soil and season. In the rhizosphere, *Saccharibacteria_unclassified, Bacillus,* and *Anabaena* were the genera with the highest number of interactions, being also the hubs with maximal out-degree (exert an effect on the other groups). *Bacillus* was the principal hub in the rhizosphere during dry season and Anabaena in the rainy season. Besides, *Ensifer, Bosea,* and *Anabaena* were the taxa with the maximal in-degree (the most affected by other members of the community). In the bulk soil, *Plantactinospora, Streptomyces,* and *Pontibacillus* were the genera with the highest number of interactions for all data, dry and rainy seasons, respectively. The hubs with the maximal out-degree were *Craurococcus, RB41_unclassified* and *Pontibacillus,* for all data, dry and rainy seasons, respectively. *Chroococcidiopsis* was the most affected by other members in the community for all data, while two genera (*Janibacter* and *Isoptericola*) and *Streptomyces* were the hubs with max-in degree in the dry and rainy seasons, respectively.

DISCUSSION

In the present investigation, the soil properties with the highest influence over the composition and structure of the microbial communities were pH, EC, moisture, OM, and C/N ratio. The soil in CC is alkaline with elevated concentrations of ions, resulting in high electrical conductivity. This characteristic can be attributed to the gypsum-rich nature of the CC soils. In previous studies, salinity has demonstrated a strong influence on the composition of the microbial communities in this Basin, not only because it is a selecting pressure by itself, but also because it decreases the availability of nutrients, which are very scarce in these soils (López-Lozano et al., 2012; Pajares et al., 2016).

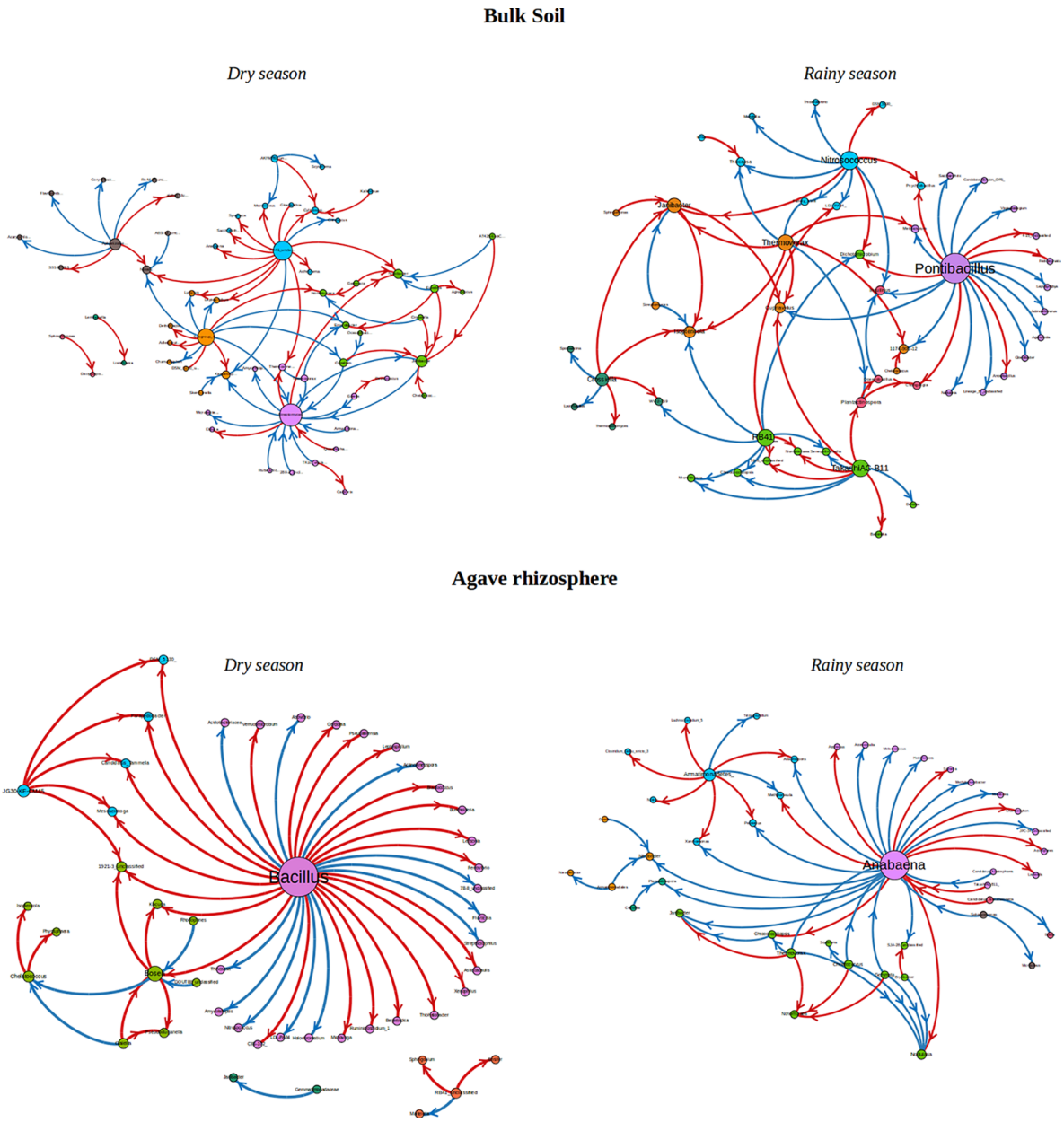


FIGURE 4 | Interaction networks found for Rhizosphere and bulk soil over two seasons. The size of the nodes represents the importance in the network, i.e. the bigger the node is the more connections it has with other OTUs. Red rows are negative interactions, while blue rows are positive interactions. The color of each node represents the community to which it belongs.

Coupled with this, the low nutrient availability to soil microbes and vegetation, observed in the low C/N ratio, suggests deficiency of soil organic C that can limit the N cycle. Despite the great heterogeneity of this arid environment, the oligotrophic conditions seem to be a general characteristic of the soils in all the Basin (López-Lozano et al., 2012; Tapia-Torres et al., 2015; Pajares et al., 2016).

The bacterial composition found in our samples clearly reflect their origin, since many of the identified taxa have been reported in arid zones (Walker and Pace, 2007; Bachar et al., 2010; Saul-Tcherkas and Steinberger, 2011; Andrew et al., 2012; López-Lozano et al., 2012; Marasco et al., 2012; Torres-Cortés et al., 2012; An et al., 2013; Kavamura et al., 2013; Desgarennes et al., 2014; Coleman-Derr et al., 2016; Postma et al., 2016). However, the presence of many bacterial groups can be related to the particular characteristics of CCB soils. For example, Acidobacteria (Fierer et al., 2007; Kielak et al., 2016), Armatimonadetes (Lee et al., 2014), Verrucomicrobia

TABLE 3 | Strongly interconnected taxa or hubs identified at genus level and their number of interactions in the community networks.

	Both soils		Rhizosphere			Bulk soil		
	Dry	Rainy	both seasons	Dry	Rainy	both seasons	Dry	Rainy
Hubs with Max Total Degree	*Janibacter, Bacillus*	*Anabaena*	*Saccharibacteria_unclassified, Bacillus*	*Bacillus*	*Anabaena*	*Plantactinospora*	*Streptomyces*	*Pontibacillus*
Max Total Degree	10	12	9	33	28	14	16	19
Hubs with Max In Degree	*Janibacter*	*Chroococcidiopsis, Asteroleplasma*	*Ensifer*	*Bosea*	*Anabaena*	*Chroococcidiopsis*	*Streptomyces*	*Janibacter, Isoptericola*
Max In Degree	8	8	5	4	7	7	10	5
Hubs with Max Out Degree	*Bacillus*	*Nonomuraea*	*Saccharibacteria_unclassified, Bacillus*	*Bacillus*	*Anabaena*	*Craurococcus*	*RB41_unclassified*	*Pontibacillus*
Max Out Degree	10	10	9	33	21	7	14	19

(Da Rocha et al., 2010; Senechkin et al., 2010; Bergmann et al., 2011), and Gemmatimonadetes (Zhou et al., 2007; Cockell et al., 2009) are commonly associated to oligotrophic places. In special, the phylum Gematimonadetes has been recognized for their high tolerance to desiccation and its ability to store phosphorus (DeBruyn et al., 2011).

Our results showed differences in bacterial diversity, community composition, potential functions, and interaction networks between the rhizosphere samples and the ones from the bulk soil, which indicate a rhizospheric effect. Higher Shannon index in the rhizosphere of other Agave species (*Agave salmiana, Agave tequilana,* and *Agave deserti*) than in bulk soil had been previously observed (Coleman-Derr et al., 2016), as well as differences in community composition (Desgarennes et al., 2014; Coleman-Derr et al., 2016). Even though these studies were focusing on differences between species and compartments of the plant (phyllosphere, endosphere, and rhizosphere), they found also significant differences between the microbial community of the zone near to the root and the microbial community of bulk soils. Despite *A. lechuguilla* is the smallest species within the genus Agave, this rizospheric effect was clearly observed. The plant presence improves the conditions for some groups of microorganisms; therefore, it changes the composition and structure of microbial communities in soil. Similar to our observations, some studies had found an increment in the relative abundance of Proteobacteria in the rhizosphere, showing a positive correlation with moisture (Marasco et al., 2012; Kavamura et al., 2013). While in bulk soil there is higher relative abundance of Actinobacteria, negatively correlated also with moisture (Saul-Tcherkas and Steinberger, 2011; Marasco et al., 2012; Kavamura et al., 2013). At genus level, for example *Bryobacter* of Acidobacteria phylum was more abundant in the rhizosphere, this might be because these microorganisms can obtain carbon from the decomposition of organic matter and nitrogen from nitrate (Kulichevskaya et al., 2010). Another example is *Mycobacterium*, which are found in sites rich in organic matter, humid, and even in outer layers of plant tissues (Kazda et al., 2010). Some of the identified bacteria are well known because they form associations with plants, like *Bradyrhizobium* and *Rhizobia* (Cararetto et al., 2014; De Souza et al., 2014) and some others carry out nitrate to nitrite reduction like *Microvirga* (Kelly et al., 2014), *Craurococcus* (Saitoh et al., 1998) and *Microlunatus*; this last one has been reported as phosphorus accumulator too (Nakamura et al., 1995). On the contrary, other groups have adaptations that could explain their presence in higher proportions in the bulk soil of CC. For example, some members of Actinobacteria phylum are spore forming, have tolerance for dry conditions, high salinity, and high temperatures (Chen et al., 2004; Goodfellow et al., 2012). *Euzebya* needs salt to grow (Kurahashi et al., 2010), *Thermoleophilum,* and *Kallotenue* have tolerance to high temperatures as well (Zarilla and Perry, 1984; Cole et al., 2013). This last one group and *Ardenticatena* are also filamentous bacteria, characteristic that allows them to retain humidity and extend through the soil matrix to get more nutrients (Cole et al., 2013; Kawaichi et al., 2013).

Furthermore, analysis of the functional capacities of the taxa identified (**Table 2**), found a higher relative abundance of bacteria with the potential to fix N in the rhizosphere of *A. lechuguilla* than in bulk soil, represented principally by *Sphingomonadales* and *Rhizobiales.* Free-living N fixers that are associated with the rhizosphere of non-symbiotic plants can represent the principal source of new N in many ecosystems, due to the lack of large populations of plants in symbiosis with N fixers (Reed et al., 2011). It is well known that N fixation is a high energy-demanding process, for this reason the lower organic matter in bulk soil can be limiting the reducing power needed to fix N.

Other functional groups that showed differences in relative abundance between rhizosphere and bulk soil were P solubilizers and denitrifiers, both being more abundant in the bulk soil. P solubilizers were represented by two unknown genera of the family *Gaiellaceae*, a group in which some species possess alkaline and acid phosphatases (Albuquerque and da Costa, 2014). However, without a better description of isolated strains within these genera is not possible differentiate if their presence is due to habitat preference or to their functional capacities. With regard to denitrifiers, multiple factors have been shown to influence the activity and abundance of this diverse functional guild, including pH (Baggs et al., 2010), soil texture (Gu et al., 2013), organic matter (OM) (Barrett et al., 2016), and inorganic

N (Niboyet et al., 2009). Our soil samples did not show significant differences in pH or texture, but had higher amounts of OM and nitrate in bulk soil. The higher relative abundance of denitrifiers observed in bulk soil may be due to a strong competition for nitrate between plants and denitrifiers in the rhizosphere.

In addition to the effect of the plant and physicochemical properties, other studies reported changes in the composition of soil microbial communities between seasons (Aguilera et al., 1999; Saul-Tcherkas and Steinberger, 2011; Lançoni et al., 2013). In this research, although the season factor was not significant for all the samples in the environmental fit perform over the NMDS analysis, in **Figure 3**, a differentiation between dry and rainy season was observed for the samples belonging to the bulk soil, this could be because the conditions in *Agave* rhizosphere remained more stable, avoiding the effect of the season over microbial communities. In a previous study in *Agave* plants, the influence of the season was only found in the microbial communities from the endosphere but not in the rhizosphere (Coleman-Derr et al., 2016). Despite the fact that soil samples used for physicochemical analyzes, came from a larger fraction of rhizospheric soil, than those taken for the molecular analyzes (which were closer to the root, hence more influenced by the plant). In this study, it was possible to observe a clear differentiation between both types of soil and seasons, as well as significant correlations between these properties and the relative abundance of some bacterial groups.

Our data show that soil type (rhizosphere or bulk soil) selects differentially bacterial traits, suggesting that abiotic filtering plays a significant role in this kind of ecosystems affected by hostile conditions (high soil electrical conductivity, water-limitation, low nutrients availability, etc.) (Goberna et al., 2014). The bulk soil exposed to high temperature fluctuations and radiation had more bacterial traits associated with tolerance to environmental stress, filtering bacteria tolerant to temperature, desiccation, salt, and with the capacity to form resistant structures (exospores, other spores, and cysts). In contrast, the "high" productive environment created in the *Agave* rhizosphere facilitates the nutrient storage and the exopolysaccharides production as competitive abilities (Goberna et al., 2014).

Our networks analysis reveals that interactions among bacterial genera followed a power-law distribution, this structure is widespread in many real-world (Internet, social, and biological) networks (Adamic and Huberman, 2000). Interestingly, we found that networks presented low complexity, characteristic that can be attributed to a dynamic community, highly variable by season, with a great number of taxa with no common co-occurrence across sites. This behavior can be the result of the great heterogeneity of the soils in the Basin, raveling the influence of stochastic processes in the assembly of these communities, another possibility is that our sampling size was too coarse to identify micro-scale covariations. However, at the scale of our sampling, the interactions detected must be strongly preserved to be distinguished. The identified modules within the networks are likely result of interactions or covariations between bacterial genera in response to shared niches in the rhizosphere and bulk soil. Topologically, we also identified network hubs and connectors, because they can function as keystone taxa that maintain the network structure (Faust and Raes, 2012). These keystone taxa are relatively more important than other taxa in the network, their loss may cause modules and networks to disassemble (Jordán et al., 2008), and thus their presence can be transcendental to maintain ecosystem stability (Sole and Montoya, 2001).

In the rhizosphere, the hubs most affected by other members of the community are involved in nutrient acquisition or enhance bioavailability of soil nutrients, *Bosea* and *Ensifer* are well-known N fixers and denitrifiers (Shinwari et al., 2019) and *Anabaena* is a filamentous Cyanobacteria with the capacity to fix N (Manjunath et al., 2016). We hypothesize that they could be subject to predation. Besides, the hubs that affect other members of the community in the rhizosphere were identified as *Saccharibacteria_unclassified, Bacillus,* and *Anabaena*. The phylum candidatus Saccharibacteria was formerly known as Candidate Division TM7 (Kindaichi et al., 2016). There are very few strains of *Saccharibacteria* isolated and characterized, for this reason, the information about their potential functions is limited. However, there are evidence that *Saccharibacteria* is related to wilt disease suppression (Zhang et al., 2016; Shen et al., 2018), but how it is involved in potential disease control remains unknown. Moreover, *Bacillus* spp. is considered to be safe microorganisms for the plants, having potent plant growth promoting traits such as IAA (indol acetic acid) production, nitrogen fixation, phosphate solubilization, and biocontrol attributes like production of hydrolytic enzymes, HCN (cyanogen), siderophores, and antibiotics (Kumar et al., 2012). Besides, *Anabaena* and other Cyanobacteria has been used in agriculture to improve the soil quality for their beneficial effects on plant health and productivity, since they produce diverse metabolites such as polysaccharides, betaines, micronutrients, and plant growth hormones (cytokinins, auxins, abscisic, and gibberellic acid). In specific, *Anabaena* extracts improve plant resistance to both biotic and abiotic stresses, including antimicrobial activity against different pathogens (Righini and Roberti, 2019). The importance of *Anabaena* in the community is enhanced during the rainy season, it has been reported that when the water content of alkali soils increases the biomass was dominated by *Anabaena* among others Cyanobacteria (Rao and Burns, 1991).

In contrast, some hubs in the bulk soil were members of Actinobacteria phylum (*Plantactinospora, Streptomyces, Janibacter*, and *Isoptericola*). Actinobacteria have a ubiquitous distribution in the biosphere being a dominant taxon in soil microbial communities (Bull, 2011). They have particular tolerance to high salinity and desiccation and have been isolated from many arid and hyper-arid deserts, including habitats considered as potential analogs of Mars (Stevenson and Hallsworth, 2014). Another characteristic of Actinobacteria is the large arsenal of secondary metabolites, nowadays about two-thirds of all naturally derived antibiotics in current clinical use, as well as antifungal, antihelmintic, and many anticancer compounds are produced by them. However, the few genomes of Actinobacteria strains isolated or

recovered from environmental metagenomic data in arid environments suggest that there is a plenteous actinobacterial bioactive chemicals to be discovered (Thumar et al., 2010; Mohammadipanah and Wink, 2016).

Other interesting hubs were *Chroococcidiopsis* and *Craurococcus*, both aerobic photosynthetic bacteria which could play an important role as primary producers in the bulk soil. *Chroococcidiopsis* are Cyanobacteria extremely resistant to desiccation and ionizing radiation (Verseux et al., 2017), during nutrient deprivation cell divisions occurred and were able to survive after one moth of starvation (Billi and Grilli Caiola, 1996). *Craurococcus* are strictly aerobic and chemoorganotrophic, which produced Chla and carotenoids only under aerobic growth conditions, these genera was isolated from mesic soils has only one species described (Saitoh et al., 1998).

Finally, the hubs with the maximum out degree were *Pontibacillus* and *RB41 unclassified*. *Pontibacillus* has been described as a salt-tolerant microbe, with the capacity to fix nitrogen; solubilize zinc, potassium, and phosphorus; produce ammonia, HCN, siderophores and other secondary metabolites (Yadav and Saxena, 2018). *Actinobacteria RB41* from soils under low-nutrient or stress conditions was shown to be important in maintaining biogeochemical and metabolic functions (Foesel et al., 2013), since they are positive correlated with nitrogen and sulfur cycling, such as nitrification, sulfide oxidation, sulfite reduction, and dimethylsulfoniopropionate degradation (Wang et al., 2019). Interestingly, all these keystone taxa not necessarily belong to the most abundant groups. These suggest that relatively rare groups in the community have an important role by their functional traits and keeping the connections on the community network.

Although community structure arises from a complex interplay between deterministic and stochastic forces, our results suggest that *A. lechuguilla* recruits specific rhizospheric microbes with functional traits that benefits the plant through growth promotion, nutrition, and control disease. This selection follows principally a deterministic processes that shapes the rizhospheric microbial communities, directed by the plant modifications around the roots (Nicolitch et al., 2017; Fitzpatrick et al., 2018) but also subjected to the influence of other environmental variables, such as seasonality and soil properties. Further analyses are needed to better understand the mechanisms by which the plants inhabiting arid lands select their rhizospheric community, for example there is few information about the role of exudates secreted by CAM (crassulaceae acid metabolism) plants in shaping microbial communities. Furthermore, knowledge of factors involved in plant–microorganism interactions in oligotrophic and saline soils should be helpful for conservation efforts and reforestation projects that pretend to use native plants in this kind of areas.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA553176.

AUTHOR CONTRIBUTIONS

This study was designed and coordinated by NL-L and VS. AE and NL-L performed the data collection, general analysis and interpretation. EO and MH performed the network analysis. NL-L and AE drafting the article. MH and VS performed a critical revision of the article. All authors contributed to the article and approved the submitted version.

FUNDING

The authors acknowledge the funding of WWF-Alianza Carlos Slim, and the support by the SEP-CONACYT Basic Science 254406 to NL-L. AE thanks to CONACyT for the scholarship 588371.

ACKNOWLEDGMENTS

We would like to thank Braulio Gutierrez and Christhopher A. Díaz for their valuable help during fieldwork and sampling collections. Joel Flores Rivas for their valuable comments and discussion. Juan Pablo Rodas, Alejandra Colunga, Carmen Rocha, and Elizabeth Cortés for the technical support during the development of the project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.01028/full#supplementary-material

REFERENCES

Adamic, L. A., and Huberman, B. A. (2000). Power-law distribution of the world wide web. *Science* 287 (5461), 2115–2115. doi: 10.1126/science.287.5461.2115a

Aguilera, L. E., Gutiérrez, J. R., and Meserve, P. L. (1999). Variation in soil microorganisms and nutrients underneath and outside the canopy of Adesmia bedwellii (Papilionaceae) shrubs in arid coastal Chile following drought and above average rainfall. *J. Arid. Environ.* 42 (1), 61–70. doi: 10.1006/jare.1999.0503

Albuquerque, L., and da Costa, M. S. (2014). *The family gaiellaceae. The Prokaryotes: Actinobacteria*. Eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin, Heidelberg: Springer), 357–360.

Alcaraz, L. D., Olmedo, G., Bonilla, G., Cerritos, R., Hernández, G., Cruz, A., et al. (2008). The genome of Bacillus coahuilensis reveals adaptations essential for survival in the relic of an ancient marine environment. *Proc. Natl. Acad. Sci.* 105 (15), 5803–5808. doi: 10.1073/pnas.0800981105

An, S., Couteau, C., Luo, F., Neveu, J., and DuBow, M. S. (2013). Bacterial diversity of surface sand samples from the Gobi and Taklamaken deserts. *Microbial Ecol.* 66 (4), 850–860. doi: 10.1007/s00248-013-0276-2

Andrew, D. R., Fitak, R. R., Munguia-Vega, A., Racolta, A., Martinson, V. G., and Dontsova, K. (2012). Abiotic factors shape microbial diversity in Sonoran Desert soils. *Appl. Environ. Microbiol.* 78 (21), 7527–7537. doi: 10.1128/AEM.01459-12

Bachar, A., Al-Ashhab, A., Soares, M. I. M., Sklarz, M. Y., Angel, R., Ungar, E. D., et al. (2010). Soil microbial abundance and diversity along a low precipitation gradient. *Microbial Ecol.* 60 (2), 453–461. doi: 10.1007/s00248-010-9727-1

Baggs, E. M., Smales, C. L., and Bateman, E. J. (2010). Changing pH shifts the microbial sourceas well as the magnitude of N 2 O emission from soil. *Biol. Fertility Soils* 46 (8), 793–805. doi: 10.1007/s00374-010-0484-6

Barrett, M., Khalil, M. I., Jahangir, M. M. R., Lee, C., Cardenas, L. M., Collins, G., et al. (2016). Carbon amendment and soil depth affect the distribution and abundance of denitrifiers in agricultural soils. *Environ. Sci. Pollut. Res.* 23 (8), 7899–7910. doi: 10.1007/s11356-015-6030-1

Bashan, Y., Puente, M. E., Rodriguez-Mendoza, M. N., Toledo, G., Holguin, G., Ferrera-Cerrato, R., et al. (1995). Survival of Azospirillum brasilense in the Bulk Soil and Rhizosphere of 23 Soil Types. *Appl. Environ. Microbiol.* 61 (5), 1938–1945. doi: 10.1128/AEM.61.5.1938-1945.1995

Bergmann, G. T., Bates, S. T., Eilers, K. G., Lauber, C. L., Caporaso, J. G., Walters, W. A., et al. (2011). The under-recognized dominance of Verrucomicrobia in soil bacterial communities. *Soil Biol. Biochem.* 43 (7), 1450–1455. doi: 10.1016/j.soilbio.2011.03.012

Billi, D., and Grilli Caiola, M. (1996). Effects of nitrogen and phosphorus deprivation on Chroococcidiopsis sp.(Chroococcales). *Algolog. Stud. Für Hydrobiol.* 83, 93–105. doi: 10.1127/algol_stud/83/1996/93

Bull, A. T. (2011). "Actinobacteria of the extremobiosphere," in *Extremophiles Handbook*. Ed. K. Horikoshi (Tokyo, Japan: Springer), 1203–1240.

Cararetо, A. L. M., de Souza, J. A. M., de Mello Varani, A., and de Macedo Lemos, E. G. (2014). "The family rhizobiaceae," in *The Prokaryotes* (Berlin, Heidelberg: Springer), 419–437.

Castillo Quiroz, D., Sáenz Reyes, J. T., Narcia Velasco, M., and Vázquez Ramos, J. A. (2013). Physical and mechanical properties of Agave lechuguilla Torr. fiber under plantations of five provenances. *Rev. Mex. Cienc. Forestales* 4 (19), 78–91.

Cerritos, R., Vinuesa, P., Eguiarte, L. E., Herrera-Estrella, L., Alcaraz-Peraza, L. D., Arvizu-Gomez, J. L., et al. (2008). Bacillus coahuilensis sp. nov., a moderately halophilic species from a desiccation lagoon in the Cuatro Cienegas Valley in Coahuila, Mexico. *Int. J. Systemat. Evol. Microbiol.* 58 (4), 919–923. doi: 10.1099/ijs.0.64959-0

Chase, J. M. (2007). Drought mediates the importance of stochastic community assembly. *Proc. Natl. Acad. Sci.* 104 (44), 17430–17434. doi: 10.1073/pnas.0704350104

Chave, J. (2004). Neutral theory and community ecology. *Ecol. Lett.* 7 (3), 241–253. doi: 10.1111/j.1461-0248.2003.00566.x

Chen, M. Y., Wu, S. H., Lin, G. H., Lu, C. P., Lin, Y. T., Chang, W. C., et al. (2004). Rubrobacter taiwanensis sp. nov., a novel thermophilic, radiation-resistant species isolated from hot springs. *Int. J. Systemat. Evol. Microbiol.* 54 (5), 1849–1855. doi: 10.1099/ijs.0.63109-0

Cockell, C. S., Olsson, K., Knowles, F., Kelly, L., Herrera, A., Thorsteinsson, T., et al. (2009). Bacteria in weathered basaltic glass, Iceland. *Geomicrobiol. J.* 26 (7), 491–507. doi: 10.1080/01490450903061101

Cole, J. K., Gieler, B. A., Heisler, D. L., Palisoc, M. M., Williams, A. J., Dohnalkova, A. C., et al. (2013). Kallotenue papyrolyticum gen. nov., sp. nov., a cellulolytic and filamentous thermophile that represents a novel lineage (Kallotenuales ord. nov., Kallotenuaceae fam. nov.) within the class Chloroflexia. *Int. J. Systemat. Evol. Microbiol.* 63 (12), 4675–4682. doi: 10.1099/ijs.0.053348-0

Coleman-Derr, D., Desgarennes, D., Fonseca-Garcia, C., Gross, S., Clingenpeel, S., Woyke, T., et al. (2016). Plant compartment and biogeography affect microbiome composition in cultivated and native Agave species. *New Phytol.* 209 (2), 798–811. doi: 10.1111/nph.13697

Compant, S., Samad, A., Faist, H., and Sessitsch, A. (2019). A review on the plant microbiome: Ecology, functions and emerging trends in microbial application. *J. Adv. Res.* 19, 29–37. doi: 10.1016/j.jare.2019.03.004

Da Rocha, U. N., Andreote, F. D., de Azevedo, J. L., van Elsas, J. D., and van Overbeek, L. S. (2010). Cultivation of hitherto-uncultured bacteria belonging to the Verrucomicrobia subdivision 1 from the potato (Solanum tuberosum L.) rhizosphere. *J. Soils Sediments* 10 (2), 326–339. doi: 10.1007/s11368-009-0160-3

De Anda, V., Zapata-Peñasco, I., Blaz, J., Poot-Hernandez, A. C., Contreras-Moreira, B., Hernandez Rosales, M., et al. (2018). Understanding the mechanisms behind the response to environmental perturbation in microbial mats: a metagenomic-network based approach. *Front. Microbiol.* 9, 2606. doi: 10.3389/fmicb.2018.02606

De Souza, J. A. M., Carareto Alves, L. M., de Mello Varani, A., and de Macedo Lemos, E. G. (2014). "The family bradyrhizobiaceae," in *The Prokaryotes* (Berlin, Heidelberg: Springer), 135–177.

DeBruyn, J. M., Nixon, L. T., Fawaz, M. N., Johnson, A. M., and Radosevich, M. (2011). Global biogeography and quantitative seasonal dynamics of Gemmatimonadetes in soil. *Appl. Environ. Microbiol.* 77 (17), .6295–.6300. doi: 10.1128/AEM.05005-11

Desgarennes, D., Garrido, E., Torres-Gomez, M. J., Pena-Cabriales, J. J., and Partida-Martinez, L. P. (2014). Diazotrophic potential among bacterial communities associated with wild and cultivated Agave species. *FEMS Microbiol. Ecol.* 90 (3), 844–857. doi: 10.1111/1574-6941.12438

Desnues, C., Rodriguez-Brito, B., Rayhawk, S., Kelley, S., Tran, T., Haynes, M., et al. (2008). Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* 452 (7185), 340–343. doi: 10.1038/nature06735

Doane, T. A., and Horwáth, W. R. (2003). Spectrophotometric determination of nitrate with a single reagent. *Analyt. Lett.* 36 (12), 2713–2722. doi: 10.1081/AL-120024647

Doornbos, R. F., van Loon, L. C., and Bakker, P. A. (2012). Impact of root exudates and plant defense signaling on bacterial communities in the rhizosphere. A review. *Agron. Sustain. Dev.* 32 (1), 227–243. doi: 10.1007/s13593-011-0028-y

Dornelas, M., Connolly, S. R., and Hughes, T. P. (2006). Coral reef diversity refutes the neutral theory of biodiversity. *Nature* 440 (7080), 80. doi: 10.1038/nature04534

Escalante, A. E., Caballero-Mellado, J., Martínez-Aguilar, L., Rodríguez-Verdugo, A., Gonzalez-Gonzalez, A., Toribio-Jimenez, J., et al. (2009). Pseudomonas cuatrocienegasensis sp. nov., isolated from an evaporating lagoon in the Cuatro Cienegas Basin in Coahuila, Mexico. *Int. J. Systemat. Evol. Microbiol.* 59 (6), 1416–1420. doi: 10.1099/ijs.0.006189-0

Faust, K., and Raes, J. (2012). Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10 (8), 538. doi: 10.1038/nrmicro2832

Fierer, N., Bradford, M. A., and Jackson, R. B. (2007). Toward an ecological classification of soil bacteria. *Ecology* 88 (6), 1354–1364. doi: 10.1890/05-1839

Fitzpatrick, C. R., Copeland, J., Wang, P. W., Guttman, D. S., Kotanen, P. M., and Johnson, M. T. (2018). Assembly and ecological function of the root microbiome across angiosperm plant species. *Proc. Natl. Acad. Sci.* 115 (6), E1157–E1165. doi: 10.1073/pnas.1717617115

Foesel, B. U., Rohde, M., and Overmann, J. (2013). Blastocatella fastidiosagen.nov., sp nov., isolated from semiarid savanna soil – the first described species of Acidobacteria subdivision 4. *Systemat. Appl. Microbiol.* 36, 82–89. doi: 10.1016/j.syapm.2012.11.002

Forster, J. C. (1995). "Soil sampling, handling, storage and analysis," *in Methods in applied soil microbiology and biochemistry*. Eds. K. Alef and P. Nannipieri. (United State of America: Academic Press), 49–121.

Gentry, H. S. (1982). *Agaves of continental North America*. (Tucson, Arizona, USA: The University of Arizona Press), 670 p.

Goberna, M., Navarro-Cano, J. A., Valiente-Banuet, A., García, C., and Verdú, M. (2014). Abiotic stress tolerance and competition-related traits underlie phylogenetic clustering in soil bacterial communities. *Ecol. Lett.* 17 (10), 1191–1201. doi: 10.1111/ele.12341

Goodfellow, M., Kämpfer, P., Busse, H. J., Trujillo, M. E., Suzuki, K. I., Ludwig, W., et al. (2012). *Bergey's Manual® of Systematic Bacteriology: Volume Five The Actinobacteria, Part A*. (New York: Springer).

Gu, J., Nicoullaud, B., Rochette, P., Grossel, A., Hénault, C., Cellier, P., et al. (2013). A regional experiment suggests that soil texture is a major control of N2O emissions from tile-drained winter wheat fields during the fertilization period. *Soil Biol. Biochem.* 60, 134–141. doi: 10.1016/j.soilbio.2013.01.029

Jordán, F., Okey, T. A., Bauer, B., and Libralato, S. (2008). Identifying important species: linking structure and function in ecological networks. *Ecol. Modell.* 216 (1), 75–80. doi: 10.1016/j.ecolmodel.2008.04.009

Kavamura, V. N., Taketani, R. G., Lançoni, M. D., Andreote, F. D., Mendes, R., and de Melo, I. S. (2013). Water regime influences bulk soil and rhizosphere of Cereus jamacaru bacterial communities in the Brazilian Caatinga biome. *PloS One* 8 (9), e73606. doi: 10.1371/journal.pone.0073606

Kawaichi, S., Ito, N., Kamikawa, R., Sugawara, T., Yoshida, T., and Sako, Y. (2013). Ardenticatena maritima gen. nov., sp. nov., a ferric iron-and nitrate-reducing bacterium of the phylum 'Chloroflexi';isolated from an iron-rich coastal

hydrothermal field, and description of Ardenticatenia classis nov. *Int. J. Systemat. Evol. Microbiol.* 63 (8), 2992–3002. doi: 10.1099/ijs.0.046532-0

Kazda, J., Pavlik, I., Falkinham, J. O.III, and Hruska, K. (2010). *The ecology of mycobacteria: impact on animal's and human's health.* (Berlin, Heidelberg: Springer).

Kelly, D. P., McDonald, I. R., and Wood, A. P. (2014). "The family methylobacteriaceae," in *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria*. (Berlin, Heidelberg: Springer), 313–340.

Kielak, A. M., Barreto, C. C., Kowalchuk, G. A., van Veen, J. A., and Kuramae, E. E. (2016). The ecology of Acidobacteria: moving beyond genes and genomes. *Front. Microbiol.* 7, 744. doi: 10.3389/fmicb.2016.00744

Kindaichi, T., Yamaoka, S., Uehara, R., Ozaki, N., Ohashi, A., Albertsen, M., et al. (2016). Phylogenetic diversity and ecophysiology of candidate phylum Saccharibacteria in activated sludge. *FEMS Microbiol. Ecol.* 92, fiw078. doi: 10.1093/femsec/fiw078

Kulichevskaya, I. S., Suzina, N. E., Liesack, W., and Dedysh, S. N. (2010). Bryobacter aggregatus gen. nov., sp. nov., a peat-inhabiting, aerobic chemo-organotroph from subdivision 3 of the Acidobacteria. *Int. J. Systemat. Evol. Microbiol.* 60 (2), 301–306. doi: 10.1099/ijs.0.013250-0

Kumar, P., Dubey, R. C., and Maheshwari, D. K. (2012). Bacillus strains isolated from rhizosphere showed plant growth promoting and antagonistic activity against phytopathogens. *Microbiolog. Res.* 167 (8), 493–499. doi: 10.1016/j.micres.2012.05.002

Kurahashi, M., Fukunaga, Y., Sakiyama, Y., Harayama, S., and Yokota, A. (2010). Euzebya tangerina gen. nov., sp. nov., a deeply branching marine actinobacterium isolated from the sea cucumber Holothuria edulis, and proposal of Euzebyaceae fam. nov., Euzebyales ord. nov. and Nitriliruptoridae subclassis nov. *Int. J. Systemat. Evol. Microbiol.* 60 (10), 2314–2319. doi: 10.1099/ijs.0.016543-0

Lançoni, M. D., Taketani, R. G., Kavamura, V. N., and de Melo, I. S. (2013). Microbial community biogeographic patterns in the rhizosphere of two Brazilian semi-arid leguminous trees. *World J. Microbiol. Biotechnol.* 29 (7), 1233–1241. doi: 10.1007/s11274-013-1286-4

Lee, K. C., Dunfield, P. F., and Stott, M. B. (2014). "The Phylum Armatimonadetes," in *The Prokaryotes*. (Berlin, Heidelberg: Springer), 447–458.

López-Lozano, N. E., Eguiarte, L. E., Bonilla-Rosso, G., García-Oliva, F., Martínez-Piedragil, C., Rooks, C., et al. (2012). Bacterial communities and the nitrogen cycle in the gypsum soils of Cuatro Ciénegas Basin, Coahuila: a Mars analogue. *Astrobiology* 12 (7), 699–709. doi: 10.1089/ast.2012.0840

Manjunath, M., Kanchan, A., Ranjan, K., Venkatachalam, S., Prasanna, R., Ramakrishnan, B., et al. (2016). Beneficial cyanobacteria and eubacteria synergistically enhance bioavailability of soil nutrients and yield of okra. *Heliyon* 2 (2), e00066. doi: 10.1016/j.heliyon.2016.e00066

Marasco, R., Rolli, E., Ettoumi, B., Vigani, G., Mapelli, F., Borin, S., et al. (2012). A drought resistance-promoting microbiome is selected by root system under desert farming. *PloS One* 7 (10), e48479. doi: 10.1371/journal.pone.0048479

Marasco, R., Mosqueira, M. J., Fusi, M., Ramond, J. B., Merlino, G., Booth, J. M., et al. (2018). Rhizosheath microbial community assembly of sympatric desert speargrasses is independent of the plant host. *Microbiome* 6 (1), 215. doi: 10.1186/s40168-018-0597-y

Martiny, J. B. H., Bohannan, B. J., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., et al. (2006). Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* 4 (2), 102–112. doi: 10.1038/nrmicro1341

Mendes, L. W., Kuramae, E. E., Navarrete, A. A., Van Veen, J. A., and Tsai, S. M. (2014). Taxonomical and functional microbial community selection in soybean rhizosphere. *ISME J.* 8 (8), 1577. doi: 10.1038/ismej.2014.17

Meyer, E. R. (1973). Late-Quarternary Paleoecology of the Cuatro Cienegas Basin, Coahuila, Mexico. *Ecology* 54 (5), 982–995. doi: 10.2307/1935565

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science* 298 (5594), 824. doi: 10.1126/science.298.5594.824

Miranda, K. M., Espey, M. G., and Wink, D. A. (2001). A rapid, simple spectrophotometric method for simultaneous detection of nitrate and nitrite. *Nitric. Oxide* 5 (1), 62–71. doi: 10.1006/niox.2000.0319

Mohammadipanah, F., and Wink, J. (2016). Actinobacteria from arid and desert habitats: diversity and biological activity. *Front. Microbiol.* 6, 1541. doi: 10.3389/fmicb.2015.01541

Morales-Luckie, R. A., Lopezfuentes-Ruiz, A. A., Olea-Mejía, O. F., Liliana, A. F., Sanchez-Mendieta, V., Brostow, W., et al. (2016). Synthesis of silver nanoparticles using aqueous extracts of Heterotheca inuloides as reducing agent and natural fibers as templates: Agave lechuguilla and silk. *Mater. Sci. Eng.: C.* 69, 429–436. doi: 10.1016/j.msec.2016.06.066

Mosqueira, M. J., Marasco, R., Fusi, M., Michoud, G., Merlino, G., Cherif, A., et al. (2019). Consistent bacterial selection by date palm root system across heterogeneous desert oasis agroecosystems. *Sci. Rep.* 9 (1), 1–12. doi: 10.1038/s41598-019-40551-4

Nakamura, K., Hiraishi, A., Yoshimi, Y., Kawaharasaki, M., Masuda, K., and Kamagata, Y. (1995). Microlunatus phosphovorus gen. nov., sp. nov., a new gram-positive polyphosphate-accumulating bacterium isolated from activated sludge. *Int. J. Systemat. Evol. Microbiol.* 45 (1), 17–22. doi: 10.1099/00207713-45-1-17

Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L., et al. (2013). Patterns and processes of microbial community assembly. *Microbiol. Mol. Biol. Rev.* 77 (3), 342–356. doi: 10.1128/MMBR.00051-12

Niboyet, A., Barthes, L., Hungate, B. A., Le Roux, X., Bloor, J. M. G., Ambroise, A., et al. (2009). Responses of soil nitrogen cycling to the interactive effects of elevated CO2 and inorganic N supply. *Plant Soil* 327, 35–47. doi: 10.1007/s11104-009-0029-7

Nicolitch, O., Colin, Y., Turpault, M. P., Fauchery, L., and Uroz, S. (2017). Tree roots select specific bacterial communities in the subsurface critical zone. *Soil Biol. Biochem.* 115, 109–123. doi: 10.1016/j.soilbio.2017.07.003

Pajares, S., Escalante, A. E., Noguez, A. M., García-Oliva, F., Martínez-Piedragil, C., Cram, S. S., et al. (2016). Spatial heterogeneity of physicochemical properties explains differences in microbial composition in arid soils from Cuatro Cienegas, Mexico. *PeerJ* 4, e2459. doi: 10.7717/peerj.2459

Philippot, L., Raaijmakers, J. M., Lemanceau, P., and Van Der Putten, W. H. (2013). Going back to the roots: the microbial ecology of the rhizosphere. *Nat. Rev. Microbiol.* 11 (11), 789–799. doi: 10.1038/nrmicro3109

Postma, A., Slabbert, E., Postma, F., and Jacobs, K. (2016). Soil bacterial communities associated with natural and commercial Cyclopia spp. *FEMS Microbiol. Ecol.* 92 (3), fiw016. doi: 10.1093/femsec/fiw016

Puente, M. E., Li, C. Y., and Bashan, Y. (2009). Rock-degrading endophytic bacteria in cacti. *Environ. Exp. Bot.* 66 (3), 389–401. doi: 10.1016/j.envexpbot.2009.04.010

Rao, D. L. N., and Burns, R. G. (1991). The influence of blue-green algae on the biological amelioration of alkali soils. *Biol. Fertility Soils* 11 (4), 306–312. doi: 10.1007/BF00335853

Reed, S. C., Cleveland, C. C., and Townsend, A. R. (2011). Functional ecology of free-living nitrogen fixation: a contemporary perspective. *Annu. Rev. Ecol. Evol. Systemat.* 42, 489–512. doi: 10.1146/annurev-ecolsys-102710-145034

Richardson, A. E., Barea, J. M., McNeill, A. M., and Prigent-Combaret, C. (2009). Acquisition of phosphorus and nitrogen in the rhizosphere and plant growth promotion by microorganisms. *Plant Soil* 321 (1-2), 305–339. doi: 10.1007/s11104-009-9895-2

Righini, H., and Roberti, R. (2019). "Algae and Cyanobacteria as Biocontrol Agents of Fungal Plant Pathogens," in *Plant Microbe Interface*. (Cham: Springer), 219–238.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4, e2584. doi: 10.7717/peerj.2584

Saharan, B. S., and Nehra, V. (2011). Plant growth promoting rhizobacteria: a critical review. *Life Sci. Med. Res.* 21 (1), 30.

Saitoh, S., Suzuki, T., and Nishimura, Y. (1998). Proposal of Craurococcus roseus gen. nov., sp. nov. and Paracraurococcus ruber gen. nov., sp. nov., novel aerobic bacteriochlorophyll a-containing bacteria from soil. *Int. J. Systemat. Evol. Microbiol.* 48 (3), 1043–1047. doi: 10.1099/00207713-48-3-1043

Sasse, J., Martinoia, E., and Northen, T. (2018). Feed your friends: do plant exudates shape the root microbiome? *Trends Plant Sci.* 23 (1), 25–41. doi: 10.1016/j.tplants.2017.09.003

Saul-Tcherkas, V., and Steinberger, Y. (2011). Soil microbial diversity in the vicinity of a Negev Desert shrub—Reaumuria negevensis. *Microbial Ecol.* 61 (1), 64–81. doi: 10.1007/s00248-010-9763-x

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75 (23), 7537–7541. doi: 10.1128/AEM.01541-09

Schulte, E. E., and Hopkins, B. G. (1996). "Estimation of soil organic matter by weight loss-on-ignition," in *Soil organic matter: Analysis and interpretation*. Eds. F. R .Magdoff, M. A. Tabatabai, E. A. Hanlon Jr. (Soil Science Society of America, John Wiley & Sons, Inc.), 46, 21–31. doi: 10.2136/sssaspecpub46.c3

Senechkin, I. V., Speksnijder, A. G., Semenov, A. M., Van Bruggen, A. H., and Van Overbeek, L. S. (2010). Isolation and partial characterization of bacterial strains on low organic carbon medium from soils fertilized with different organic amendments. *Microbial Ecol.* 60 (4), 829–839. doi: 10.1007/s00248-010-9670-1

Shahzad, R., Khan, A. L., Bilal, S., Asaf, S., and Lee, I. J. (2018). What is there in seeds? Vertically transmitted endophytic resources for sustainable improvement in plant growth. *Front. Plant Sci.* 9, 24. doi: 10.3389/fpls.2018.00024

Shaw, G. T.-W., Pao, Y.-Y., and Wang, D. (2016). MetaMIS: a metagenomic microbial interaction simulator based on microbial community profiles. *BMC Bioinf.* 17, 488. doi: 10.1186/s12859-016-1359-0

Shen, G., Zhang, S., Liu, X., Jiang, Q., and Ding, W. (2018). Soil acidification amendments change the rhizosphere bacterial community of tobacco in a bacterial wilt affected field. *Appl. Microbiol. Biotechnol.* 102 (22), 9781–9791. doi: 10.1007/s00253-018-9347-0

Shi, S., Nuccio, E. E., Shi, Z. J., He, Z., Zhou, J., and Firestone, M. K. (2016). The interconnected rhizosphere: high network complexity dominates rhizosphere assemblages. *Ecol. Lett.* 19 (8), 926–936. doi: 10.1111/ele.12630

Shinwari, Z. K., Tanveer, F., and Iqrar, I. (2019). "Role of Microbes in Plant Health, Disease Management, and Abiotic Stress Management," in *Microbiome in Plant Health and Disease*. (Singapore: Springer), 231–250.

Silva-Montellano, A., and Eguiarte, L. E. (2003). Geographic patterns in the reproductive ecology of Agave lechuguilla (Agavaceae) in the Chihuahuan desert. I. Floral characteristics, visitors, and fecundity. *Am. J. Bot.* 90 (3), 377–387. doi: 10.3732/ajb.90.3.377

Sole, R. V., and Montoya, M. (2001). Complexity and fragility in ecological networks. *Proc. R. Soc. London Ser. B.: Biol. Sci.* 268 (1480), 2039–2045. doi: 10.1098/rspb.2001.1767

Souza, V., Espinosa-Asuar, L., Escalante, A. E., Eguiarte, L. E., Farmer, J., Forney, L., et al. (2006). An endangered oasis of aquatic microbial biodiversity in the Chihuahuan desert. *Proc. Natl. Acad. Sci.* 103 (17), 6565–6570. doi: 10.1073/pnas.0601434103

Spaepen, S., Vanderleyden, J., and Remans, R. (2007). Indole-3-acetic acid in microbial and microorganism-plant signaling. *FEMS Microbiol. Rev.* 31 (4), 425–448. doi: 10.1111/j.1574-6976.2007.00072.x

Stevenson, A., and Hallsworth, J. E. (2014). Water and temperature relations of soil Actinobacteria. *Environ. Microbiol. Rep.* 6, 744–755. doi: 10.1111/1758-2229.12199

Tapia-Torres, Y., López-Lozano, N. E., Souza, V., and García-Oliva, F. (2015). Vegetation-soil system controls soil mechanisms for nitrogen transformations in an oligotrophic Mexican desert. *J. Arid. Environ.* 114, 62–69. doi: 10.1016/j.jaridenv.2014.11.007

Thumar, J. T., Dhulia, K., and Singh, S. P. (2010). Isolation and partial purification of an antimicrobial agent from halotolerant alkaliphilic Streptomyces aburaviensis strain Kut-8. *World J. Microbiol. Biotechnol.* 26, 2081–2087. doi: 10.1007/s11274-010-0394-7

Torres-Cortés, G., Millán, V., Fernández-González, A. J., Aguirre-Garrido, J. F., Ramírez-Saad, H. C., Fernández-López, M., et al. (2012). Bacterial community in the rhizosphere of the cactus species Mammillaria carnea during dry and rainy seasons assessed by deep sequencing. *Plant Soil* 357 (1-2), 275–288. doi: 10.1007/s11104-012-1152-4

Vellend, M. (2010). Conceptual synthesis in community ecology. *Q. Rev. Biol.* 85 (2), 183–206. doi: 10.1086/652373

Verseux, C., Baqué, M., Cifariello, R., Fagliarone, C., Raguse, M., Moeller, R., et al. (2017). Evaluation of the resistance of Chroococcidiopsis spp. to sparsely and densely ionizing irradiation. *Astrobiology* 17 (2), 118–125. doi: 10.1089/ast.2015.1450

Walker, J. J., and Pace, N. R. (2007). Phylogenetic composition of Rocky Mountain endolithic microbial ecosystems. *Appl. Environ. Microbiol.* 73 (11), 3497–3504. doi: 10.1128/AEM.02656-06

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microb.* 73 (16), 5261–5267. doi: 10.1128/AEM.00062-07

Wang, T., Hao, Y., Zhu, M., Yu, S., Ran, W., Xue, C., et al. (2019). Characterizing differences in microbial community composition and function between Fusarium wilt diseased and healthy soils under watermelon cultivation. *Plant Soil* 438 (1-2), 421–433. doi: 10.1007/s11104-019-04037-6

Yadav, A. N., and Saxena, A. K. (2018). Biodiversity and biotechnological applications of halophilic microbes for sustainable agriculture. *J. Appl. Biol. Biotechnol. Vol.* 6 (1), 48–55. doi: 10.7324/JABB.2018.60109

Yan, Y., Kuramae, E. E., de Hollander, M., Klinkhamer, P. G., and van Veen, J. A. (2017). Functional traits dominate the diversity-related selection of bacterial communities in the rhizosphere. *ISME J.* 11 (1), 56. doi: 10.1038/ismej.2016.108

Zarilla, K. A., and Perry, J. J. (1984). Thermoleophilum album gen. nov. and sp. nov., a bacterium obligate for thermophily and n-alkane substrates. *Arch. Microbiol.* 137 (4), 286–290. doi: 10.1007/BF00410723

Zhalnina, K., Louie, K. B., Hao, Z., Mansoori, N., da Rocha, U. N., Shi, S., et al. (2018). Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat. Microbiol.* 3 (4), 470. doi: 10.1038/s41564-018-0129-3

Zhang, C., Lin, Y., Tian, X., Xu, Q., Chen, Z., and Lin, W. (2016). Tobacco bacterial wilt suppression with biochar soil addition associates to improved soil physiochemical properties and increased rhizosphere bacteria abundance. *Appl. Soil Ecol.* 112, 90–96. doi: 10.1016/j.apsoil.2016.12.005

Zhou, J., Gu, Y., Zou, C., and Mo, M. (2007). Phylogenetic diversity of bacteria in an earth-cave in Guizhou Province, Southwest of China. *J. Microbiol.* 45 (2), 105–112.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 11 August 2020
doi: 10.3389/fpls.2020.01224

Integral Projection Models and Sustainable Forest Management of *Agave inaequidens* in Western Mexico

Ignacio Torres-García[1,2], *Alejandro León-Jacinto*[†], *Ernesto Vega*[2], *Ana Isabel Moreno-Calles*[1] *and Alejandro Casas*[2*]

[1] Environmental Transdisciplinary Studies, Escuela Nacional de Estudios Superiores, Universidad Nacional Autónoma de México, Morelia, México, [2] Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, Morelia, México

OPEN ACCESS

Edited by:
Jim Leebens-Mack,
University of Georgia, United States

Reviewed by:
Vania Jiménez-Lobato,
Autonomous University of Guerrero, Mexico
Alejandra Moreno-Letelier,
National Autonomous University of Mexico, Mexico

***Correspondence:**
Alejandro Casas
acasas@cieco.unam.mx

[†]Deceased

Specialty section:
This article was submitted to Functional Plant Ecology, a section of the journal Frontiers in Plant Science

Received: 02 March 2020
Accepted: 27 July 2020
Published: 11 August 2020

Citation:
Torres-García I, León-Jacinto A, Vega E, Moreno-Calles AI and Casas A (2020) Integral Projection Models and Sustainable Forest Management of *Agave inaequidens* in Western Mexico. Front. Plant Sci. 11:1224. doi: 10.3389/fpls.2020.01224

In México, at least 37 *Agave* species are extracted from wild populations for producing distilled spirits. This activity involves harvesting mature agaves just before producing their inflorescences, which cancels sexual reproduction of plants used. The increasing demand of agaves spirits in national and international markets is determining a strong pressure on wild populations, most of them lacking adequate management. In addition, the dynamics of agave populations may be affected by natural phenomena like oscillation of rainfall regimes, which affects the recruitment of agave seedlings, or the scarcity of pollinators that may affect seed production and general population dynamics. We studied the demography of wild populations of *Agave inaequidens* to analyze critical conditions for populations recovery, modelling the effects of rainfall trends on the demographic performance of this species, and exploring response of populations to hypothetical extraction regimes and reforestation efforts. Our study was performed in four well-conserved wild populations in Central Western Mexico, each population was sampled in a plot of about one hectare composed by 10 subplots 50 x 5 m (2500 m^2). Populations were monitored yearly between 2011 and 2013, measuring plant size, reproductive individuals, and fecundity. Data were analyzed through integral projection models by using the IPMpack for R, to perform prospective analyses. We in addition constructed stochastic models to explore the possible influence of rainfall variation on species demography, using data for the drier and wetter years of the study period. Population growth varied from λ=1.003 to λ=0.899 among populations and years, and exceptionally λ=0.559 after a fire event. Low rainfall decreases λ values, indicating especial limitations to harvesting agaves during dry years whose frequency most probably will increase. In general, extraction rates from 10% to 30% of mature individuals are viable to maintain λ above 1, and these rates may be higher if new plants are introduced in populations. Depending on levels of extraction, our models suggest that it is necessary to carry out actions of reforestation, and *in situ* management according to the trends found in each site. This is one indispensable condition to maintain λ close to or greater than 1.

Sustainable extraction of wild agaves is possible, some communities are already carrying out a repertoire of goods practices in this direction, but together with ecological criteria and good management techniques, strict regulations and social organization are needed to achieve it.

Keywords: Agavoideae, mescal, demography, forest management, Michoacán, sustainability

INTRODUCTION

Mexico is the territory that hosts the higher species diversity of the genus *Agave* in the world (García-Mendoza et al., 2019). It is the main setting of natural evolutionary divergence of this genus and the principal area of management and domestication of agaves (Colunga-GarcíaMarín et al., 2017). The interactions between such diversity and the extraordinarily rich human cultures inhabiting this territory started thousands of years ago. Archaeobotanical remains, dating back nearly 10,000 years B.P., witness these long interactions (McNeish, 1967; Gentry, 1982), including gathering, incipient management of populations and, eventually, domestication of some species. The most ancient uses recorded like food, rope making and textile manufacturing, are all current. A total of 22 use categories of agaves (Torres-García et al., 2019) are practiced in rural Mexican societies. In the past two centuries some of these uses were developed in immense industries, such as the case of the sap-fermented beverage called pulque, henequen fibre, tequila and more recently the mescal, raicilla, and other agave spirits industries. Some of these activities have experimented a boom and then a collapse (Bowen and Valenzuela-Zapata, 2009; Colunga-GarcíaMarín et al., 2007; Ramírez-Rodríguez, 2018; Valenzuela-Zapata and Macias-Macias, 2014). In the last three decades, diverse negative socio-environmental phenomena related to the boom and intensification of production, trade, exportation, and consumption of mescal are evident. The Mexican regulatory organisms and official norms that rule the tequila and mescal appellations of origin (DOT and DOM for their acronyms in Spanish, respectively), have allowed a vertiginous incursion of investors into the traditional communities that produce mescal, which has determined substantial changes in their forms of production and marketing of this beverage. These changes have determined a strong pressure over rural mescal producers that dramatically increases the extraction of agaves from forests and harvesting agave crops intensively cultivated. Among the main negative trends it is now possible to see the depletion of wild populations and natural habitats, the transformation of both forests and traditional crop fields to establishing agave monocultures maintained with chemical inputs, which, in the case of *Agave tequilana* var. *azul* F.A.C. Weber, are called "blue deserts" (Torres-García et al., 2019). These phenomena are generated by the demand of mescal in the big cities of Mexico and the international markets. Another negative consequence of the intensive extractive regimes is the alteration of interactions in biotic communities, for instance, the removal or damage to nurse plants, and impact on flower visitors. Thus, the foraging patterns of the main pollinators of agaves, bats, and diverse communities of birds and insects that visit agave flowers and have a role in pollination, are all being affected by over-extraction of mature agaves, which compromises the pollination effectivity and the general biological diversity in numerous ecosystems where the agaves occur (Rocha et al., 2006; Trejo-Salazar et al., 2016). It has been documented by Estrella-Ruiz (2008) and Baraza-Ruiz and Estrella-Ruiz (2008) that in areas where the offer of nectar decreases, bats scarcely visit them, and move to populations with a higher nectar offer. This fact determines that populations offering few flowers are scarcely visited and seed production is low or unlikely.

In México, 37 *Agave* species have been documented to be extracted exclusively from wild populations for producing distilled spirits, mainly adult plants just before sexual reproduction, most of them without targeted management (Torres et al., 2015a). The traditional process of elaborating mescal involves four stages, the first one is the harvest, a critical step determining the amount and quality of mescal and the recovering of populations used for production. The second stage is baking agave stems, in which carbohydrates of the plant tissue are lysed into smaller molecules and the smoked flavour is acquired. The third stage is fermentation, in which sugar is transformed into alcohol. And finally, distillation, through which the alcohol of mescal is separated. Depending on the species used, the weight and concentration of carbohydrates may vary, as well as the efficiency of the distillation methods and the preferences of tastes that are related to alcohol concentration. And all these aspects influence the variable amount of mescal that can be obtained per individual plant of agave. In our fieldwork, and based on information from mescal producers, we have recorded that variation may range from 0.5 litres per agave in *A potatorum* Zucc., *A. marmorata* Roezl., and *A. karwinskii* Zucc., 2.0 litres in *A. maximiliana* Baker, 2.0 to 15.0 litres in *A. rhodacantha* Trel., 4.5 litres in *A. cupreata* Trel. & A. Berger, and 10 litres in *A. angustifolia* Haw.

In Western Mexico, at least nine *Agave* species have been used to produce distilled beverages which, depending on the area they occur and are used, receive different local names like "raicilla", "tepe", "tutsi", "barranca", "tusca", "vino de cerro", "vino de mezcal", among others, all of them grouped in the generic term mescal. The main species involved in these distillates are *A. tequilana* var. *azul*, *A. rhodacantha*, *A. americana* var. *subtilis* (Trel.) Valenz.-Zap. & Nabhan that are domesticated and cultivated, as well as *A. angustifolia*, *A. maximiliana*, *A. durangensis* Gentry, *A. bovicornuta* Gentry, *A. cupreata*, and *A. inaequidens* K. Koch., which grow wild in regional forests.

We studied the species *A. inaequidens*, which belongs to the Crenatae group of the Agavoideae (Gentry, 1982; APG III, 2009), a group of species well known for their rare or null production of clonal shoots, being the sexual monocarpic event their only way of reproduction (Gentry, 1982). Throughout their life cycle,

individuals of these agaves accumulate carbohydrates, especially in their stems and foliar bases, a process that prepare the plants to sexual reproduction and end their life, using most of their reserves to generate massive floral panicles, flowers, nectar, pollen and, after pollinated, seeds. Agaves of this species reach maturity in 12 to 25 years depending on the substrate and exposure to solar radiation where plants grow (Torres-García, 2016). On average, a mature individual of *A. inaequidens* may produce nearly 400,000 fertile seeds. However, to produce mescal this unique event is cancelled by cutting the floral shoot just before it begins to develop. We have estimated that only in five municipalities, of northern Michoacán, nearly 12,000 *A. inaequidens* individual plants are harvested from wild populations every year (Torres et al., 2015b), which cancels the production of approximately 4.8 billion seeds. This activity, therefore, severely affects the regional populations and their potential to regenerate after harvesting such high number of reproductive organisms.

In this study, we focused our attention on analyzing the population dynamics of *Agave inaequidens* in order to establish ecological criteria for recommending sustainable forms of using mature plants for mescal production. Our study, therefore, looks for understanding population biology of this and other agave species, as well as supporting the construction of regulations and designing strategies and environmental education programmes for using and conserving agaves of this species and others with similar life history traits and problems. Not only human impact on reproduction may affect the dynamics of agave populations. For instance, previous studies estimated that oscillation of rainfall regimes affects the recruitment of agave seedlings (Jordan and Nobel, 1979; Nobel, 1992). These studies documented that episodic events of high precipitation, which also determined episodes of high recruitment. In addition, the scarcity of pollinators associated to both natural and human factors may affect seed production and the general population dynamics (Torres et al., 2015a). In some species, amounts of humidity and presence or not of nurse plants may determine the success or not of seedlings recruitment and the maintenance of populations regeneration (Rangel-Landa et al., 2015). Therefore, documenting the biological processes determining reproduction and the indispensable conditions for ensuring seedling establishment, survival and growth are relevant for identifying management patterns needed for conservation of agaves (Torres et al., 2013; Torres et al., 2015a).

Demographic studies of non-timber forest products (NTFP) have been useful for identifying crucial aspects of management based on ecological criteria for sustainable use. A high number of studies on ecological bases for sustainable management of NTFP have found in demography a useful tool not only for analyzing theoretical problems in relation to adaptability, vulnerability and conservation ecology, but also as a way to analyze consequences of impacts of human activities on populations and ecosystems and, importantly, a helpful approach to design strategies for sustainable use of biotic resources (Shackleton et al., 2015). Studies on monocarpic perennial plants have been carried out for several species (Metcalf et al., 2003; Metcalf et al., 2006), which emphasize that the reproductive strategy of these plants is directly related to the individuals size, larger plants yielding larger numbers of seeds and likewise increasing the number of seedlings. Case studies using matrix models of *Agave* species under human exploitation are relatively scarce [e.g., *Agave cupreata* (Illsley et al., 2007), *A. marmorata* (Jiménez-Valdés et al., 2010), *A. potatorum* (Torres et al., 2015a), and *A. angustifolia* (Arias-Medellín et al., 2016)], compared with the large number of species used. However, the few studies available for this genus allow identifying some general demographic patterns. The most remarkable aspects are that survival and growth are the processes that mostly contribute to the finite growth rate (λ), whereas fecundity has low contribution. Studies available also coincide that management actions are urgent on populations that are over-exploited, in order to procure maintaining the demographic equilibrium and contribute to the conservation of these valuable resources. Management actions such as identifying harvest thresholds of reproductive individuals based on ecological criteria, as well as protection of individuals of the lower size categories, and enhancing recruitment of other stages have been proposed for some species (Martin et al., 2011; Torres et al., 2015a). In addition, some studies emphasize the need of protecting particular nurse plants, which are crucial for ensuring establishment of agave species depending on interactions of facilitation (Torres et al., 2013; Rangel-Landa et al., 2015). However, matrix models study the populations performance commonly by using discrete categories based on the researcher methodology and may not represent satisfactorily the biological cycle of a species.

The studies referred to above have mainly used matrix models (Caswell, 2001; Ramula et al., 2009; Gaoue and Ticktin, 2010; Zuidema et al., 2010; Schmidt et al., 2011; Gaoue et al., 2011). But more recently, integral projection models (IPM´s) that do not categorize populations into discrete stages but use data on continual actual plant sizes have been used to study the demography of some NTFP species (Ellner and Rees, 2006; Merow et al., 2014). However, hitherto, these models have not been used for analyzing population dynamics of *Agave* species (Kuss et al., 2008). In this study we use these models for analyzing populations of *A. inaequidens* and the conditions for sustainable use of its populations. We studied the demography of wild populations of *A. inaequidens* to analyze critical conditions for populations recovery, the effects of rainfall trends on the demographic performance and responses of populations to hypothetical extraction regimes and reforestation efforts. In addition to population biology topics, this study aspires to generate information useful to agave managers and decision makers.

METHODS

Study Species

Agave inaequidens, locally named "maguey bruto" or "maguey alto", is a rosetophyllous monocarpic species that forms medium to large-sized rosettes. It is the most abundant agave species in

the Trans-Mexican Volcanic Belt, inhabiting areas at elevation ranges from 1850 to 2600 m (Torres-García, 2016). Its main reproductive system is sexual, but it may have axillary asexual propagules, whose reproduction may be active in response to damages of the apical meristem, but this type of events is rare. Its flowers are mainly pollinated by bats (*Leptonycteris* spp.), orioles (*Icterus* spp.) and some insects (*Apis mellifera*) (León, 2013). Studies on reproduction of other species of the Crenatae group suggest that this species most probably is self-incompatible. One single capsule produces on average 205 ± 75 (n=57) viable seeds, a panicle is composed on average by 21 ± 10 (n=16) umbels, and reproductive individuals produce on average 417,356 ± 202,176 (n=27) viable seeds (León, 2013). According to mescal producers who sow agave seeds and take care of their plantations, depending on the site where plants are established *A. inaequidens* may reach maturity in 12 to 25 years. In addition to humidity and soil quality, solar radiation received by a plant may influence its size and reproductive capacity (Nobel and Valenzuela, 1987; Nobel, 2003). The studied species grows mainly in temperate habitats such as subtropical scrub, oak forest, pine-oak forest and fir forest, the populations mainly having an aggregated spatial distribution in patches called "magueyeras", but in some areas it may show scattered spatial distribution (Gentry, 1982; Torres-García, 2016). This agave represents a multipurpose resource for numerous rural communities. Torres et al. (2015b) and Valenzuela-Zapata et al. (2011) documented that this agave is used for about 34 different specific uses, among them the extraction of complete individuals for producing the distilled spirit mescal, the most important cultural and economic activity currently carried out with this species.

Traditional production of mescal with *A. inaequidens* in the study area is carried out from October to May, the dry season. Rain limits agave harvest and mescal production because it hinders the movement of donkeys that carry the harvested agaves, wets the wood used to bake agaves, and floods the rustic roasting pits. According to León (2013), the reproductive phenology of *A. inaequidens* (**Figure 1**) starts with the emergence and development of inflorescences, starting in May and ending in November. When the inflorescence is emerging and it is no more than 2 m tall, the managers cut the inflorescence to stop its development, a practice called "capado" or castration. But the stems are not harvested until October. People harvest reproductive individuals and those that they identify that will bloom the following year, to complete a batch. One batch is composed on average by 147 agaves (n=29 producers) (Torres et al., 2015b). To complete a batch, people harvest agaves from different sites, including plantations, backyard fences, home gardens, managed and unmanaged forests, within their community or bought agave stems to other communities. The proportion of agaves from these sites is highly variable, depending on their availability and social arrangements. To produce one litre of mescal from this species, 25 to 30 kg of agave are needed. A harvested individual (stem and foliar bases), or "piña", weights on average 70 to 80 kg, meaning that about two litres of mescal can be obtained from one individual agave.

Study Sites

This study was carried out in Michoacán (**Figure 1**), in Central Western Mexico, where we selected four well-conserved wild populations, without signs of extraction. Once we obtained the permit by local authorities, we established permanent study plots of nearly 1 ha each. Each plot was divided into 10 parallel subplots of 50 x 5 m (2,500 m^2), each subplot was marked with metal labels and geo-positioned to make easier further visits. Each study plot was monitored for two consecutive years or periods. Ahead, we show the results of monitoring per period (first or 1 monitored period and second or 2 monitored period).

Permanent Plots

1. Piedra de Indio is a natural protected area of the municipality of Morelia, located at elevations around 2414 m, with a northern exposure, the slope inclination varying from 30° to 35° (UTM 14Q 268684-2170422), where vegetation is *Pinus-Quercus* forest. The plot studied in this site will be called ahead "Piedra de Indio". 2. The Ejido Pino Real in the municipality of Charo, where vegetation is also *Pinus-Quercus* forest, at elevations around 2349 m, with north-northeastern exposure, the slope inclination varying between 15° and 30° (UTM 14Q 289118-2172533). This plot will be called "Pino Real" throughout the text. 3. The Ejido Cuanajo in the municipality of Pátzcuaro, where vegetation types are *Quercus* and *Pinus-Quercus* forests, at elevations around 2541 m, with south-southeastern exposure, and slope inclinations varying between 35° and 45° (UTM 14Q 235534-2157777). This plot will be further called "Cuanajo". 4. The Ejido of Quiroga in the municipality of Quiroga, where

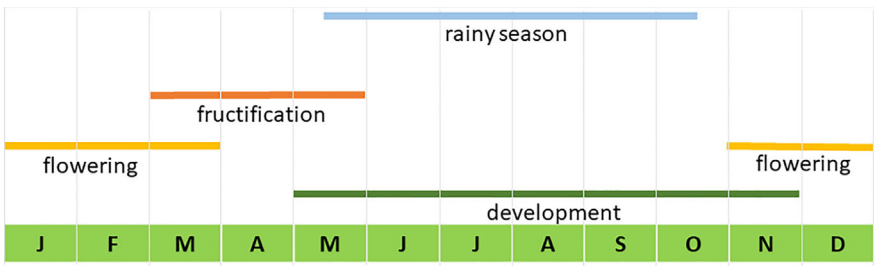


FIGURE 1 | Temporal phenology of Agave inaequidens in the study area. Modified from León, 2013.

vegetation is *Quercus* forest, at elevations around 2475 m, with south-southeastern exposure, the slope inclination varying from 15° to 25° (UTM 14Q 235534-2157777). This plot will be called "Icuacato" (**Figure 2**). The annual mean precipitation in the two first plots is 1360 mm and in the other two is 1200 mm. We obtained this information from WorldClim BIO12 (Fick and Hijmans, 2017).

Data Collection and Characterization of Size

The four populations were monitored yearly, between 2011 and 2013. We used the method of Total Foliar Area (TFA) developed by Torres et al. (2015a) for characterizing plant size in demographic studies of *Agave potatorum*. This method consists in direct measurement of the rosette size, which is used to characterize the functional state of the individuals. The TFA is calculated by measuring length and width of four leaves of four different whorl levels, then estimating the average leaf area based on calculating an ellipsoid area, the geometrical form of leaves. The average area of the ellipsoid (AAE) was then multiplied by the leaves number (LN). Therefore, the TFA is the product of AAE * LN. To adjust the data to a normal distribution, these were transformed into their natural logarithm (logTFA), similarly as in other demographic studies of monocarpic perennials (Metcalf et al., 2003). Therefore, in the three monitoring times of our study, the following data were recorded for each agave plant: Number of leaves, length, and width of four leaves of different strata of the rosette. The individuals and leaves measured were labelled and consecutively numbered.

Reproduction and Fecundity

Based on studies on reproductive phenology by León (2013), we identified the seasons of development, flowering, and fructification of the panicles (**Figure 1**), and we designed a strategy to characterize the seeds production. The first step was to estimate the number of capsules per umbel. A total of 120 umbels from all studied

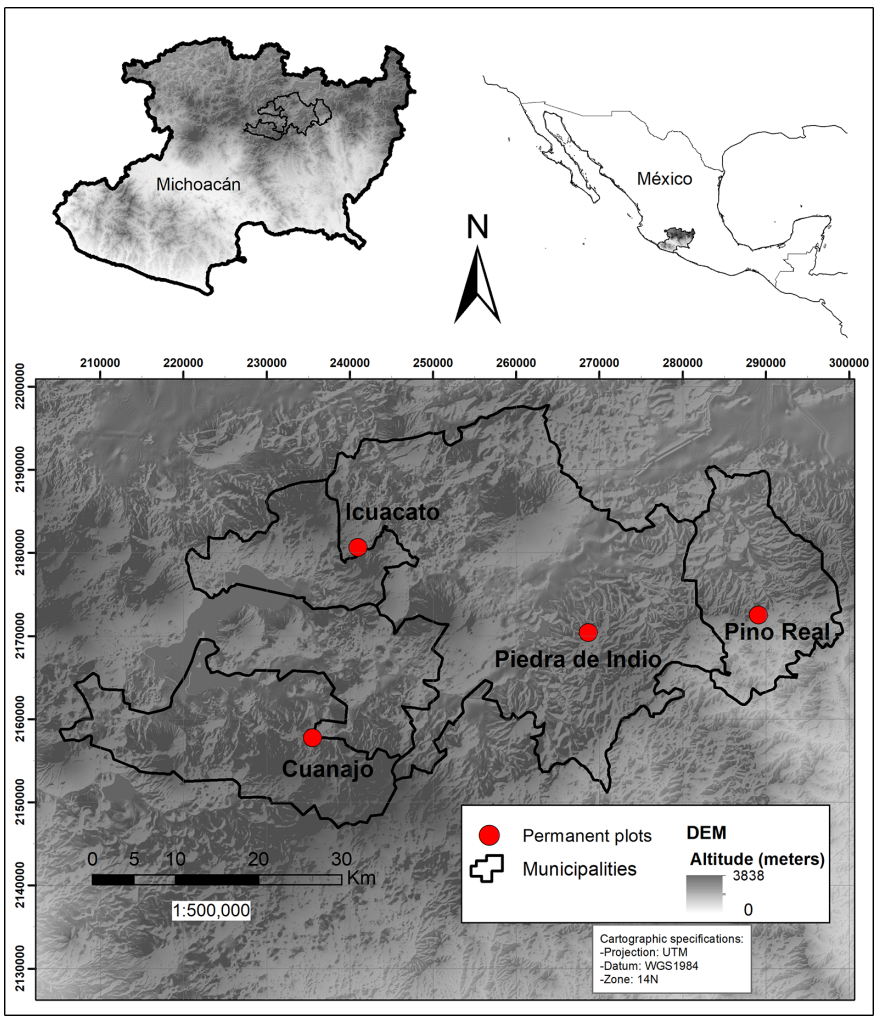


FIGURE 2 | Map of the study area and the location of the four permanent plots in the Trans-Mexican Volcanic Belt.

populations was counted, and we recorded on average 96 fruits per umbel. Then, for estimating the average number of seeds per capsule, we sampled 57 capsules from three of the populations studied ("Cuanajo", "Icuacato" and "Piedra de Indio"). After capsules dehiscence, we counted the total number of seeds per fruit, identifying viable and unviable seeds by their color. Based on previous experiences we knew that black seeds were viable whereas those with clear color were not. We calculated on average a 375 ± 73 seeds per capsule, and about 36,000 seeds per umbel. In each plot, during the monitoring events, we recorded the number of reproductive individuals, the number of umbels per panicle and the estimated number of seeds per flowering individual.

The second step to estimate fecundity, was carrying out experiments of natural establishment in each population, determining the proportion of seeds that were established one year after their germination. The experiments were established *in situ*, using two gridded (grid square size: 50 cm) subplots (5 x 2.5 m) beneath reproductive individuals, which are areas influenced by seed rain. The number of seeds, viable and unviable were counted in each subplot. One month later, the seedlings with cotyledon inside the subplots were recorded in 2012. One year later, the number of one leaf seedlings established in 2013 was recorded. Unfortunately, in three of the populations these experiments were altered by livestock and confident results were obtained only from one population. The successful subplots experiments were those established in the population "Piedra de Indio". This rate was included in the IPM´s of the four populations since it was the only real-confident data that we could obtain.

IPM´s

Critical life history traits of *Agave inaequidens* were characterized to construct the IPM´s. This plant species is monocarpic, reproducing only by sexual means. During their development, plants grow continually increasing their volume, accumulating carbohydrates. When mature, these reserves are directed to produce a massive inflorescence, flowers, nectar, and seeds. The size of the panicle is proportional to the plant size. *A. inaequidens* is pollinated mainly by *Leptonycteris* bat species (León, 2013). The reproduction causes the death of an agave, but also the production of hundreds of thousands of seeds, and the recruitment rate will depend on the sites and substrate where seeds arrive, the amount of seed predation, and the intensity of precipitation. For this agave, at least two months of high humidity favor germination of practically all viable seeds that were not predated; therefore, the formation of seed banks is unlikely, similarly as in almost all studies in this topic conducted with *Agave*.

After conducting surveys of the annual population dynamics, IPM´s were constructed using the IPMpack package for R (Easterling et al., 2000; Ellner and Rees, 2006; Metcalf et al., 2013; Ortiz-Rodríguez, 2015; R Development Core Team, 2010). This analysis is based on a kernel representing the growth probabilities between continuous size stages that are conditioned by the survival and offspring production. In this study, populations are structured by a set of continuous variability, according to the TFA values of individual plants. For analyzing such structure, we used the following equation:

$$n(y,t+1) = \int_L^U K(y,x)n(x,t)dx = \int_L^U [P(x,y) + F(x,y)]n(x,t)dx \quad (1)$$

Where $n(y,t+1)$ is the size distribution of the established and recruited plants at time $t+1$; $n(x,t)$ is the size distribution of plants at time t; L and U are the IPM in the lowest and highest size limits. The kernel K is divided into the two sub-kernels P and F. The P sub-kernel representing the growth and survival transitions, the F sub-kernel describing the *per-capita* contribution of the reproductive individuals given the recruitment density in the next census. To build K, the growth, survival, and fecundity functions of P and F were calculated from regression models based on the data recorded for each population and period (see **Table 1** and **Appendix 1**). The model was then built by using the midpoint rule (Ellner and Rees, 2006; Zuidema et al., 2010) in order to achieve a numerical integration and obtaining a matrix of 100 x 100 dimensions.

Prospective Analysis

Elasticity Analyses

Through the models referred to above, we performed elasticity analysis (De Kroon et al., 1986; Metcalf et al., 2013) to identify which plant sizes have a stronger effect on the λ values per site and year.

Numerical Simulations of Management Strategies

Based on information about traditional forms of management of *A. inaequidens* (Torres et al., 2015b), we carried out numerical simulations of their possible effects on population dynamics. We simulated the effect of *in situ* management actions *sensu* Blancas et al. (2010), such as tolerance of reproductive individuals, their extraction at different rates, and the enhancement of young plants (individuals 6 to 12 months old grown in nurseries), under different introduction efforts. Simulations were conducted combining variables and scenarios to characterize the effect of different management strategies on λ. Firstly, we calculated the basic K of each population, and then simulations were performed in two nested cycles. The external cycle simulates the extraction of 0 to 100% reproductive plants through 5% intervals. The internal cycle simulates different introduction efforts from 25 to 300 young plants through intervals of 25 plants. The external cycle affects the K diagonal of the reproductive plants, which represents the survival. The value of the diagonal was reduced in 5% intervals, then the percentage of reduction was calculated and subtracted from the initial value. The internal cycle affects the population vector, which allows identifying the sizes of young individuals that should be reintroduced; the total number of plants to be introduced was divided by the number of categories. To calculate the number of plants per category, the augmented categories were added to the initial population vector. With this modified vector and K, we calculated λ for the resulting populations by using 100 iterations.

TABLE 1 | Models of the demographic processes used to construct IPM´s for the first (1st) and second (2nd) monitoring periods (first and second consecutive years of the study, respectively).

Site (monitoring period)	Survival	Growth	Fecundity	Density/size range
Cuanajo (1st period)	y~x+x2: D^2 = 0.051; AIC=267.94	y~x+x2+x3: D^2 = 0.94; AIC=165.48	y~x: D^2 = 0.441; AIC=40.176	480 ind
(2nd period)	y~x+x2: D^2 = 0.056; AIC=248.69	y=0.08708x^2+0.36079x R^2 = 0.85	Null	TFA (cm^2)
				3–126,950
				(log) TFA
				0.52–5.10
Icuacato	y~x+x2: D^2 = 0.154; AIC=339.59	y=0.93871x R^2 = 0.85	y~x: D^2 = 0.44; AIC=40.716	524 ind
(1st period)				TFA (cm^2)
(2nd period)	y~x+x2+x3:D^2 = 0.117; AIC=469.37	y~x+x2+x3: D^2 = 0.83; AIC=491.91	y~x+x2: D^2 = 0.41; AIC=62.597	1–154,608
				(log) TFA
				-0.29–5.19
Piedra de Indio	y~x+x2: D^2 = 0.176; AIC=119.37	y=0.94086x R^2 = 0.959	y~x: D^2 = 0.288; AIC=25.129	158 ind
(1st period)				TFA (cm^2)
(2nd period)	y~x: D^2 = 0.161; AIC=87.277	y=0.70846 x^2+0.03731 x R^2 = 0.967	Null	1–190,336
				(log) TFA
				0.80–5.28
Pino Real	y~x+x2: D^2 = 0.158; AIC=57.287	y~x+x2+x3: D^2 = 0.937; AIC=-30.812	y~x: D^2 = 0.367; AIC=11.488	137 ind
(1st period)				TFA (cm^2)
(2nd period)	y~x: D^2 = 0.007; AIC=39.412	y=0.96131 x R^2 = 0.979	y~x+x2: D^2 = 0.99; AIC=6.0126	10–15,225
				(log) TFA
				1–5.18

Densities correspond to the number of individuals in 2,500 m^2 sampled units, and the range size of individuals were calculated based on Total Foliar Area (TFA) and logarithm TFA. x= plant size at time t.

Influence of Rainfall Variation

To explore in a simple way the possible influence of rainfall variation on species demography, data from the nearest climatic stations to each population were analyzed. To find out if a rainfall series could be represented with a pdf Gamma, we firstly fitted the series to a Gamma distribution (fitdistr function in MASS package) and then, through a test of Kolmogorov-Smirnoff, we determined if the series correspond to the pdf gamma (ks.test function in stats package). When data did not adjust to a Gamma probability, we used a pdf normal (see **Appendix 2**).

For representing two possible, simple, scenarios of change of rainfall patterns (one drier than the other), we used rainfall data recorded during the period when the demographic data were collected. We used the rainfall thresholds corresponding to the first and third years of sampling (519.8 mm and 744.7 mm, respectively). We calculated simulations of 1,000 events of rainfall for each population studied and for each precipitation threshold using parameters of the fitted pdf. For each condition we identified the number of events below (nb) and over (na) the threshold.

We performed projections of each population using the series of simulated rainfall. The demographic model used was: $n(t) = Mx * n(t-1)$ where n is the vector of the population sizes. The kernel Mx used depends on the value of the simulated rainfall. If the rainfall value at the time (t) is lower than the threshold, we used the kernel of the dry period, but if it is higher, we used the kernel of the wet period. The stochastic growth rate was calculated as: $\lambda rand = avg(\frac{n(t)}{n(t-1)})$, considering the last 500 iterations of the series of populational vectors n(t).

For characterizing the effect of the variation of the rainfall simulations over the management strategies, we used the numbers of events below (nb) and over (na) the threshold and the kernels simulating the extraction and introduction in the analyzed populations.

The average matrix was calculated as: $Mavg = ((ml * na) + (m2 * nb))/(na+nb)$ where m1 and m2 are the kernels of management for each annual period analyzed. With the average matrix it was possible to calculate the variation $\{d1 = na*(ml - Mavg)^2; d2 = nb*(m2 - Mavg)^2; v = \frac{di+d2}{na+nb}\}$ and the standard error $sd = sqrt(v); se = \frac{sd}{sprt(na+nb)}$ of each kernel for each population and rainfall threshold.

Integrated Stochastic Rainfall Model

To provide broader scenarios of environmental patterns and to provide general helpful recommendations to agave managers, with the two rainfall patterns studied we calculated two average kernels with the four individual kernels for each period. We looked for obtaining a general view of how management strategies modify growth rates.

RESULTS

Population Dynamics: IPM´s

In the population Cuanajo, a total of 480 agave plants were monitored. In the first period 16 agaves bloomed and the analysis projected λ=0.975 for the period 2011-2012, and λ= 0.899 for the period 2012-2013, when no plants bloomed. In the population Icuacato, a total of 524 agaves were monitored. In the first period six mature agaves bloomed and λ= 1.003 was projected; 10 agaves bloomed in the second one; λ=0.559 as consequence of wildfire that severely affected plants of the smaller sizes.

In the population Piedra de Indio, 158 agave plants were monitored, three of them bloomed and λ=0.976 was projected, but no plant flowered in the second period and had λ=0.955. In the population Pino Real a total of 137 agaves were monitored. In the first period one agave bloomed and the population had λ= 0.952, and in the second period one agave bloomed and the population had λ= 0.966.

Prospective Analysis

Elasticity Analysis

In the population Cuanajo, during the first period of the analysis we found that the individuals of intermediate and larger sizes had a greater effect on λ, whereas the intermediate stages were more relevant in the second period. In the population Icuacato, the larger stages were more relevant in the first period and the smaller stages during the second period that suffered the wildfire (**Figure 3**).

In the population Piedra de Indio, the individuals of intermediate and larger stages had a greater effect on λ in the first period, whereas the larger stages were more relevant in the second period. In the population Pino Real, the larger stages were more important in the first period, whereas the intermediate and the larger intermediate stages were relevant in the second period (**Figure 4**).

Management and Population Dynamics Simulations

For the population Cuanajo, during the first period studied, the projected management simulations showed that extraction rates from 10% to 30% of mature individuals may maintain λ > 1, only if 200 to 300 agave plants of the lower sizes (plantlets 6 to 12 months old grown in nurseries), are introduced into the population, respectively (**Figure 5A**) (**Table 2** shows the amount of mature agaves, that represent the percentage of extraction). For the second period, no combination of introduction and extraction allowed to maintain λ > 1 (**Figure 5B**).

In the population Piedra de Indio, during the first period, the extraction rate of 5% of mature individual would maintain λ > 1 if more than 200 agaves of the lower sizes are reintroduced to the population, and up to 60% if > 300 agaves are introduced into the population (**Figure 5C** and **Table 3** show the amount of mature agaves representing the percentage of extraction). In the second period, the extraction of 5% would be sustainable if more than 200 young agaves are introduced to the population, and from 40% up to 100% if > 275 young agaves are introduced into the population (**Figure 5D**, and **Table 4** show the amount of mature agaves, that represent the percentage of extraction).

For the population Pino Real, the simulations showed that for the first period that an extraction rate of 20% to 40% of mature individuals would maintain λ > 1 if more than 25 to 275 agaves are introduced in the population (**Figure 5E** and **Table 5** show the amount of mature agaves that represent the percentage of extraction). In the second period no combination of management strategies allowed to maintain λ > 1 (**Figure 5F**).

For the population Icuacato, in the first period studied, an extraction rate of 5% of adult plants can be compensated if more than 50 young agaves (6 to 12 months old) are reintroduced into the population, an extraction rate of 10% of adult plants may be sustainable if more than 75 young plants are reintroduced, and up to 100% extraction of adult plants if 275 or more young agaves are

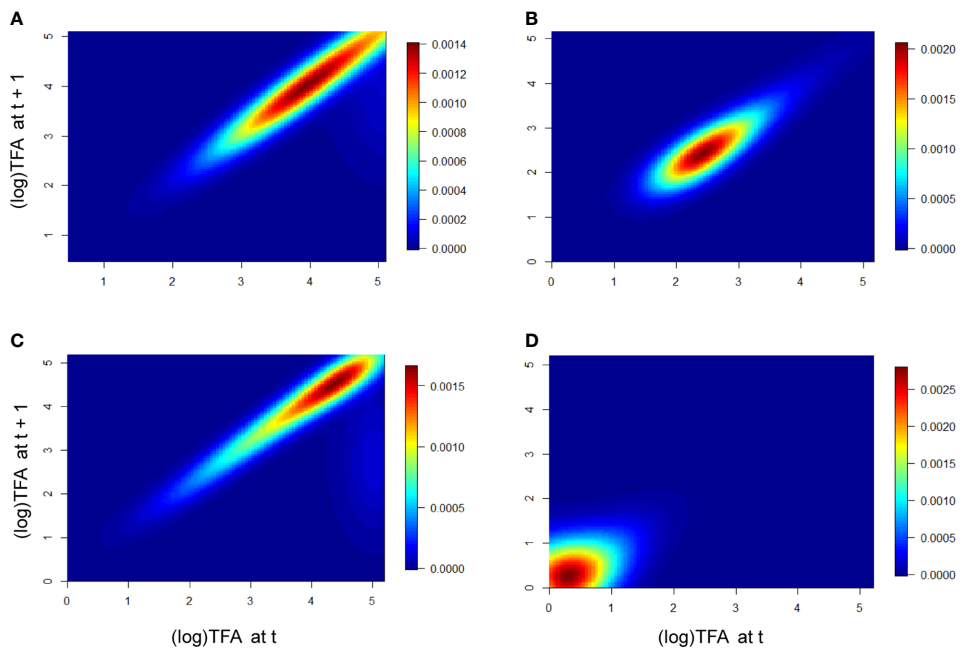


FIGURE 3 | Elasticity of the populations Cuanajo and Icuacato. Cuanajo **(A)** first period, **(B)** second period. Icuacato **(C)** first period, **(D)** second period (wildfire). Warmer colors represent size stages that mostly contributed to λ.

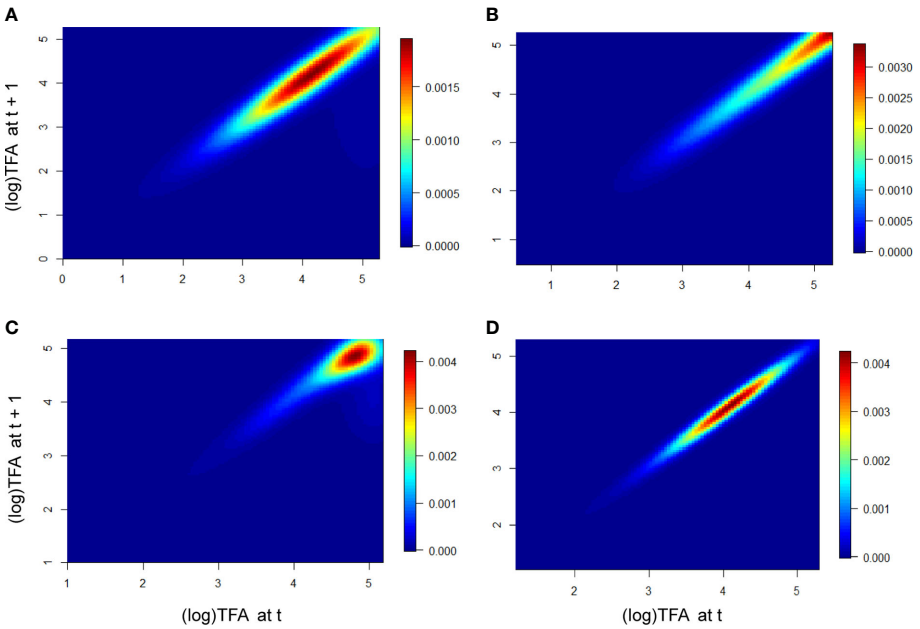


FIGURE 4 | Elasticity of the populations Piedra de Indio and Pino Real. Piedra de Indio **(A)** first period, **(B)** second period. Pino Real **(C)** first period, **(D)** second period. Warmer colors represent size stages that mostly contributed to λ.

introduced to the population; all these scenarios would maintain λ > 1 (**Figure 5G**, and **Table 6** show the amount of mature agaves that represent the percentage of extraction). For the second period, no combination of introduction and extraction allowed to maintain λ > 1 (**Figure 5H**).

Stochastic Rainfall Model

The projected variation of the numerical simulations in the stochastic rainfall models renders or reproduces the average patterns (i.e., the magnitude of variation does not change depending on the extent of exploitation or reintroduction). Upper rainfall thresholds projected a higher λ value, except in population Pino Real (**Figure 5** upper shaded sheets, **Table 7**). These models showed that management simulations may be influenced by rainfall patterns, in some cases lower rain scenarios determine that no extraction would be recommendable. In the first period (**Figure 5A**), we can find combinations of extraction and reintroduction that produce growth rates greater than one. In general, in all the populations, to maintain λ in equilibrium or above one, is needed a combination of low extraction with reintroduction of many individuals. However, in the population of Pino Real (scenario A), it seems that the population could support a maximum extraction rate of 20% to 40%, regardless of introduction rates. The numerical simulations on scenario B suggest that growth rates are lower than one.

Integrated Stochastic Rainfall Model

A general scenario can be projected from the average models for general recommendations. In the first period, a threshold of 10% to 30% of extraction combined with an introduction of 175 to 300 individuals can be performed to maintain λ above 1 (**Figure 6A**). This model showed also that management simulations may be influenced by rainfall patterns, in some cases, scenarios with lower rainfall determine that no extraction would be recommendable. In the second period, no combination of management actions allows to maintain λ above 1 (**Figure 6B**). This pattern may be influenced by the integration of the kernel of the second period of Icuacato population were the wildfire occurred.

DISCUSSION

Population Dynamics and Elasticity

Our models showed that in the first period of the study, λ is close to the population equilibrium, i.e., the mortality rate is balanced with the recruitment rate and if those conditions are maintained, the populations would remain stable throughout time. The elasticity analyses showed that for all models we constructed, the alterations in the vital rates of the larger size categories, i.e., the nearly mature and already mature agaves that are harvested for mescal production have greater effects on λ, as it is shown in our prospective analyses. Only in the population Pino Real, λ values, in both study periods, had values slightly below the equilibrium, suggesting a decreasing trend, but these values were not far from the unity. Similarly as found in other works with *Agave* species of semiarid zones of Mexico (Illsley et al., 2007; Jiménez-Valdés et al., 2010; Torres et al., 2015a; Arias-Medellín et al., 2016), survival and growth are the vital rates that mostly contribute to λ in *A. inaequidens*, but in this study we

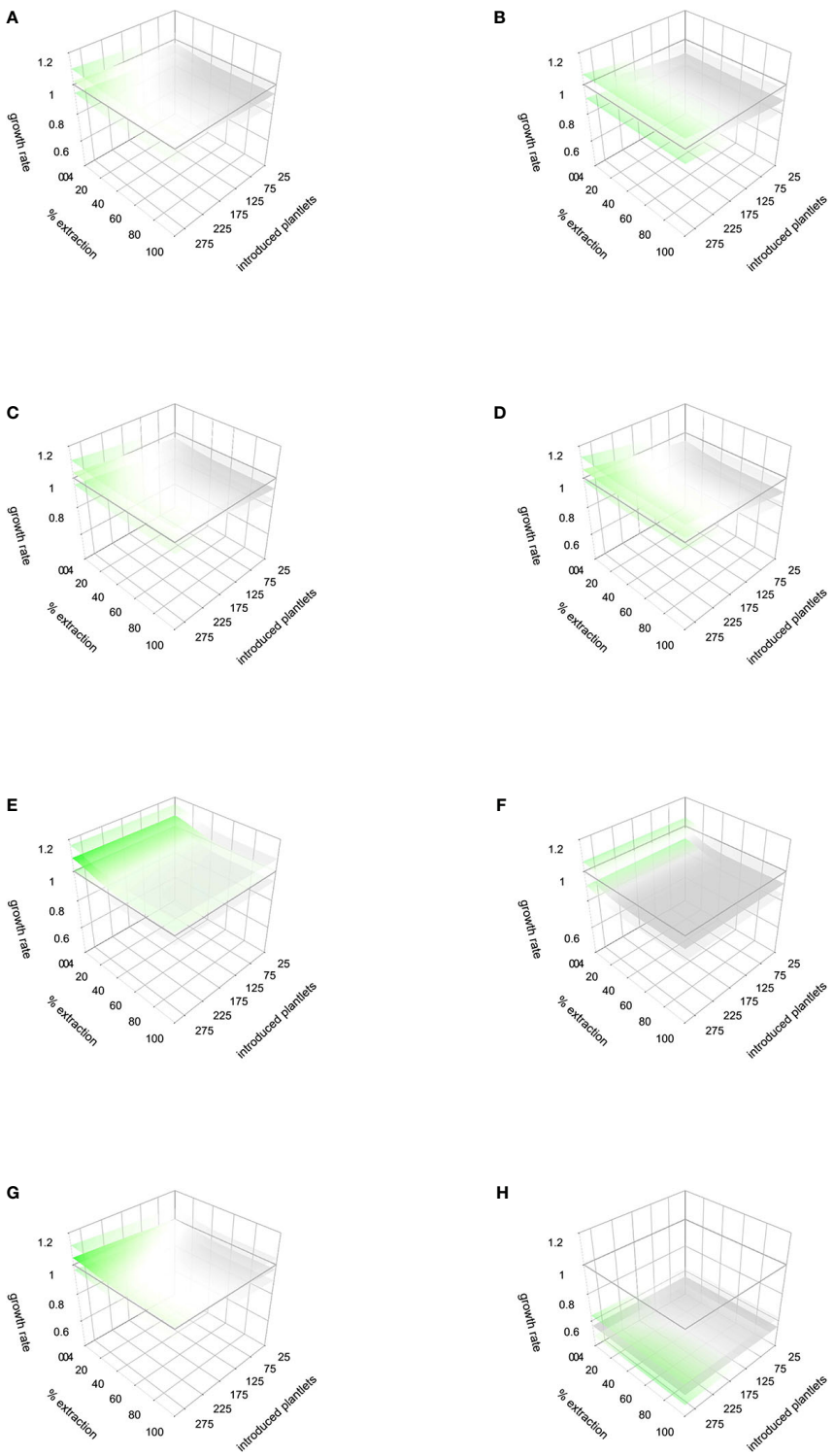


FIGURE 5 | Response to management simulations and stochastic rainfall models for the two monitored periods (first and second years of the study, respectively). Cuanajo **(A)** first period, **(B)** second period. Icuacato **(C)** first period, **(D)** second period (Forest fire). Piedra de Indio: **(E)** first period, **(F)** second period. Pino Real: **(G)** first period, **(H)** second period. Values of λ below 1 are gray colored, values equals 1 are white colored, and λ values over 1 are green colored. Lower and upper λ responses to the lower and upper rainfall thresholds are represented in shaded sheets.

TABLE 2 | Different combinations of thresholds of management actions to maintain λ above unity in the population Cuanajo during the first monitoring period of the study.

Management action	First period		
Extraction of mature individuals	10% 26 ind.	20% 52 ind.	30% 78 ind.
Reforestation (number of plantlets needed)	200	250	300

TABLE 3 | Different combinations of thresholds of management actions to maintain λ above unity in the population Piedra de Indio during the first monitoring period of the study.

Management action	First period				
Extraction of mature individuals	5% 2 ind.	15% 8 ind.	25% 14 ind.	30% 16 ind.	60% 33 ind.
Reforestation (number of plantlets)	200	225	250	275	300

TABLE 4 | Different combinations of thresholds of management actions to maintain λ above unity in the population Piedra de Indio during the second monitoring period of the study.

Management action	Second period		
Extraction of mature individuals	20% 11 ind.	35% 19 ind.	40%–100% 22–56 ind.
Reforestation (number of plantlets needed)	225	250	275

TABLE 5 | Different combinations of thresholds of management actions to maintain λ above unity in the population Pino Real during the first monitoring period of the study.

Management action	First period
Extraction of mature individuals	20%–40% 11–22 ind.
Reforestation (number of plantlets)	25 to 275

TABLE 6 | Different combinations of thresholds of management actions to maintain λ above 1 in the population Icuacato in the first monitoring period.

Management action	First period						
Extraction of mature individuals	10% 9 ind.	20% 18 ind.	25% 22 ind.	30% 27 ind.	40% 36 ind.	50% 45 ind.	100% 91 ind.
Reforestation (number of plantlets needed)	75	100	150	175	200	250	275

found that fecundity also showed a significant contribution to λ, a fact that can be explained based on findings by Metcalf et al. (2003). These authors documented demographic patterns of monocarpic perennials and analyzed them as a function of individual's size, finding a direct relationship between plant size and the number of seeds and seedlings established. *A. inaequidens* has larger individuals than other species whose demography has been studied (*A. angustifolia*, *A. cupreata*, *A. marmorata*, and *A. potatorum*), and in addition, it occurs in temperate zones with higher rainfall regimes than the semiarid areas where the other species are distributed. The greater the number of seeds in a more favorable environment, and the higher fertility values are expected to contribute to the population dynamics of this species.

TABLE 7 | Growth rate values (λ) (average, variation, and standard errors) in populations at different rainfall thresholds.

Site	Rainfall threshold (mm of annual rain)	Avg growth rate (λ)	SE	λ >1	λ <1
Cuanajo	519.8	1.1376	0.0121	599	401
Cuanajo	744.7	1.2090	0.0145	602	398
Icuacato	519.8	1.1230	0.0103	624	376
Icuacato	744.7	1.2142	0.0149	497	503
Piedra	519.8	1.0472	0.0057	562	438
Piedra	744.7	1.0848	0.0090	557	443
Pino Real	519.8	0.9916	0.0013	7	993
Pino Real	744.7	0.9677	0.0025	76	924

The last two columns show the number of times the λ value showed in the model. The sum of the two values equals 1,000, the simulation size.

A. inaequidens is totally dependent on its reproduction *via* seeds, unlike other agave species that have active asexual propagation and produce stoloniferous shoots during their growth, as in the cases of *A. tequilana*, *A. angustifolia*, and *A. salmiana*. This trait has defined that the management and propagation of those species, predominantly through the transplantation of suckers, which has determined unnecessary to allow individuals blooming to obtain seeds. This management type allows to obtain agaves with the same characteristics of the mother plant, a useful aspect when seeking to unify production. However, it has great repercussion on decreasing genetic variation, which in turn determines low capacity of resistance against pests and effects of climate change. Exclusive sexual reproduction of *A. inaequidens* allows and enhances the conservation of high levels of genetic variation, as shown in recent studies (Figueredo et al., 2015; Figueredo-Urbina et al., 2017).

The stochastic rainfall models as well as the integrated rainfall model help to identify how much oscillations in rainfall may affect the growth of agaves populations. As general trends, these models suggest that dry years would be particularly severe for harvesting reproductive agaves. The unpredictability of rainfall can be used as a justification for lowering harvest rates, in order to maintain agave populations in the long term. But in addition, this research line deserves special attention since climate change may be increasing the variation of climate regimes (IPPC, 2015).

Implications for Sustainable Extraction

Regional *Agave inaequidens* managers and mescal producers in northern Michoacán, practice several management strategies, which represent a gradient of management complexity that goes from simple gathering to intensive *ex situ* cultivation. *In situ* management in wild populations in forests is carried out by a small number of managers. In those populations, people use to transplanting juvenile agave plants from undesirable to desirable places, weeding competing plants, let standing mature individuals for seed production and natural dispersal. They, in addition, collect seeds and sow them in rustic nurseries, and some of the resulting plantlets

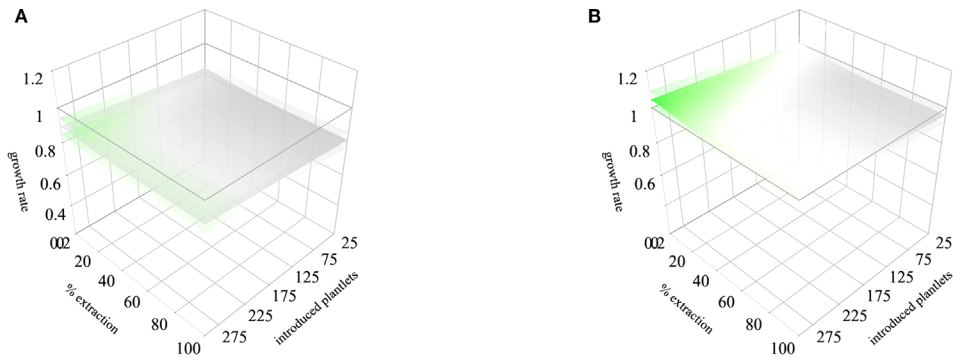


FIGURE 6 | Integrated stochastic rainfall models for the two monitored periods (first and second years of the study, respectively. The scenario **(A)** corresponds to the first monitored period and scenario **(B)** corresponds to the second monitored period. Lower and upper lambda responses to the lower and upper rainfall thresholds are represented in shaded sheets. Lower λ values are in gray, and higher in green.

are targeted to *ex situ* cultivation, but some others are destined to *in situ* reforestation efforts (Torres et al., 2015b). Our findings of the previous and current studies suggest that, considering the number of agaves per batch of mescal production, (on average 147 mature agaves), and based on our prospective models, each population have different extraction thresholds. The main recommendation, based on our models and the integration of the rainfall patterns in wild populations, is to harvest in a precautionary threshold of 10 to 30% of mature individuals in areas of 2,500m^2, tolerating individuals, preferably the larger ones, since seeds production is larger in these individuals, to maintain natural seed production and recruitment of new individuals. We recommend emphatically to let standing larger individuals, particularly of this species, since an historical decrease in the size of wild individuals have been documented as a result of harvesting the larger individuals for mescal production (Torres et al., 2015b). In this way, promoting that larger individuals' reincorporation to the populations, would determine higher yield of seeds and stems to the mescal production. This threshold must be accompanied with the sowing of seeds, nursing, and introduction agave plantlets to wild populations from 175 to 300 individuals. These reforestation efforts must be carried out just before the rainy season starts in order to enhance the survival of this younger agaves.

Communities of managers of *Agave inaequidens*, and producers of mescal in northern Michoacán are rural "ejidatarios", and their territories comprise hundreds of hectares with wild populations of this species, which means that a greater area than at present could be managed, and the impact on populations would be buffered if more populations are used extracting fewer individuals per population. Since managers relay only on wild populations to complete a batch (the rest are agaves from plantations and other agroforestry systems), the harvesting thresholds discussed could be used as a guide in different populations within the communitarian territories. Other aspects are relevant when discussing social and economic aspects of the problems related to using agaves. For instance, the diversification of species used, the diversification of sources of agaves, the diversification of activities supporting local households' economy, and, importantly, diversifying markets looking for fair principles of commercialization, are all important issues for designing sustainable alternatives.

We have identified the average number of mature agaves to complete a batch, and the variation of environmental conditions (interannual and long-term variations) and their effects on populations dynamics. In addition, we have identified market pressures motivating increasing extraction of mature agaves and responses of management techniques developed by people to attenuate the effects of extraction on forests. Cultivation of agaves emerged as an alternative, but cultivators of agave plantations have drastic economic problems, because of plant mortality, pests, and slow growth of agaves. These factors have led people to use chemical inputs, which has determined economic and ecological negative impacts. We are therefore before a complex socio-ecological system that requires social and ecological holistic approaches to be faced. We have documented details of the management techniques developed by local people, and these can be consulted in our study (Torres et al., 2015b). All these techniques are valuable tools for solving problems, and our ecological studies, some of them reported here, provide elements to complement such techniques, considering biological and ecological factors that may help to finer adjustments of the techniques for specific situations. But particularly important are the social agreements and regulations for an organized way of using common resources. Basic principles proposed by Ostrom (1990); Dietz et al. (2003); Ostrom et al. (2012), as well as the valuable experiences of social organization proliferating in Mexican rural communities and ejidos are of great value. This is for instance the case of the organization *Sansekan Tinemi* and those of other organizations of Nahua people from Guerrero (Martin et al., 2011; Illsley et al., 2018). Therefore, academic ecological studies find in local people's organizational experiences the possibility to produce knowledge useful to answer scientific questions but also to solve problems.

CONCLUSIONS

Depending on the different levels of extraction and the areas needed to supply of raw material for the production of distilled spirits, our models suggest that it is necessary to carry out actions of reforestation, and *in situ* management of agaves according to

the trends found in each site. This is one indispensable condition to maintain λ close to or greater than 1. The models also indicate that management actions must be specific, according to characteristics of each population, like density and structure, and the fluctuating climatic conditions.

The accelerated demand of mescal production and the patterns of exploitation that have predominated, determine in some cases a drastic depletion of wild agave populations. In some of the analyzed localities, these conditions drive to carry immediate management actions. Although our models were carried out in particular systems and climatic conditions for the sampling years, and has limitations in characterizing fecundity, there is an urgent need to implement comprehensive management actions. It is feasible to propose actions and monitoring their success or the responses of populations to such management practices. And, according to the patterns found in these monitoring activities, to continue or to reconsider adjustments of actions, based on an adaptive management framework principles (Christensen et al., 1996).

DATA AVAILABILITY STATEMENT

All the datasets generated for this study can be found by asking to the first author.

AUTHOR CONTRIBUTIONS

IT-G: Conception and design of the research, fieldwork gathering data, data analysis, writing the manuscript, and reviewing several versions of manuscript. AL-J: Fieldwork gathering data. EV: Analysis design, data analyzing, analysis interpretation, and writing the manuscript. AM-C: Writing the manuscript and reviewing of manuscript. AC: Conception and design of the research, analysis interpretation, and writing and reviewing several versions of manuscript.

FUNDING

We thank the Dirección General de Asuntos del Personal Académico (DGAPA) for the postdoctoral scholarship awarded to the first author. Also, for financial support to the PAPIIT projects IN206217 and IN206520.

We also thank the Consejo Nacional de Ciencia y Tecnología CONACYT for financial support to the project A1-S-14306, and to the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) for supporting the project RG023.

We also thank the Red Temática de Sistemas Agroforestales de México (RedSAM) project 293348 and the Red Temática de Productos Forestales No Maderables (RED-PFNM), project 293914 of CONACYT for the support provided to this research.

ACKNOWLEDGMENTS

This study is dedicated to the loving memory of the second author: Alejandro León Jacinto. We also want to say sincerely thank you to the communities of Michoacán that own the wild populations studied and that let us to perform this study in their land. Special thanks to the authorities of Quiroga, Pino Real and Cuanajo, Michoacán.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.01224/full#supplementary-material

REFERENCES

Arias-Medellín, L. A., Bonfil, C., and Valverde, T. (2016). Demographic analysis of *Agave angustifolia* (Agavaceae) with an emphasis on ecological restoration. *Bot. Sci.* 94 (3), 513–530. doi: 10.17129/botsci.525

Baraza-Ruiz, E., and Estrella-Ruiz, J. P. (2008). Manejo sustentable de los recursos naturales guiado por proyectos científicos en la mixteca poblana mexicana. *Rev. Ecosistemas* 17 (2), 3–9.

Blancas, J., Casas, A., Rangel-Landa, S., Moreno-Calles, A.II, Torres, I., Peréz-Negrón, E., et al. (2010). Plant management in the Tehuacán-Cuicatlán Valley Mexico. *Econ. Bot.* 64, 287–302. doi: 10.1007/s12231-010-9133-0

Bowen, S., and Zapata, A. V. (2009). Geographical indications, terroir, and socioeconomic and ecological sustainability: the case of Tequila. *J. Rural Stud.* 25, 108–119. doi: 10.1016/j.jrurstud.2008.07.003

Caswell, H. (2001). *Matrix population models: Construction, analysis and interpretation. 2nd edn* (Sunderland, MA: Sinauer Associates, Inc.).

Christensen, N. L., Bartuska, A. M., Brown, J. H., Carpenter, S., Antonio, C. D., Francis, R., et al. (1996). The Report of the Ecological Society of America Committee on the Scientific Basis for Ecosystem Management. *Ecol. Appl.* 6 (3), 657–691. doi: 10.2307/2269460

Colunga-GarcíaMarín, P., Larqué-Saavedra, A., Eguiarte, L., Largué, S. A., and Zizumbo-Villarreal, D. (2007). *En lo ancestral hay futuro: del tequila, los mezcales y otros agaves*. (Mérida: Centro de Investigación Científica de Yucatán, AC. Consejo Nacional de Ciencia y Tecnología-Comisión Nacional para el Conocimiento y Uso de la Biodiversidad), 402 pp.

Colunga-GarcíaMarín, P., Torres-García, I., Casas, A., Figueredo-Urbina, C. J., Rangel-Landa, S., Delgado-Lemus, A., et al. (2017). "Los agaves y las prácticas mesoamericanas de aprovechamiento, manejo y domesticación," in: *Domesticación en el Continente Americano Vol 2*. Eds. A. Casas, J. Torres-Guevara and F. Parra (Lima: Universidad Nacional Autónoma de México-Universidad Nacional Agraria de La Molina). pp. 273–308.

De Kroon, H., Plaisier, A., van Groenendal, J., and Caswell, H. (1986). Elasticity: the relative contribution of demographic parameters to population growth rate. *Ecology* 67, 1427–1431. doi: 10.2307/1938700

Dietz, T., Ostrom, E., and Stern, P. C. (2003). The struggle to govern the commons. *Science* 302 (5652), 1907–1912. doi: 10.1126/science.1091015

Easterling, M., Ellner, S. P., and Dixon, P. (2000). Size-Specific Sensitivity: Applying a New Structured Population Model. *Ecology* 81 (3), 694–708. doi: 10.1890/0012-9658(2000)081[0694:SSSAAN]2.0.CO;2

Ellner, S. P., and Rees, M. (2006). Integral projection models for species with complex demography. *Am. Nat.* 167 (3), 410–428. doi: 10.1086/499438

Estrella-Ruiz, P. (2008). *Efecto de la explotación humana en la biología de la polinización de Agave salmiana y Agave potatorum en el Valle de Tehuacán-Cuicatlán*. (México: UNAM). (M. Sc. thesis).

Fick, S. E., and Hijmans, R. J. ,. (2017). WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37 (12), 4302–4315. doi: 10.1002/joc.5086

Figueredo, C. J., Casas, A., González-Rodríguez, A., Nassar, J. M., Colunga-GarcíaMarín, P., and Rocha-Ramírez, V. (2015). Genetic structure of coexisting wild and managed agave populations: implications for the evolution of plants under domestication. *AoB Plants* 7, plv114. doi: 10.1093/aobpla/plv114

Figueredo-Urbina, C. J., Casas, A., and Torres-García, I. (2017). Morphological and genetic divergence between *Agave inaequidens*, *A. cupreata* and the domesticated *A. hookeri*. Analysis of their evolutionary relationships. *PloS One* 12 (11), e0187260. doi: 10.1371/journal.pone.0187260

Gaoue, O. G., and Ticktin, T. (2010). Effects of harvest of Non Timber Forest Products and ecological differences between sites on the demography of African Mahogany. *Conserv. Biol.* 24 (2), 605–614. doi: 10.1111/j.1523-1739.2009.01345.x

Gaoue, O. G., Sack, L., and Ticktin, T. (2011). Human impacts on leaf economics in heterogeneous landscapes: The effect of harvesting Non-Timber Forest Products from African Mahogany across habitats and climates. *J. Appl. Ecol.* 48 (4), 844–852. doi: 10.1111/j.1365-2664.2011.01977.x

García-Mendoza, A. J., Franco-Martínez, I. S., and Sandoval-Gutiérrez, D. (2019). Cuatro especies nuevas de Agave (Asparagaceae, Agavoideae) del sur de México. *Acta Bot. Mex.* 126, 1–18. doi: 10.21829/abm126.2019.1461

Gentry, H. S. (1982). *Agaves of Continental North America* (Tucson, AZ: University of Arizona Press).

Illsley, C., Vega, E., Pisanty, I., Tlacotempa, A., García, P., Morales, P., et al. (2007). "Maguey papalote: Hacia el manejo campesino sustentable de un recurso colectivo en el trópico seco de Guerrero, México," in *En lo ancestral hay futuro: del tequila, los mezcales y otros agaves*, (Merida: Centro de Investigación Científica de Yucatán, AC. Consejo Nacional de Ciencia y Tecnología-Comisión Nacional para el Conocimiento y Uso de la Biodiversidad), 319–338.

Illsley, C., Torres-García, I., Hernández-López, J. J., Morales-Moreno, P., Varela-Álvarez, R., Ibañez-Couch, I., et al. (2018). *Manual de manejo campesino de magueyes mezcaleros forestales* (México: Grupo de Estudios Ambientales AC).

Intergovernmental Panel on Climate Change (IPPC). (2015). *Climate Change 2014: Mitigation of Climate Change Vol. 3* (Cambridge, United Kingdom: Cambridge University Press).

Jiménez-Valdés, M., Godínez-Alvarez, H., Caballero, J., and Lira, R. (2010). Population dynamics of *Agave marmorata* Roezl. under two contrasting management systems in central Mexico. *Econ. Bot.* 64, 149–160. doi: 10.1007/s12231-010-9117-0

Jordan, P. W., and Nobel, P. S. (1979). Infrequent establishment of seedlings of *Agave deserti* (Agavaceae) in the northwestern Sonoran Desert. *Am. J. Bot.* 66 (9), 1079–1084. doi: 10.1002/j.1537-2197.1979.tb06325.x

Kuss, P., Rees, M., Ægisdóttir, H. H., Ellner, S. P., and Stöcklin, J. (2008). Evolutionary demography of long-lived monocarpic perennials: A time-lagged integral projection model. *J. Ecol.* 96 (4), 821–832. doi: 10.1111/j.1365-2745.2007.0

León, A. (2013). *Aspectos de la fenología, visitantes florales y polinización de Agave inaequidens Koch.ssp. inaequidens (Agavaceae), en el estado de Michoacán." Dissertation/Bachelor* (Morelia, Michoacán, México: Universidad Michoacana de San Nicolas de Hidalgo).

MacNeish, R. S. (1967). "A summary of the subsistence." in: *The prehistory of the Tehuacán Valley. Volume one: Environment and subsistence*. Ed. D. S. Byers (Austin, TX: University of Texas Press). pp: 290–331.

Martin, M. P., Peters, C. M., Palmer, M.II, and Illsley, C. (2011). Effect of habitat and grazing on the regeneration of wild *Agave cupreata in* Guerrero, Mexico. *For. Ecol. Manage.* 26 (8), 1443–1451. doi: 10.1016/j.foreco.2011.06.045

Merow, C., Dahlgren, J. P., Metcalf, C. J. E., Childs, D. Z., Evans, M. E. K., Jongejans, E., et al. (2014). Advancing population ecology with integral projection models: A practical guide. *Methods Ecol. Evol.* 5 (2), 99–110. doi: 10.1111/2041-210X.12146

Metcalf, C. J. E., Rose, K. E., and Rees, M. (2003). Evolutionary demography of monocarpic perennials. *Trends Ecol. Evol.* 18 (9), 471–480. doi: 10.1016/S0169-5347(03)00162-9

Metcalf, C. J. E., Rees, M., Alexander, J. M., and Rose, K. (2006). Growth-survival trade-offs and allometries in rosette-forming perennials. *Funct. Ecol.* 20 (2), 217–225. doi: 10.1111/j.1365-2435.2006.01084.x

Metcalf, C. J. E., McMahon, S. M., Salguero-Gómez, R., and Jongejans, E. (2013). IPMpack: An R Package for Integral Projection Models. *Methods Ecol. Evol.* 4 (2), 195–200. doi: 10.1111/2041-210x.12001

Nobel, P. S., and Valenzuela, A. G. (1987). Environmental responses and productivity of the cam plant, *Agave tequilana*. *Agric. For. Meteorol.* 39 (4), 319–334. doi: 10.1016/0168-1923(87)90024-4

Nobel, P. S. (1992). Annual variations in flowering percentage, seedling establishment and ramet production for a desert perennial. *Int. J. Plant Sci.* 153 (1), 102–107. doi: 10.1086/297011

Nobel, P. S. (2003). *Environmental biology of agaves and cacti* (Cambridge, UK: Cambridge University Press).

Ortiz-Rodríguez, I. A. (2015). *Aumento poblacional de una palma tropical y sus consecuencias sobre la comunidad arbórea: Exploración de factores causales usando modelos integrales de proyección." Dissertation/masters* (Ciudad de México: Universidad Nacional Autónoma de México).

Ostrom, E., Chang, C., Pennington, M., and Tarko, V. (2012). "The Future of the Commons-Beyond Market Failure and Government Regulation," in *Institute of Economic Affairs Monographs* (London, UK: The Institute of Economic Affairs).

Ostrom, E. (1990). "Governing the commons," in *The evolution of institutions for collective actions* (Cambridge, U.K: Cambridge University Press).

R Development Core Team. (2010). *R: A language and environment for statistical computing*. (Vienna, Austria: R Foundation for Statistical Computing).

Ramírez-Rodríguez, R. (2018). *La querella por el pulque: auge y ocaso de una industria mexicana, 1890-1930*. (Zamora, Michoacán: El Colegio de Michoacán), pp.507.

Ramula, S., Rees, M., and Buckley, Y. M. (2009). Integral projection models perform better for small demographic data sets than matrix population models: A case study of two perennial herbs. *J. Appl. Ecol.* 46 (5), 1048–1053. doi: 10.1111/j.1365-2664.2009.01706.x

Rangel-Landa, S., Casas, A., and Dávila, P. (2015). Facilitation of *Agave potatorum*: an ecological approach for assisted population recovery. *For. Ecol. Manage.* 347, 57–74. doi: 10.1016/j.foreco.2015.03.003

Rocha, M., Good-Avila, S. V., and Molina-Freaner, F. (2006). Pollination biology and adaptive radiation of Agavaceae, with special emphasis on the genus Agave. *Aliso* 22 (1), 329–344. doi: 10.5642/aliso.20062201.27 http://web.ecologia.unam.mx/laboratorios/evolucionmolecular/homes/pdfslab/Rocha-etal_2006.pdf.

Schmidt, I. B., Mandle, L., Ticktin, T., and Gaoue, O. G. (2011). What do matrix population models reveal about the sustainability of non-timber forest product harvest? *J. Appl. Ecol.* 48 (4), 815–826. doi: 10.1111/j.1365-2664.2011.01999.x

Shackleton, C. M., Pandey, A. K., and Ticktin, T. (2015). *Ecological sustainability for non-timber forest products: dynamics and case studies of harvesting* (New York, NY: Routledge).

The Angiosperm Phylogeny Group III (2009). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161 (2), 105–121. doi: 10.1111/j.1095-8339.2009.00996.x

Torres, I., Casas, A., Delgado-Lemus, A., and Rangel-Landa, S. (2013). Aprovechamiento, demografía y establecimiento de *Agave potatorum* en el Valle de Tehuacán, México: Aportes ecológicos y etnobiológicos para su manejo sustentable. *Zonas Áridas* 15 (1), 92–109. doi: 10.21704/za.v15i1.110.

Torres, I., Casas, A., Vega, E., Martínez-Ramos, M., and Delgado-Lemus, A. (2015a). Population dynamics and sustainable management of mescal agaves in central Mexico: *Agave potatorum* in the Tehuacán-Cuicatlán Valley. *Econ. Bot.* 69 (1), 26–41. doi: 10.1007/s12231-014-9295-2

Torres, I., Blancas, J., León, A., and Casas, A. (2015b). TEK, local perceptions of risk, and diversity of management practices of *Agave inaequidens* in Michoacán, Mexico. *J. Ethnobiol. Ethnomed.* 11 (1), 61. doi: 10.1186/s13002-015-0043-1

Torres-García, I. (2016). "Aprovechamiento de agaves mezcaleros en el centro de México: Una aproximación socio–ecológica para su manejo sustentable." PhD dissertation in *Posgrado en Ciencias Biológicas* (Morelia, Michoacán, México: Universidad Nacional Autónoma de México).

Torres-García, I., Rendón-Sandoval, F. J., Blancas, J., Casas, A., and Moreno-Calles, I. (2019). The genus Agave in agroforestry systems of Mexico. El género Agave en los sistemas agroforestales de México. *Bot. Sci.* 97, 261–288. doi: 10.17129/botsci.2202

Trejo-Salazar, R. E., Eguiarte, L. E., Suro-Piñera, D., and Medellin, R. A. (2016). Save our bats, save our tequila: industry and science join forces to help bats and agaves. *Natural Areas J.* 36 (4), 523–530. doi: 10.3375/043.036.0417

Valenzuela-Zapata, A. G., Lopez-Muraira, I., and Gaytán, M. S. (2011). Traditional knowledge, *Agave inaequidens* (Koch) conservation, and the charro lariat artisans of San Miguel Cuyutlán, Mexico. *Ethnobiol. Lett.* 2, 72–80. doi: 10.14237/ebl.2.2011.24

Valenzuela-Zapata, A. G. , and Macías-Macías, A. (2014). *La indicación geográfica Tequila: lecciones de la primera denominación de origen mexicana*. (México DF: Comisión Nacional para el Conocimiento y uso de la Biodiversidad), pp. 125

Zuidema, P. A., Jongejans, E., Chien, P. D., During, H. J., and Schieving, F. (2010). Integral projection models for trees: a new parameterization method and a validation of model output. *J. Ecol.* 98 (2), 345–355. doi: 10.1111/j.1365-2745.2009.01626.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer [AM-L] declared a shared affiliation, though no other collaboration, with the authors [IT-G, AL-J, EV, AM-C, AC] to the handling Editor.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 08 September 2020
doi: 10.3389/fpls.2020.524812

Morphological Diversity and Genetic Relationships in Pulque Production Agaves in Tlaxcala, Mexico, by Means of Unsupervised Learning and Gene Sequencing Analysis

Laura Trejo[1]*, *Miguel Reyes*[2], *Daniela Cortés-Toto*[2], *Elvira Romano-Grande*[1] *and Lizbeth L. Muñoz-Camacho*[1]

[1] Laboratorio de Biodiversidad y Cultivo de Tejidos Vegetales, Instituto de Biología, Universidad Nacional Autónoma de México, Tlaxcala, Mexico, [2] Departamento de Actuaría, Física y Matemáticas, Universidad de las Américas Puebla, Puebla, Mexico

OPEN ACCESS

Edited by:
Karolina Heyduk,
University of Hawaii, United States

Reviewed by:
Ryan Stewart,
Brigham Young University,
United States
Jose Alberto Narvaez-Zapata,
Instituto Politécnico Nacional, Mexico

***Correspondence:**
Laura Trejo
laura.trejo@st.ib.unam.mx

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: 06 January 2020
Accepted: 20 August 2020
Published: 08 September 2020

Citation:
Trejo L, Reyes M, Cortés-Toto D, Romano-Grande E and Muñoz-Camacho LL (2020) Morphological Diversity and Genetic Relationships in Pulque Production Agaves in Tlaxcala, Mexico, by Means of Unsupervised Learning and Gene Sequencing Analysis. Front. Plant Sci. 11:524812. doi: 10.3389/fpls.2020.524812

Pulque is one of the oldest fermented beverages, with its origins dating back to pre-Hispanic Mexico. Recently, public consumption has increased. However, the majority of *Agave* plantations for pulque production have disappeared or been abandoned in recent decades. To create strategies for the conservation and production of pulque agaves, it is necessary to first determine their taxonomic identities and to better understand their genetic and morphological diversity. Despite the historical importance of pulque in Mexico, little attention has been placed on the study of *Agave* plants used for its production. Therefore, we analyzed the morphological diversity of vegetative characters of nine landraces of two *Agave* species (*A. salmiana* and *A. mapisaga*) which are widely cultivated for pulque production in Tlaxcala, Mexico. The analysis of morphological characters showed that the landraces largely clustered based on classic taxonomic relationships. One cluster of landraces associated with *Agave mapisaga* var. *mapisaga* and another with *A. salmiana* subsp. *salmiana*, but with the exception of *A. salmiana* subsp. *salmiana* "Ayoteco", which is more closely related with *A. mapisaga* var. *mapisaga*. Additionally, we analyzed the genetic relationships between 14 landraces and wild individuals using molecular markers (*trnL* and *ITS*). The identified genetic variants or haplotypes and genetic pools mainly corresponded with the species. In the case of "Ayoteco", incongruence between markers was observed. Low selection intensity, genetic flow events, and the plasticity of morphological traits may explain the high number of landraces without clear differences in their morphological diversity (vegetative characters) or genetic pools. The use of reproductive traits and massive sequencing might be useful for identifying possible morphological and genetic changes in the *Agave* landraces used for pulque production.

Keywords: plant domestication, landrace, morphology, molecular markers, *Agave salmiana*, *Agave mapisaga*

INTRODUCTION

The *Agave* L. genus has great economic, cultural, and ecological importance in Mexico and beyond. For over 10,000 years, agaves have provided many products, such as food, fiber, construction materials, and beverages (Gentry, 1982; Colunga-GarcíaMarín et al., 2007). From an ecological perspective, agaves are the keystone species in their native environments, providing food, refuge, and water to various organisms as well as preventing soil erosion (Gentry, 1982). The genus *Agave*, commonly known as maguey, is comprised of 206 species endemic to the Americas (Gentry, 1982; García-Mendoza, 2011). It has a recent origin (7.8–10.1 mya; Good-Avila et al., 2006), with Mexico being the primary center of diversity of the genus. Nearly 75% (159) of *Agave* species are found in Mexico (Gentry, 1982; García-Mendoza, 2011). It has been suggested that *Agave* species rapidly diversified with the coinciding of increased aridization of portions of North America (Good-Avila et al., 2006) with the increased diversity of pollinator species, which favored hybridization of *Agave* species (Good-Avila et al., 2006; Flores-Abreu et al., 2019).

One of the most important products obtained from *Agave* plants is pulque, a fermented beverage of pre-Hispanic origin. However, little to no research has been carried out on the *Agave* species used for pulque production, making it difficult to develop adequate conservation, management, and production strategies. Pulque agaves (*magueyes pulqueros*) are enormous plants that can measure more than 5 m in diameter and 10 m in height, including their inflorescence. These plants are mainly adapted to the elevation and climatic conditions of central Mexico. The main species utilized in pulque production are *Agave americana* L., *Agave inaequidens* Koch, *Agave mapisaga* Trel., *Agave salmiana* Otto ex Salm-Dyck subsp. *salmiana*, and *Agave salmiana* Otto ex Salm-Dyck subsp. *tehuacanensis* (Karw. Ex Salm-Dyck) García-Mend. (García-Mendoza, 2011). One study found more than 60 common names for pulque agaves (Mora-López et al., 2011), complicating the study of their diversity.

Pulque is produced by the alcoholic fermentation of *Agave* sap, which is called aguamiel. Its consumption, along with its benefits, has been recorded in at least eight codices since pre-Hispanic times (Gonçalves de Lima, 1956). According to the Aubin Codex, the Aztecs discovered pulque in the year 7 Acatl (1187 A.D.) during their long pilgrimage from Aztlán in search of Tenochtitlan (which is now Mexico City; Gonçalves de Lima, 1956). During the pre-Hispanic period, pulque played a major role in the cultural development of the region, especially for populations living in the Mexican highlands. For example, pulque was offered as tribute to the Aztecs by those they conquered (Guerrero-Guerrero, 1985). However, on a regular basis, this beverage was supposedly only consumed by the elderly, the sick, lactating women, and high-level dignitaries. Being inebriated by pulque was considered improper and a cause for punishment (Gonçalves de Lima, 1956). Alternatively, the Florentine Codex states that pulque was sold in the so-called *tianguis*, or open-air markets, and was consumed by those who participated in religious ceremonies as well as manual laborers (Sahagún, 1979).

Such was the economic importance of pulque that its production was allowed to continue in New Spain despite the prohibition of many other native beverages, such as, for example, mezcal. At the time of independence, 30,000 barrels of pulque were being produced every year (Guerrero-Guerrero, 1985). By the beginning of the Mexican Revolution in 1910, however, agrarian reform led to a 45% reduction in production which was also a result of land redistribution, *Agave* overexploitation, competition from other alcoholic beverages (mainly beer), and an intense smear campaign (Guerrero-Guerrero, 1985; Ramírez-Rancaño, 2000). From that point, pulque was considered a filthy, unsanitary, and low-class beverage dangerous for human health (Ramírez-Rancaño, 2000). Given the decrease in consumption, *Agave* plants have gradually disappeared from the Mexican landscape, with the majority of production fields being abandoned or substituted by other crops (Ramírez-Rancaño, 2000; Alvarez-Duarte et al., 2018).

Despite these historical challenges, pulque has had a rebound in recent years due to the efforts of the Mexican government to promote it through fairs, its consumption among young people and also due to the existence of economic resources to promote the cultivation of agaves used to produce it (Ramírez, 2014). Various studies have shown that pulque is, in fact, safe for consumption and nutritious, with high levels of carbohydrates, mineral salts, vitamins, and amino acids (Escalante et al., 2008; Lappe-Oliveras et al., 2008; Ortiz-Basurto et al., 2008). It also contains prebiotic and probiotic microorganisms that help improve intestinal microbiota (Moreno-Vilet et al., 2014). Many other products are obtained from pulque agaves including: inulin-type fructans; xylitol sweetener (Alvarado et al., 2014); agave syrup; distilled liquors, such as mezcal; leaves for cooking barbacoa (Gentry, 1982; Colunga-GarcíaMarín et al., 2007); *mixiotes*, or leaf cuticles for wrapping and cooking meat; the edible white worm (*Aegiale hesperians* Walker) and red worm (*Hipota agavis* Blásquez); and maguey mushrooms (*Pleurotus opuntiae* Durieu et Levy; Gentry, 1982; Alvarez-Duarte et al., 2018).

Given the context provided above, the present study seeks to contribute to the understanding of the morphological and genetic diversity of traditional pulque *Agave* species and landraces in Tlaxcala, the second-most important state in Mexico in terms of pulque production (Álvarez-Duarte et al., 2018; Ramírez-Manzano, 2020), through an analysis of morphological characters and chloroplast and nuclear DNA markers. The resulting data are fundamental for the conservation of the landrace diversity of pulque *Agave* species and for improving their production in central Mexico.

MATERIALS AND METHODS

Study Area

The state of Tlaxcala is located in the central eastern region of Mexico between 19° 06' 18'' and 19° 43' 44'' N and 97° 37' 31'' and 98° 42' 30'' W, with an elevation range of 2200 to 4,400 m above sea level. It comprises 0.2% of the area of Mexico

(399,113.7 ha) and has a population of 1,272,847. Its climate can be characterized as semi-arid and cool temperate, with an annual precipitation of 720 mm (INEGI, 2016). The main soil type is cambisol. Agriculture is one of the main economic activities and corresponds with 74% of the state's area; the main crops are corn (*Zea mays* subsp. *mays* L.), beans (*Phaseolus vulgaris* L.), barley (*Hordeum vulgare* L.), and potato (*Solanum tuberosum* L.). Nearly 41% of the land area is comprised of grasslands, and another 9% of forests, wherein the main arboreal species are *Pinus teocote* Schiede ex Schltdl., *Juniperus depeana* Steud., *Abies religiosa* (Kunth) Schltdl & Cham., and *Quercus laurina* Humb & Bonpl (INEGI, 2016).

Study System

Fourteen *Agave* landraces used for pulque production in Tlaxcala and three wild individuals (plants from wild populations that are not used for pulque production) were analyzed (**Table 1** and **Table S1**). Landraces are plants that are only grown in certain ecogeographic areas that have adapted to the environmental conditions of these areas as well as traditional management practices and uses (Casañas et al., 2017). Both landraces and wild individuals were identified at the species and subspecies level (**Table 1** and **Table S1**) following García-Mendoza's (2011) and Gentry's (1982) taxonomic keys. However, there are no taxonomic keys specifically for identifying landraces, which are generally identified by their common names (names assigned by people). These landraces are described in the present study according to our field work observations in nine municipalities, 10 localities, and 14 producers (see **Supplementary Data 1**, **Figure 1**).

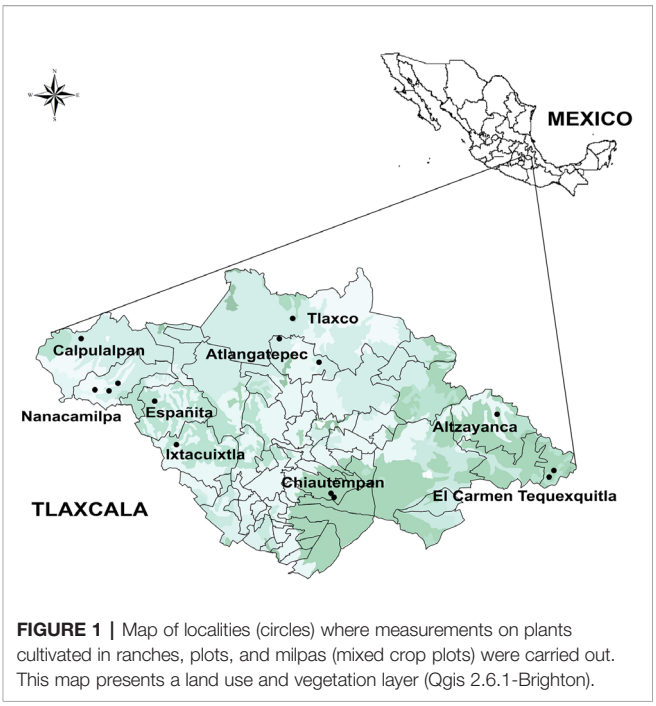


FIGURE 1 | Map of localities (circles) where measurements on plants cultivated in ranches, plots, and milpas (mixed crop plots) were carried out. This map presents a land use and vegetation layer (Qgis 2.6.1-Brighton).

Morphological Diversity

Nine *Agave* landraces were analyzed (**Figure 2**): *Agave salmiana* subsp. *salmiana* "Amarillo", "Ayoteco", "Colorado", "Chalqueño", "Chino", "Manso", "Prieto", and "Xilomelt" and *Agave mapisaga* var. *mapisaga* "Palmilla". Twenty-two populations of 13–35 individuals were included in the study (N = 478 individuals; **Table 1**; https://zenodo.org/record/3976297#.Xy35gChKjIU). The number of landraces and analyzed individuals depended on the availability of the number of mature *Agave* plants close to initiating inflorescence emergence.

Twenty morphological variables were measured (**Figure 3**) in each individual plant. Of these variables, spine shape and tooth

TABLE 1 | Morphological diversity of nine *Agave* landraces analyzed, species, places of origin and number of individuals by population.

Species	"Landrace"	Municipality	Community	Individuals
A. salmiana subsp. *salmiana*	"Amarillo"	Chiautempan	San Pedro Tlalcuapan	14
A. salmiana subsp. *salmiana*	"Ayoteco"	Nanacamilpa	Nanacamilpa	27
A. salmiana subsp. *salmiana*	"Ayoteco"	Nanacamilpa	Nanacamilpa	25
A. salmiana subsp. *salmiana*	"Colorado"	Atlangatepec	Villa Alta	10
A. salmiana subsp. *salmiana*	"Colorado"	Nanacamilpa	Nanacamilpa	10
A. salmiana subsp. *salmiana*	"Chalqueño"	Nanacamilpa	Nanacamilpa	24
A. salmiana subsp. *salmiana*	"Chino"	Ixtlacuixtla	Alpotzonga	25
A. salmiana subsp. *salmiana*	"Manso"	Altzayanca	Altzayanca	26
A. salmiana subsp. *salmiana*	"Manso"	Atlangatepec	Villa Alta	21
A. salmiana subsp. *salmiana*	"Manso"	Calpulalpan	Xultepec	25
A. salmiana subsp. *salmiana*	"Manso"	Chiautempan	San Pedro Tlacoapan	25
A. salmiana subsp. *salmiana*	"Manso"	El Carmen Tequexquitla	El Carmen Tequexquitla	18
A. salmiana subsp. *salmiana*	"Manso"	El Carmen Tequexquitla	El Carmen Tequexquitla	17
A. salmiana subsp. *salmiana*	"Manso"	Ixtlacuixtla	Alpotzonga	25
A. salmiana subsp. *salmiana*	"Manso"	Nanacamilpa	Nanacamilpa	34
A. salmiana subsp. *salmiana*	"Manso"	Nanacamilpa	Nanacamilpa	35
A. salmiana subsp. *salmiana*	"Manso"	Nanacamilpa	Nanacamilpa	19
A. salmiana subsp. *salmiana*	"Manso"	Tlaxco	Hacienda Xochuca	25
A. salmiana subsp. *salmiana*	"Prieto"	Chiautempan	San Pedro Tlalcuapan	25
A. salmiana subsp. *salmiana*	"Xilomelt"	Españita	Álvaro Obregón	13
Agave mapisaga var. *mapisaga*	"Palmilla"	Atlangatepec	San Pedro Ecatepec	13
Agave mapisaga var. *mapisaga*	"Palmilla"	Españita	Álvaro Obregón	22

FIGURE 2 | Agaves used in the production of pulque. **(A–J)** *Agave salmiana* subsp. *salmiana*: **(A)** "Amarillo," **(B)** "Ayoteco," **(C)** "Colorado," **(D)** "Chalqueño," **(E)** "Chino," **(F)** "Manso," **(G)** "Matecón," **(H)** "Prieto," **(I)** "Púa Larga," **(J)** "Xilomelt." **(K)** *Agave salmiana* subs *tehuacanensis* Tepezorra and **(L)** *Agave mapisaga* var. *mapisaga* "Palmilla."

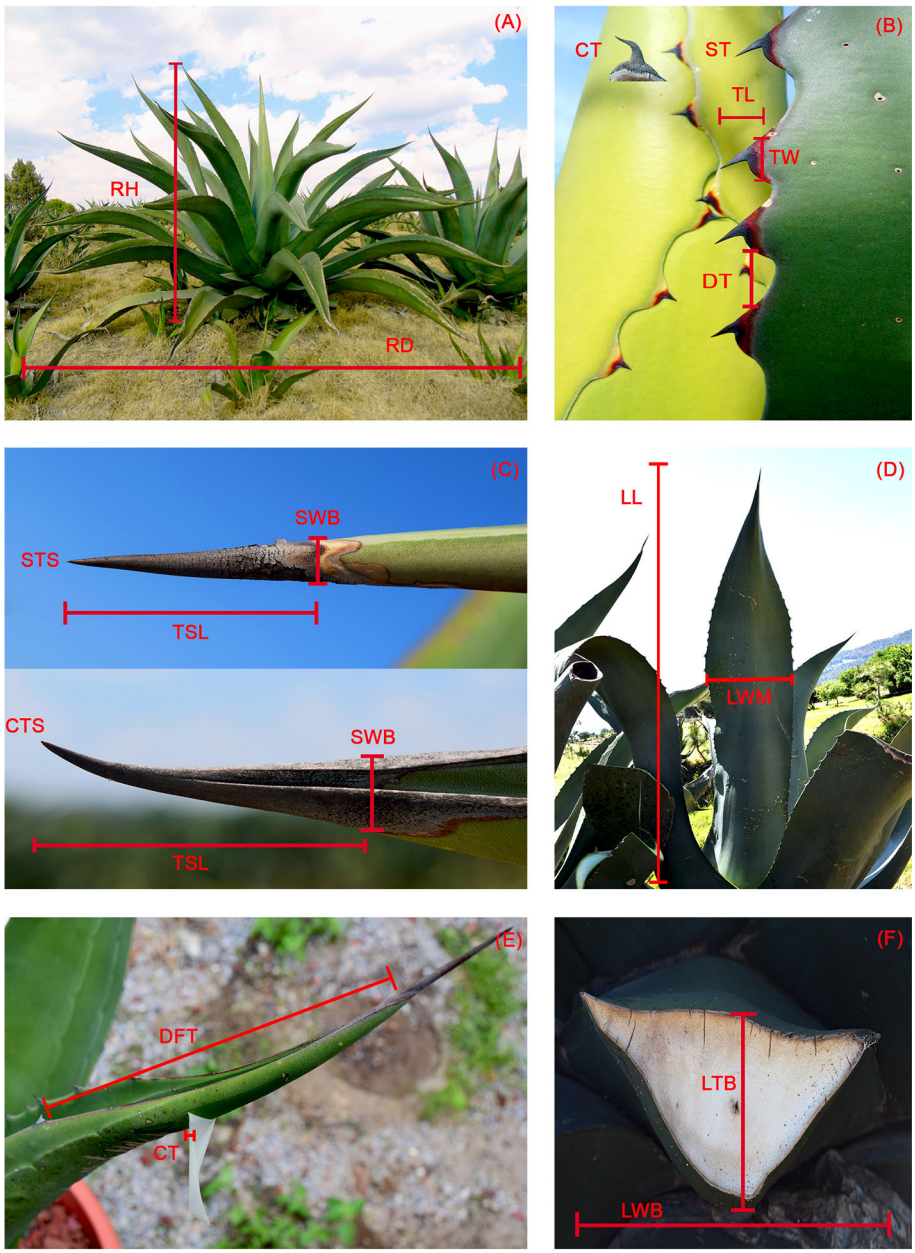


FIGURE 3 | Morphological characters. **(A)** RH, rosette height; RD, rosette diameter. **(B)** CT, curved teeth; ST, straight teeth; TL, tooth length; TW, tooth width; DT, distance between teeth. **(C)** STS, straight terminal spine; CTS, curved terminal spine; SWB, terminal spine, width at base; TSL, terminal spine length. **(D)** LL, leaf length; LWM, leaf, width at mid length. **(E)** DFT, distance to first tooth; CT, cuticle thickness. **(F)** LTB, leaf thickness at base; LWB, Leaf width at base.

shape were binary (0, 1), and total number of leaves, number of rigid leaves, and number of flaccid leaves were counts. The remaining variables were intervals except for the percentage of rigid leaves, which was treated as an ordinal variable with five levels since most leaves were rigid (i.e., 77% of observations had entirely rigid leaves and 82% of the observations had at least 95% rigid leaves). It was calculated from three variables (total number of leaves, number of rigid leaves, number of flaccid leaves). The levels were as follows: 1) less than 60% rigid leaves, 2) 60 to 70% rigid leaves, 3) 70 to 80% rigid leaves, 4) 80 to 90% rigid leaves, and 5) more than 90% rigid leaves.

Pearson linear correlations were calculated between interval variables to detect redundant information and, in this way, describe the morphology of the plants with a smaller number of variables. The correlations are shown on the heat map in **Figure 4**. Variables with correlations above $R = 0.11$ appear collapsed in the same group, all of them providing redundant information. Based on correlations and a PCA (Gabriel, 1971;

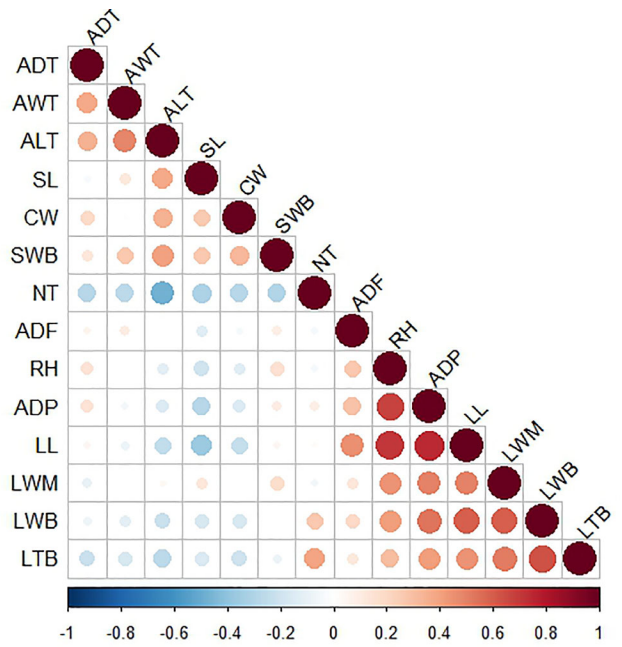


FIGURE 4 | Heat map of linear correlations between interval variables. Intense red corresponds with a highly positive correlation and intense blue with a highly negative correlation. Size of the circles indicate the magnitude of the correlation coefficient: larger circles indicate greater correlation, being the maximum size (in the diagonal) equivalent to a coefficient of 1. RH, Rosette height (cm); ADP, Average diameter of plant (cm); TNL, Total number of leaves; NRL,Number of rigid leaves; NFL, Number of flaccid leaves; CW, Cuticle width (mm), LL, Leaf length (cm); LWB, Leaf width at base (cm); LWN, Leaf width at middle (cm); LTB, Leaf thickness at base (cm); SL, Spine length (mm); SWB, Spine width at base (mm); SS, Spine shape; ADF, Average distance of the first tooth (cm); NT, Number of teeth; ADT, Average distance between teeth (mm); AWT, Average width of teeth (mm); ALT, Average length of teeth (mm); TS, Tooth shape; PRL, Percentage of rigid leaves.

Gabriel, 1980), we selected three variables (average distance of the first tooth, leaf length, and leaf thickness at the base) from the first group in addition to three variables (average length of teeth, cuticle width, and spine length) from the second group (see **Figure 4** and **Figure S1**). We also included two binary variables (spine shape and tooth shape) and the ordinal variable (percentage of rigid leaves). Consequently, we were able to describe the morphology of the plants with just nine variables (i.e., the morphology of each individual plant is represented by a 9-dimensional vector) (Nguyen and Holmes, 2019).

In order to generate clusters of plants based on morphological data, we used a hierarchical agglomerative clustering algorithm (Rokach and Maimon, 2005) along with the Gower's metric (Gower, 1971), which is suitable for calculating distances between mixed data (i.e., individuals described using different types of variables, such as binary, ordinal, and interval; Kaufman and Rousseeuw, 2005). We calculated these distances between our mixed data using the daisy function in the cluster package (Maechler et al., 2018) of R software (R Core Team, 2019). Once the dissimilarity matrix was obtained, we used the hierarchical clustering algorithm in the agnes function of the aforementioned package. For the linkage criterion, we used the Ward's method (Ward, 1963).

The visualization of the clustering structure is useful for deciding the number of clusters. For this, we used a dendrogram, which is a graphical tool that shows the distance at which observations merged. We also used the agglomerative coefficient, which describes the strength of the obtained clustering structure, ranging from zero to one. Values close to one suggest a strong clustering structure, and values close to zero a weak structure (Kaufman and Rousseeuw, 2005). Both the dendrogram and agglomerative coefficient can be obtained *via* the plot fuction applied to a proper agnes object.

For a visual impression of the detected clusters, we used an Andrews plot. High-dimensional data can be represented defining a finite Fourier series, so that each multivariate vector is represented by a curve in a two-dimensional space. Due to the periodicity of Fourier series, the *x*-axis is the interval from $-\pi$ to π (roughly from −3.14 to 3.14), while the *y*-axis is the value of the specific finite Fourier series (defined in each situation) when evaluated over the interval from $-\pi$ to π (Andrews, 1972). If there is structure in the data, it should be visible in the Andrews curves.

The curve algorithm was implemented using the Andrews function in the Andrews package in R software (Myslivec, 2012).

Genetic Relationships

Twenty individuals of *A. salmiana* subsp. *salmiana*, *A. salmiana* subsp. *tehuacanensis*, and *A. mapisaga* var. *mapisaga* were included in the relationship analysis. Seventeen belonged to the landraces, and three were wild types (*A. salmiana* subsp. *salmiana* Manso Listado, *A. salmiana* subsp. *salmiana* Prieto Silvestre, and *A. salmiana* subsp. *tehuacanensis* Tepezorra; **Table S1**). Tissue from the tip of a leaf was collected from each individual and kept in silica gel. DNA was extracted from each sample using liquid nitrogen and the DNeasy Plant Mini Kit extraction kit (Qiagen, Hilden, Germany). The quantity and quality of the obtained DNA were measured using a Colibri microvolume spectrometer (Titertek-Berthold, Pforzheim, Germany).

Ten primers previously used to screen for interspecific variation in *Agave* were tested (Good-Avila et al., 2006; Scheinvar et al., 2017; Flores-Abreu et al., 2019). Only two of them presented variation (**Table S2**): the $trnL^{(UAA)}$ intron (c–d) of cpDNA (Taberlet et al., 1991) and the internal transcribed spacers (*ITS*) 1–4 (Bayer et al., 1996). The $trnL^{(UAA)}$ primer was amplified by polymerase chain reaction (PCR) under the following protocol: 95°C/7 min and 35 cycles of 95°C/1 min, 52.4°C/60 s, and 72°C/90 s followed by one cycle of 72°C/7 min and one cycle of 4°C/10 min. The nuclear ribosomal DNA *ITS* 1–4 were sequenced under the following protocol: 94°C/2 min followed by 30 cycles of 94°C/1 min, 57°C/90 s, 72°C/2 min; an elongation period of 72°C/10 min; and one cycle of 4°C/10 min. The PCR was performed for both pairs of primers, each reaction had a final volume of 25 μL and contained: 2 μL of DNA (10–100 ng), 1 μL of 10 μM for each pair of primer, and 21 μL (1X) of GoTaq Hot Start Colorless Master Mix (Promega, Madison, USA). The PCR products were run in an electrophoresis chamber in agarose gel and purified with QIAquick PCR purification kits (Qiagen). The sequences were obtained in the

Biodiversity and Health Genomic Sequencing Laboratory of the Institute of Biology, Universidad Nacional Autónoma de México (UNAM). To prepare the samples, we used 0.4 μL of big dye terminator v. 3.1, 2 μL of 5x buffer, 4 μL of water, 1 μL of primer (10 μM), and 3 μL of the purified amplificate. The samples were placed in the PCR, and the following program was used: 30 cycles at 96°C/10 s, 50°C/5 s, and 60°C/4 min. Once the cycle series ended, the samples were purified with Centri-Sep (Thermo Fisher Scientific) following the manufacturer's specifications. To the purified samples, 25 μL of EDTA (0.5 mM) were added, and the samples on the plate were run in an Applied Biosystems 3730xL sequencer (ThermoFisher Scientific). The generated sequences were uploaded to GenBank (see **Table S1**).

Sequences were cleaned and aligned by eye with the support of BioEdit 7.2.6.1. Haplotype networks per marker and per combined matrix were obtained, and a parsimony analysis was performed using the TCS 1.21 program (Clement et al., 2000). The gaps were considered as missing data, and the network was constructed at a 95% confidence level.

For each marker and combined matrix, a Bayesian clustering method was applied with the support of the STRUCTURE v. 2.3.4 program. For these analyses, an independent allelic frequencies linkage model was used (Pritchard et al., 2000; Falush et al., 2003). A total of 350,000 Markov Chain Monte Carlo (MCMC) repetitions were performed, with 150,000 burn-in periods for each run. Ten runs were designed to estimate from 1 to the 5 populations. The optimum number (K) of genetic clusters was obtained by calculating the ΔK statistic using an Evanno test in the STRUCTURE HARVESTER program (Evanno et al., 2005; Earl and vonHoldt, 2012). A bar plot was made to represent the individual assignment probability obtained for K in the R software.

RESULTS AND DISCUSSION

Morphological Diversity

In the different analyses carried out under the unsupervised learning method, we mainly observed two clusters related with each species. All of the landraces of *A. salmiana* subsp. *salmiana*, with the exception of "Ayoteco", were mainly observed in one of the clusters, and the other cluster was mainly composed of *A. mapisaga* var. *mapisaga* and *A. salmiana* subsp. *salmiana* "Ayoteco". This confirms that the analysis of only the vegetative characters can distinguish the *Agave* species but not landraces of commonly used for pulque production. However, the landrace "Ayoteco" recognized in classical taxonomy as *A. salmiana* subsp. *salmiana* is closer to *A. mapisaga* var. *mapisaga* based on morphometric evidence, as provided below.

The dendrogram (**Figure 5**) is overall divided into two large morphological zones: a first zone indicated by cluster 1 and a second zone indicated by merger of cluster 2 and cluster 3. However, a further split of the merged cluster 2 and 3 is reasonable as well. Clusters 2 and 3 consisted of landraces of *A. salmiana* subsp. *salmiana*, with the exception of "Ayoteco", and cluster 1 contains mainly *A. mapisaga* var. *mapisaga* and *A. salmiana* subsp. *salmiana* "Ayoteco". **Table 2** shows how the individuals of the landraces are distributed in the three clusters as well as the total observations in each. To formalize the distribution of the individual plants according to landrace and cluster, we performed a cluster analysis using the same agglomerative method and the Ward's method with the row-wise vectors in **Table 2B** (**Figure 6**).

Table 3 shows the means of the interval variables per cluster. The main morphological contrasts occurred between individuals in cluster 1 and the merged cluster 2 and cluster 3 (**Table 3**). The

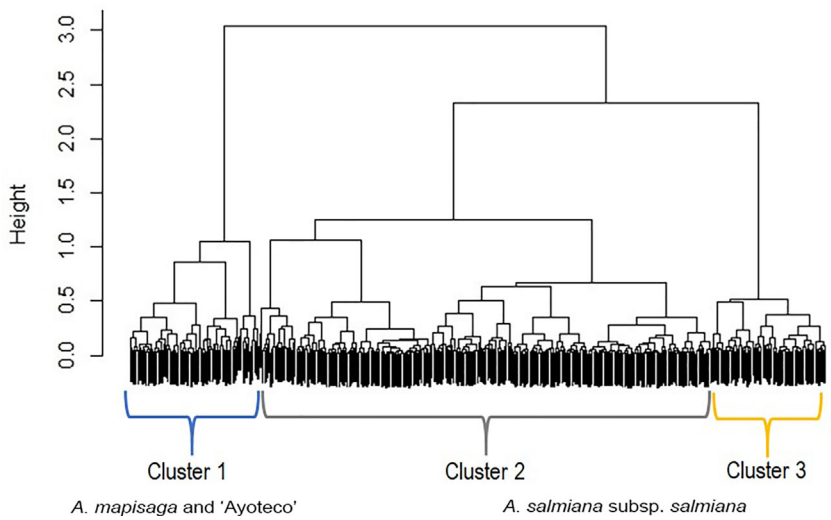


FIGURE 5 | Dendrogram showing the fusions of individual observations when clustering according to the "agnes" function of the R cluster library using Gower's dissimilarities and Ward's method. Agglomerative coefficient: 0.9855. The black area on the bottom represents the unclustered data, while the black vertical lines that protrude into the white area show the heights (dimensionless distance) at which the clusters were formed.

TABLE 2 | (A) Distribution of the different individual plants according to landrace and cluster, in absolute numbers. **(B)** Proportions of individual plants distributed (row wise) according to landrace and cluster.

Landrace	(A) Absolutes				(B) Proportions			
	Cluster 1	Cluster 2	Cluster 3	Total	Cluster 1	Cluster 2	Cluster 3	Total
A. mapisaga var. *mapisaga* "Amarillo"	0	**14**	0	14	0.00	**1.00**	0.00	1.00
A. salmiana subsp. *salmiana* "Ayoteco"	**29**	23	0	52	**0.56**	0.44	0.00	1.00
A. salmiana subsp. *salmiana* "Colorado"	2	**10**	8	20	0.10	**0.50**	0.40	1.00
A. mapisaga var. *mapisaga* "Chalqueño"	7	**16**	1	24	0.29	**0.67**	0.04	1.00
A. salmiana subsp. *salmiana* "Chino"	3	**18**	4	25	0.12	**0.72**	0.16	1.00
A. salmiana subsp. *salmiana* "Manso"	25	**188**	57	270	0.09	**0.70**	0.21	1.00
A. salmiana subsp. *salmiana* "Prieto"	3	**13**	9	25	0.12	**0.52**	0.36	1.00
A. salmiana subsp. *salmiana* "Xilomelt"	0	**13**	0	13	0.00	**1.00**	0.00	1.00
A. mapisaga var. *mapisaga* "Palmilla"	**20**	15	0	35	**0.57**	0.43	0	1.00
Total	89	310	79	478				

The highest values for each landrace were shown in bold.

landraces of cluster 1 (mainly *A. mapisaga* var. *mapisaga* and *A. salmiana* subsp. *salmiana* "Ayoteco") had, on average, shorter average teeth length, shorter spine length, and thinner cuticles; they also have a longer distance to the first tooth, greater leaf thickness at the base, and longer leaves. On the other hand, the means of the individuals in cluster 2 and 3 (mainly landraces of *A. salmiana* subsp. *salmiana*, with the exception of "Ayoteco") showed some differences yet tended to be more similar.

Individuals in cluster 1 also differed from those in cluster 2 and 3 in terms of spine and tooth shape (**Table 4**). Cluster 1 (mainly *A. mapisaga* var. *mapisaga* and *A. salmiana* subsp. *salmiana* "Ayoteco") is dominated by individuals with curved spines and straight teeth. Cluster 2 (mainly *A. salmiana* subsp. *salmiana* "Amarillo", "Chalqueño", "Chino", "Manso", and "Xilomelt") is basically formed by individuals with straight spines and straight teeth. Meanwhile, in cluster 3 (mainly *A. salmiana* subsp. *salmiana* "Colorado" and "Prieto"), all individuals have straight spines and curved teeth. Morphological differences between clusters are also evident in the percentage of rigid leaves. Ninety-five percent of individuals in cluster 3 had at least 90% rigid leaves, but only 66% of individuals in cluster 1 had rigid leaves. Meanwhile, cluster 2 is between clusters 1 and 3 in terms of percentage of rigid leaves (see **Table 5**). Therefore, it was possible to morphologically separate species of pulque agaves.

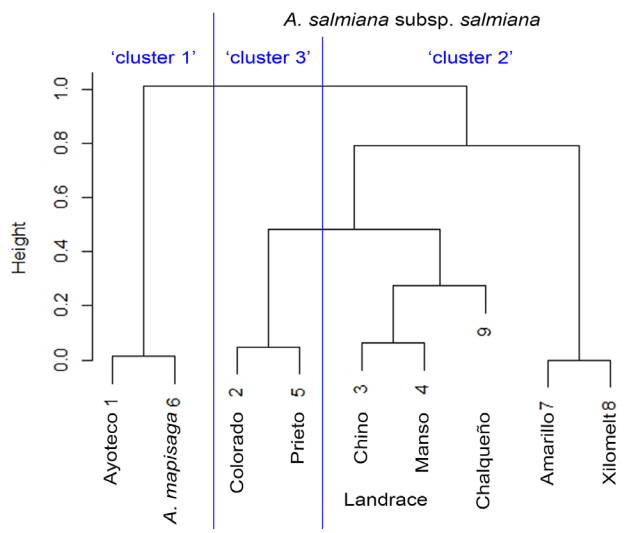


FIGURE 6 | Similarity dendrogram between varieties. The percentage distributions in **Table 2** were introduced as vectors in the "agnes" function of the R cluster library using Ward's method.

Figure 7 shows the actual clusters in two dimensions. It is important to remember that the morphological vectors of the individual plants are 9-dimensional, so we cannot visualize them precisely. To have a visual impression of both the data and clusters, it is necessary to use visualization techniques such as the Andrews plot, whose main, and perhaps only, purpose is to visualize high dimensional data. This plot allowed us to confirm that the clusters are not only geometrically but visually differentiable and, as already anticipated in the dendrogram, clusters 2 and 3 are more similar, as can be seen in the greater similarity of the oscillatory patterns of the observations of these clusters.

The majority of the studies on the morphological diversity of agaves for pulque production are based solely on the analysis of vegetative characters (Alfaro-Rojas et al., 2007; Mora-López et al., 2011; Figueredo et al., 2014; Alvarez-Ríos et al., 2020; **Table S3**). Similar to the present study, the variability of morphological vegetative characters in these latter studies is distributed and grouped at the species level with respect to classical taxonomy (Mora-López et al., 2011; Alvarez-Ríos et al., 2020; **Table S4**). Mora-López et al. (2011) observed that *A. salmiana* and *A. mapisaga* var. *mapisaga* are separated into two groups, similar to the findings herein. Alfaro-Rojas et al. (2007) analyzed six stem characters and found that "Manso", "Ayoteco", and "Verde" formed a single cluster that differed from "Xilomelt". Notably, these latter authors also found that "Ayoteco" belonged to the morphological group of *A. salmiana* in contrast with our findings that it is closest to *A. mapisaga* var. *mapisaga*. This reflects the difficulty of identifying plants by only their common names, as it is possible that the "Ayoteco" studied by Alfaro-Rojas et al. (2007) is not the same that we analyzed

TABLE 3 | Averages of the morphological interval variables per cluster.

	Cuticle width *(mm)*	Leaf length *(cm)*	Leaf thickness at base *(cm)*	Spine length *(mm)*	Average distance of the first tooth *(cm)*	Average length of teeth *(mm)*
Cluster 1						
Mean	0.09	226.4	20.4	49.1	24.4	4.9
Standard deviation	0.02	38.6	3.9	13.7	7.2	1.8
Cluster 2						
Mean	0.11	196.6	18.7	61.0	20.8	6.0
Standard deviation	0.02	36.5	4.4	11.9	6.5	2.2
Cluster 3						
Mean	0.10	184.1	16.9	66.4	20.7	7.3
Standard deviation	0.01	33.4	5.0	11.9	6.1	2.7

TABLE 4 | Percentage of individual plants with straight spine, curve spine, straight tooth and curve tooth per cluster.

Cluster	Straight spine	Curve spine	Straight tooth	Curve tooth
Cluster 1	0	100	84	12
Cluster 2	100	0	99	1
Cluster 3	100	0	0	100

TABLE 5 | Percentage distribution of the ordinal variable (percentage of rigid leaves) per cluster.

Percentage of rigid leaves per cluster	*1 (≤60%)*	*2 (>60% and ≤70%)*	*3 (>70% and ≤80%)*	*4 (>80% and ≤90%)*	*5 (>90%)*
Cluster 1	6	3	15	10	66
Cluster 2	1	3	3	6	87
Cluster 3	0	0	0	5	95

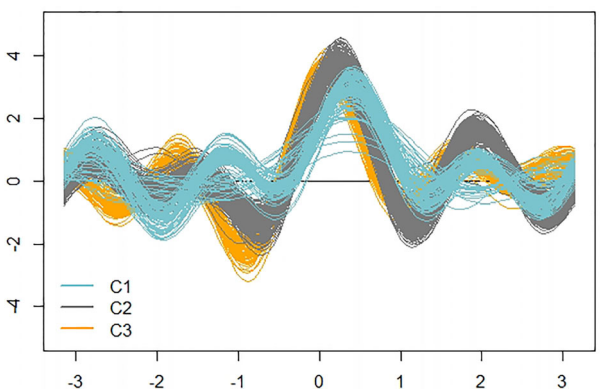


FIGURE 7 | Andrews curves of the 478 observations distinguished according to cluster. Blue: cluster 1; Gray: cluster 2; Orange: cluster 3.

herein or that this cultivar presents variability in vegetative characters throughout its distribution.

The analysis of reproductive, biochemical, and genetic characters could improve the taxonomic resolution of the different landraces. Figueredo et al. (2014, 2017) analyzed vegetative, reproductive, and genetic characters, distinguishing wild populations of *A. inaequides* from cultivated ones as well as species such as *A. hookeri* and *A. cupreata*. Few studies have analyzed reproductive characters due to the difficulty of obtaining flowers, fruits, and seeds from pulque agaves. Growers halt their development to obtain aguamiel or, if flowers bloom, people collect them as a food source (**Tables S3** and **S4**).

Comparison between previous studies and ours is not an easy task because of differences between studies. For example, each study measured morphological characters in different ways, and the most significant characters also tended to be different in each study (**Tables S3** and **S4**). The number of measured plants has varied from 2 to 100; however, in most of cases, just a few plants were measured. In addition, many authors did not identify species and landraces using classical taxonomic keys. Rather, the majority of studies only reported the common name of the taxa, which complicates comparisons given that these names may or may not correspond with the same taxa across studies. These issues could be resolved by standardizing the methods for measuring morphological, biochemical, and genetic characters. In addition, studies should be rigorous in their taxonomic identification efforts and aspire to create taxonomic keys for identifying landraces in the future in order to avoid the ambiguity caused by using only common names.

Some authors have also pointed out that the management and artificial selection of certain plants could have modified certain characters, reflecting the characteristics desired by growers (Vargas-Ponce et al., 2007; Mora-López et al., 2011). Growers tend to select for larger plants (longer leaves and thicker leaf bases), smaller teeth, greater distance between teeth, greater distance from the apical spine to the first tooth, and shorter apical spines. Cluster 1 ("Ayoteco" and *A. mapisaga* var. *mapisaga*) could be expressing these characters of interest. Notably, we also observed that the subspecies *salmiana*, in particular the "Manso" landrace, is still cultivated in a very traditional manner in Tlaxcala. On the other hand, *A. mapisaga* and "Ayoteco" could be the landraces most modified by growers. However, these differences could also be due to pre-existing differences between species resulting from their

evolutionary histories. The shape of the spine and teeth could be explained by the latter, as these do not appear to be characters of interest but rather distinctive characters of the landraces (**Table 4**). To better understand the changes caused by artificial selection, it would be necessary to make comparisons with wild relatives.

The analysis of morphological diversity has many practical applications for the management, production, and conservation of the *Agave* landraces used for pulque production. Through these studies, the variability of morphological characters in production fields can be described, possibly enabling the identification of different landraces as morphological groups with particular characters, which has taxonomic implications. The characterization of this diversity is fundamental for the establishment of *in situ* and *ex situ* conservation strategies. On the other hand, the lack or loss of diversity can lead to an increase in the vulnerability of plants to pests and diseases and decrease their resistance to climate change.

Genetic Relationships

The low resolution obtained from the *Agave* genome fragments is most likely due to the recent origin of the genus, around 6–10 million years ago, and the high hybridization rates occurring as a consequence of pollinator-mediated genetic flow (Good-Avila et al., 2006; Scheinvar et al., 2017; Flores-Abreu et al., 2019). In our first attempt to use genetic markers to distinguish pulque-producing agaves, we only found enough variation to differentiate between species using one plastid marker and one nuclear marker.

A 13 base-pair (bp) inversion was found in the *trnL*$^{(UAA)}$ intron of cpDNA at the 1,209 position in the total matrix for *A. mapisaga* and *A. salmiana* subsp. *salmiana* "Ayoteco". Therefore, in building our *trnL* haplotype network (**Figure 8A**), *A. mapisaga* and "Ayoteco" were grouped as haplotype H1. Using the *ITS* 1–4 marker, we only found a single base change (C–A) at position 602 in a 635-bp-long matrix, and polymorphisms were observed in *A. mapisaga*. Thus, the *ITS*

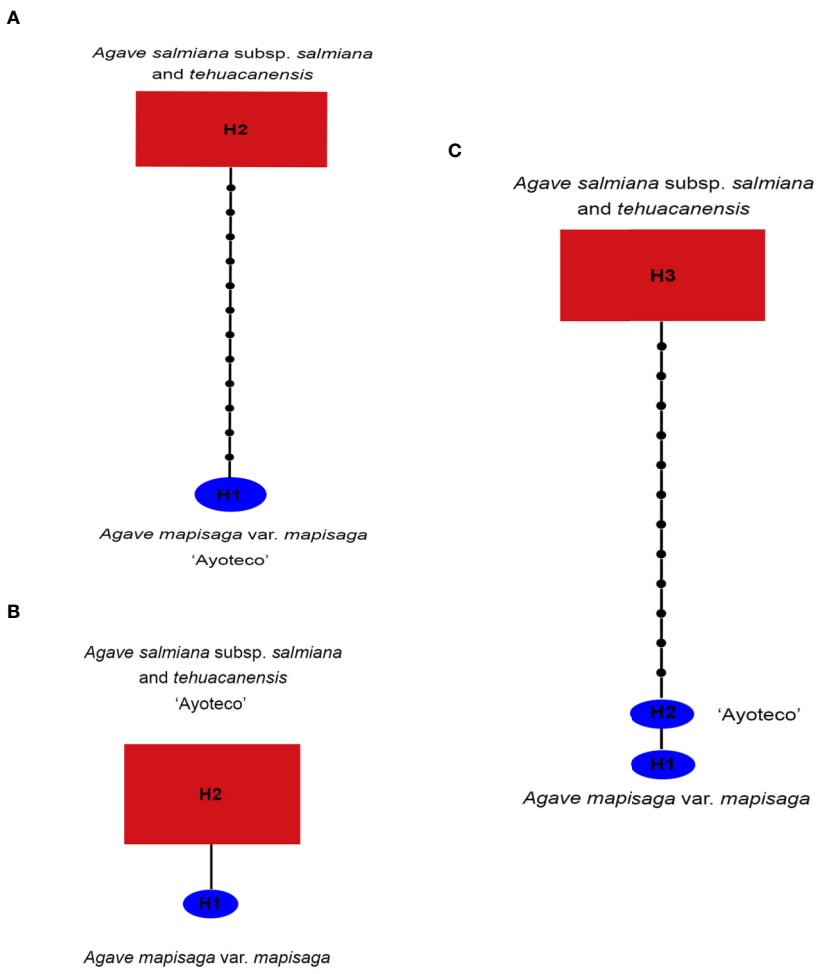


FIGURE 8 | Haplotype network from the **(A)** *trnL*, **(B)** *ITS*, and **(C)** combined *trnL* and *ITS* matrices obtained in the TCS program (Clement et al., 2000). The size of the red rectangles and blue ellipses represent the haplotype frequency (**Table S1**).

haplotype network (**Figure 8B**) indicated a haplotype per species separated by a single mutation, grouping "Ayoteco" with *A. salmiana*. Finally, the combined matrix showed three haplotypes (**Figure 8C**): *A. mapisaga* (H1) separated by only one mutation from "Ayoteco" (H2) and the remaining *A. salmiana* taxa (H3). This latter haplotype is separated from the other two by various steps.

In the STRUCTURE assignment analysis, we obtained a ΔK probability value supporting two clusters in *trnL* intron and combined matrices (**Figures 9C–F**). Each cluster or genetic pool corresponded with a species. Specifically, cluster 1 corresponded with *A. salmiana* subsp. *salmiana* and *A. salmiana* subsp. *tehuacanensis* and cluster 2 with *A. mapisaga* var. *mapisaga*. Only *A. salmiana* subsp. *salmian* "Ayoteco" was the exception and was shown to share the same genetic pool as *A. mapisaga* var. *mapisaga* (cluster 2). In the case of the *ITS* marker, only one cluster was identified with a K = 1 (**Figure 9A**). For this last marker, there was no ΔK obtained because there was no variance

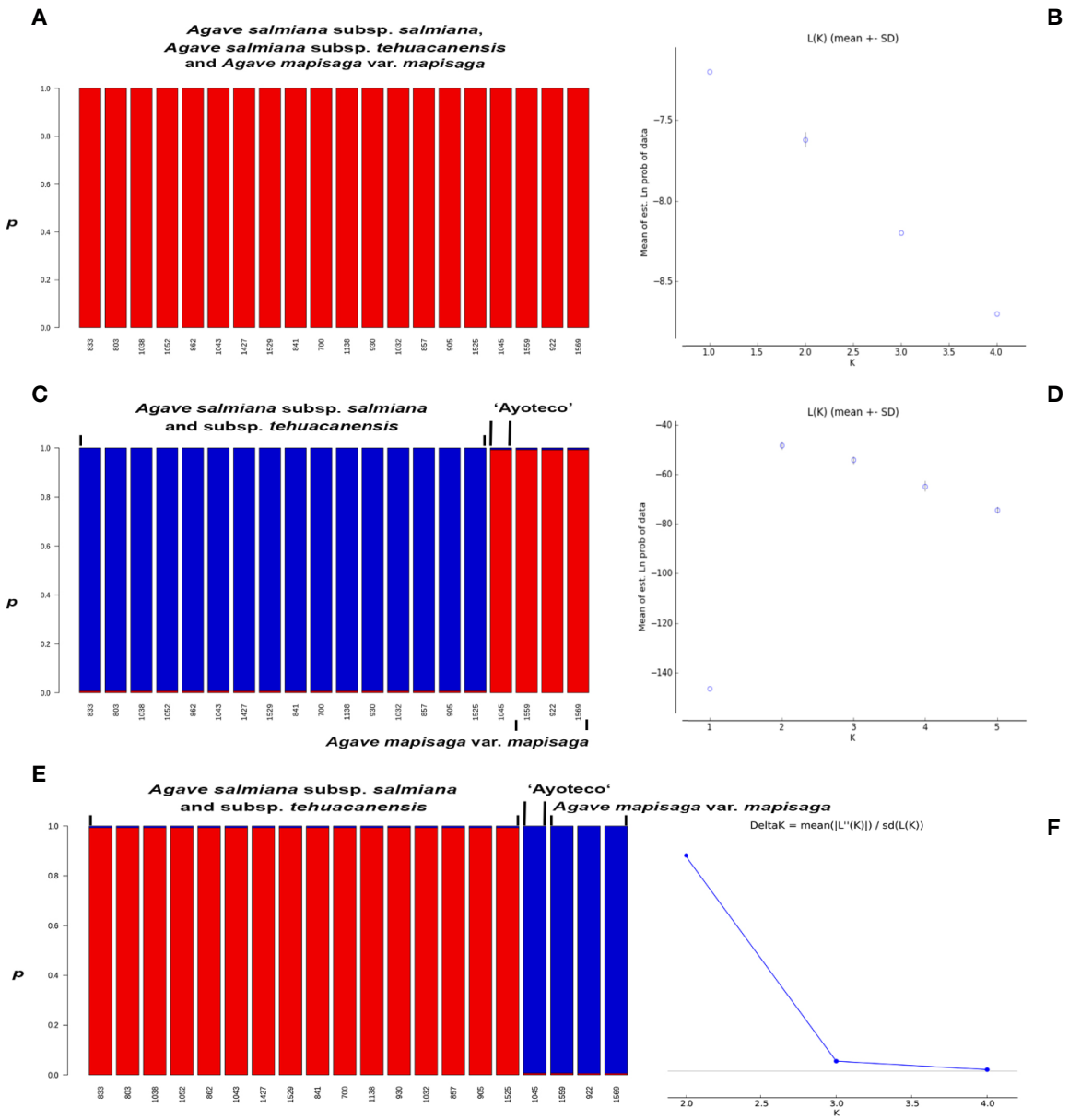


FIGURE 9 | Individuals' assignment analysis using the Bayesian statistic with the *p* co-ancestry coefficient in STRUCTURE and probability values for **(A, B)** *ITS* (K=1), **(C)** *trnL* (K=2), and **(D, E)** combined matrix (K=2); with the optimum genetic cluster number (K) based on Evanno's ΔK statistic calculated in the STRUCTURE HARVESTER program **(F)**, except for *ITS*. The Evanno's ΔK and probability values graphs for *trnL* are quite similar to the ones from the combined matrix, these were not presented.

found between the different runs, and the highest probability value reached 1 (**Figure 9B**).

As previously mentioned, classical taxonomy based on morphological traits identifies "Ayoteco" as *A. salmiana* subsp. *salmiana*, even though this landrace shares traits with both species. However, the morphological analysis grouped "Ayoteco" and *A. mapisaga* var. *mapisaga*, even though they are not identical. Therefore, with respect to "Ayoteco," the genetic data are not congruent, although the combined data analysis supports its close relationship with *A. mapisaga* var. *mapisaga*.

The history retrieved from each genome (either plastid or nuclear) tends to be different because of its origin, evolutionary processes, and manner of inheritance. Chloroplasts are haploid with maternal inheritance, almost nonexistent recombination, and a lower mutation rate compared to the nuclear genome, but higher than nuclear ribosomal DNA (Palmer et al., 1988; Wheeler et al., 2014; Ferradini et al., 2017). Many plant studies have easily found sufficient phylogenetic resolution from plastid genome markers (Yang et al., 2013; Wheeler et al., 2014).

There is a much discussion over whether the results of combined analyses are accurate since two different evolutionary histories are combined. However, these histories are not entirely independent because they have the same species history (Gadagkar et al., 2005; Degnan and Rosenberg, 2009). Some studies have found well-supported trees with high resolution from analyzing total combined data (Gadagkar et al., 2005; Degnan and Rosenberg, 2009). In the plastid genome, inversions are more rare than single nucleotide substitutions (Palmer et al., 1988); thus, in this case, the use of the *trnL* intron could be more plausible than the *ITS* marker. Ultimately, these approaches represent hypotheses that can be re-addressed in future studies to search for further phylogenetic resolution for the taxa in question, as the genetic data presented herein are not conclusive. It is necessary to explore a larger number of *Agave* species (paying special attention to the *Salmianae* group, Gentry, 1982) and a larger set of molecular markers to find ones that would be capable of providing a more complete and resolved phylogenetic history in order to better understand the genetic diversity of pulque-producing *Agave* species and landraces.

Currently, most studies on the *Agave* genus have been carried out at the population level using DNA fingerprint markers. Using microsatellite molecular markers (SSRs), Figueredo et al. (2015) found that *A. inaequides* (used for mezcal and pulque production) shows high genetic diversity and low structuring and that genetic diversity was mainly associated with the environment or the samples' region of origin. This could imply that there is variability among populations of landraces or throughout their geographic distribution, which highlights the need to carry out further studies on their morphological and genetic diversity throughout their geographic distribution. Later, once again using microsatellite molecular markers (SSR), Figueredo et al. (2017) observed that the species used for mezcal and pulque production (*Agave inaequides*, *Agave hookeri*, and *Agave cupreata*) tended to form genetic pools corresponding with the species level, even though *A. hookeri* is mixed with cultivated *A. inaequides* plants. In the present study, *A. salmiana* and *A. mapisaga* could similarly represent separate genetic pools, with the exception of "Ayoteco" (**Figure 9**). Alvarez-Ríos et al. (2020) also found through SSR markers that *A. salmiana* and *A. mapisaga* have distinguishable genetic pools by species, with moderate genetic flow between these species. A cluster showing similarities to *Agave americana* was also found, possibly a hybrid of *A. salmiana* and *A. americana*. Microsatellites are especially useful for the study of interpopulational variation. In particular, chloroplast microsatellites (cpSSRs) are useful for testing hypothesis on the hybridization between species or populations (Wheeler et al., 2014). However, cpSSRs have been little used in studies on agave and could be one alternative for distinguishing the genetic pools of *A. salmiana* and *A. mapisaga* and possible introgression events.

Because of their low selection intensity, the landraces studied herein may be assigned to genetic clusters at the species level in contrast to extreme examples like *Agave tequilana* "Azul", which is intensively and continuously cloned to produce tequila. *A. tequilana* "Azul" is recognized as having an identity distinct from other possibly closely related species (Trejo et al., 2018). With respect to pulque agaves, there is a possibility that morphological vegetative traits are plastic and reflect growers' knowledge of different landraces and the common names given to different forms. Accordingly, such plasticity may be environmental and directed by the selection of plants or, occasionally, may result from genetic flow between them when plants eventually flower following abandonment or growers forget to harvest the aguamiel.

CONCLUSIONS

The perception of the diversity of pulque agaves is high among growers, as reflected by the high number of common names. However, most morphological studies have only analyzed vegetative characters and have only been able to differentiate landraces at the species level. The sequencing analyses carried out herein enabled genetic variants generally associated with genetic pools at the species level to be identified. However, the landraces from the pulque production agaves could be different from each other, since they have origin in different wild plants and have different uses. Furthermore, crop abandonment allows for genetic flow events. In contrast, the intensive production of certain landraces through restrictive asexual reproduction or the complete abandonment of certain landraces could lead to their eventual disappearance.

The identity of *A. salmiana* subsp. *salmiana* "Ayoteco" is still not clear, this might be a hybrid between *A. salmiana* and *A. mapisaga* or we just might be observing gene flow between "Ayoteco" and *A. mapisaga*. In order to acquire further understanding about pulque landraces' taxonomic identity, it is necessary to include more characters in the analysis, including reproductive and biochemical characters, as well as polymorphisms in a single nucleotide (SNP). It is also recommendable to perform

population analyses of *Agave* landraces throughout their geographic distribution and compare these to their wild relatives.

In Tlaxcala, the conservation of *Agave* landraces with a local distribution can be strengthened through their taxonomic identification and continued local use. However, the conservation of their vast richness also depends on avoiding their replacement with other landraces that could lead to their extinction, such as in the Yucatan Peninsula, where local landraces (henequen) have been displaced, or in Jalisco, where traditional varieties once used for tequila production have been substituted by the blue agave.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the GenBank (see **Table S2**).

AUTHOR CONTRIBUTIONS

LT participated in collecting, taking morphological characters measurements, obtaining sequences, analyzing kin relationships, writing and editing the article. MR and DC-T did the statistical analysis of morphological characters, and participated in writing and editing the article. ER-G took part in collecting, taking morphological characters measurements and making some of the figures. LM-C also took part in collecting and measuring morphological characters.

ACKNOWLEDGMENTS

We thank CONACyT Support for the Strengthening and Development of the Scientific and Technological Infrastructure for grant 26,866, and "Catedras CONACyT" project No. 59. We thank SEFOA for their support in the maintenance of the work area. We are grateful to the Biology Institute, UNAM, for its financial support to scientific publishing. The authors thank Eribel Bello Cervantes, Estefania Briones Dumas, Yulia Hernández Hernández, Gerardo D. Amezquita Hernández, Nayely López Rodríguez, Mariana E. Sánchez Martínez, and Adriana Montoya for their support with the field work. We also acknowledge the producers and growers who granted us access to their crop fields. Finally, we thank Laura Márquez and Nelly López of LaNABio for their assistance in obtaining the sequences and Jaime Gasca for his support during analysis stage and suggestions for the writing of the article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.524812/full#supplementary-material

REFERENCES

Alfaro-Rojas, R. G., Legaria, J. P., and Rodríguez, J. E. (2007). Diversidad genética en poblaciones de agaves pulqueros (*Agave* spp.) del nororiente del Estado de México. *Rev. Fitotec. Mex.* 30, 1–12.

Alvarado, C., Camacho, R. M., Cejas, R., and Rodríguez, J. A. (2014). Profiling comercial *Agave* fructooligosaccharides using ultrafiltration and high performance thin layer cromatography. *Rev. Mex. Ing. Quim.* 13, 417–427.

Alvarez-Duarte, M. C., Garcia-Moya, E., Suárez-Espinosa, J., and Luna-Cavazos, M. (2018). Traditional knowledge, cultivation and use of maguey pulquero in municipalities of Puebla and Tlaxcala. *Polibotanica* 45, 205–222. doi: 10.18387/polibotanica.45.15

Alvarez-Ríos, G., Pacheco-Torres, F., Figueredo-Urbina, C. J., and Casas, A. (2020). Mnagement, morphological and genetic diversity of domesticated agaves in Michoacán, México. *J. Ethno. Biomed.* 16:3. doi: 10.1186/s13002-020-0353-9

Andrews, D. F. (1972). Plots of high-dimensional data. *Biometrics* 28, 125–136. doi: 10.2307/2528964

Bayer, R. J., Soltis, D. E., and Soltis, P. S. (1996). Phylogenetic inferences in *Antennaria* (Asteraceae: Gnaphalieae: Cassiniinae) based on sequences from Nuclear Ribosamal DNA internal transcribed spacers (ITS). *Am. J. Bot.* 83, 516–527. doi: 10.2307/2446220

Casañas, F., Simo, J., Casals, J., and Prohens, J. (2007). Toward an Evolved Concept of Landrace. *Front. Plant Sci.* 8, 145. doi: 10.3389/fpls.2017.00145

Clement, M., Posada, D., and Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9, 1657–1659. doi: 10.1046/j.1365-294x.2000.01020.x

Colunga-GarcíaMarín, P., Zizumbo-Villarreal, D., and Martinez-Torres, J. (2007). "Tradiciones en el aprovechamiento de los agaves mexicanos: una aportación a la protección legal y conservación de su diversidad biológica y cultural," in *En lo ancestral hay futuro: del tequila, los mezcales y otros agaves*. Eds. P. Colunga-GarciaMarin, A. Larqué Saavedra, L. E. Eguiarte and D. Zizumbo-Villarreal (Merida, Yuc: CICY, A.C), 229–248.

Degnan, J. H., and Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 24, 332–340. doi: 10.1016/j.tree.2009.01.009

Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Escalante, A., Giles-Gómez, M., Hernández, G., Córdova-Aguilar, M. S., López-Munguía, A., Gosset, G., et al. (2008). Analysis of bacterial community during the fermentation of pulque, a traditional Mexican alcoholic beverage, using a polyphasic approach. *Int. J. Food Micribiol.* 124, 126–134. doi: 10.1016/j.ijfoodmicro.2008.03.003

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587. doi: 10.1111/j.1471-8286.2007.01758.x

Ferradini, N., Lancioni, H., Torricelli, R., Russi, L., Ragione, I. D., Cardinali, I., et al. (2017). Characterization and phylogenetic analysis of ancient italian landraces of pear. *Front. Plant Sci.* 8:751:751. doi: 10.3389/fpls.2017.00751

Figueredo, C. J., Casas, A., Colunga-GarcíaMarín, P., Nassar, J. M., and González-Rodríguez, A. (2014). Morphological variation, management and domestication of 'maguey alto' (*Agave inaequidens*) and 'maguey manso' (*A. hookeri*) in Michoacán, México. *J. Ethno. Biomed.* 10, 66. doi: 10.1186/1746-4269-10-66

Figueredo, C. J., Casas, A., González-Rodríguez, A., Nassar, J. M., Colunga-GarcíaMarín, P., and Rocha-Ramírez, V. (2015). Genetic structure of coexisting wild and managed agave populations: implications for the evolution of plants under domestication. *AoB Plants* 7:plv114. doi: 10.1093/aobplan/plv114

Figueredo, C. J., Casas, A., and Torres-García, I. (2017). Morphological and genetic divergence between *Agave inaequidens*, *A. cupreata* and the domesticated *A. hookeri*. Analysis of their evolutionary relationships. *PloS One* 12, 11. doi: 10.1371/journal.pone.0187260

Flores-Abreu, I. N., Trejo-Salazar, R. E., Sánchez-Reyes, L. L., Good, S. V., Magallón, S., García-Mendoza, A., et al. (2019). Tempo and mode in coevolution of *Agave sensu* lato (Agavoideae, Asparagaceae) and its bat pollinators, Glossophaginae (Phyllostomidae). *Mol. Phylogenet. Evol.* 133, 176–188. doi: 10.1016/j.ympev.2019.01.004

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58, 453–467. doi: 10.1093/biomet/58.3.453

Gabriel, K. R. (1980). "Biplot," in *Encyclopedia of Statistical Sciences*. Eds. N. L. Johnson and S. Kotz (New York: John Wiley and Sons), 147–173.

Gadagkar, S. R., Rosenberg, M. S., and Kumar, S. (2005). Inferring species phylogenies from multiple genes: Concatenated sequence tree versis consensus gene tree. *J. Exp. Zool. B. Mol. Dev. Evol.* 304B, 64–74. doi: 10.1002/jez.b.21026

García-Mendoza, A. J. (2011). "Agavaceae," in *Flora del Valle de Tehuacán-Cuitlatán. Fascículo 88* (Mexico: Instituto de Biología, UNAM).

Gentry, H. S. (1982). *Agaves of Continental North America* (Tucson: The University of Arizona Press).

Gonçalves de Lima, O. (1956). *El Maguey y el pulque en los códices Mexicanos* (Mexico: Fondo de Cultura Económica).

Good-Avila, S. V., Souza, V., Gaut, B. S., and Eguiarte, L. E. (2006). Timing and rate of speciation in *Agave* (Agavaceae). *PNAS* 103, 9124–9129. doi: 10.1073/pnas.0603312103

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics* 27, 857–871. doi: 10.2307/2528823

Guerrero-Guerrero, R. (1985). *El pulque* (Mexico: INAH/Joaquín Mortiz).

Instituto Nacional de Estadística and Geografia e informática (2016). *Anuario estadístico y geográfico de Tlaxcala* (Mexico: INEGI). Available at: http://evaluacion.tlaxcala.gob.mx/images/stories/documentos/planea/estadistica/ae/ea2016.pdf (Accessed 12 de june de 2019).

Kaufman, L., and Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis* (New York: Wiley).

Lappe-Oliveras, P., Moreno-Terrazas, R., Arrizón-Gaviño, J., Herrera-Suárez, T., García-Mendoza, A., and Gschaedler-Mathis, A. (2008). Yeasts associated with the production of Mexican alcoholic nondistilled and distilled *Agave* beverages. *FEMS Yeast. Res.* 8, 1–16. doi: 10.1111/j.1567-1364.2008.00430.x

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2018). *cluster: Cluster Analysis Basics and Extensions*, R package version 2.0.7-1.

Mora-López, J., Reyes-Agüero, J., Flores-Flores, J., Peña-Valdivia, C., and Aguirre-Rivera, J. (2011). Variación morfológica y humanización de la Sección Salmianae del género *Agave*. *Agrociencia* 45, 465–477.

Moreno-Vilet, L., Garcia-Hernandez, M. H., Delgado-Portales, R. E., Corral-Fernandez, N. E., Cortez-Espinosa, N., Ruiz-Cabrera, M. A., et al. (2014). *In vitro* assessment of agave fructans (*Agave salmiana*) as prebiotics and immune system activators. *Int. J. Biol. Macromol.* 63, 181–187. doi: 10.1016/j.ijbiomac.2013.10.039

Myslivec, J. (2012). *andrews: Andrews curves. R package version 1.0*. Available at: https://CRAN.R-project.org/package=andrews.

Nguyen, L. H., and Holmes, S. (2019). Ten quick tips for effective dimensionality reduction. *PloS Comput. Biol.* 15 (6), e1006907. doi: 10.1371/journal.pcbi.1006907

Ortiz-Basurto, R. I., Pourcelly, G., Doco, T., Williams, P., Dornier, M., and Belleville, M. P. (2008). Analysis of the main components of the aguamiel produced by the maguey-pulquero (*Agave mapisaga*) throughout the harvest period. *J. Agric. Food Chem.* 56, 3682–3687. doi: 10.1021/jf072767h

Palmer, J. D., Jansen, R. K., Michaels, H. J., Chase, M. W., and Manhart, J. R. (1988). Chloroplast DNA variation and plant phylogeny. *Ann. Mo. Bot. Gard.* 75, 1180–1206. doi: 10.2307/2399279

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

R Core Team (2019). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.R-project.org/.

Ramírez, A. M. (2014). "El pulque, un bien biocultural y sustentable," in *El maguey y el pulque en la región central de México*. Ed. Y. Ramos-Galicia (Mexico: Gobierno del Estado de Tlaxcala), 83–98.

Ramírez-Manzano, S. I., Bye, R., García-Moya, E., and Romero-Manzanares, A. (2020). Aprovechamiento del maguey pulquero en Nanacamilpa, Tlaxcala, Mexico. *Rev. Etnb.* 18, 65–76.

Ramírez-Rancaño, M. (2000). *Ignacio Torres Adalid y la industria pulquera* (Mexico: Instituto de investigaciones Sociales de la UNAM. Editores Plaza y Valdés).

Rokach, L., and Maimon, O. (2005). *Clustering Methods. in Data Mining and Knowledge Discovery*. Eds. O. Maimon and L. Rokach (Boston, M.A: Springer). Handbook.

Sahagún, B. (1979). *Historia general de las cosas de Nueva España. Manuscrito de la Colección Palatina de la Biblioteca Medica Laurenziana, Florencia [reproducción facsimilar: Códice Florentino]* (México, D.F: Archivo General de la Nación).

Scheinvar, E., Gámez, N., Castellanos-Morales, G., Aguirre-Plater, E., and Eguiarte, L. E. (2017). Neogene and Pleistocene history of *Agave lechuguilla* in the Chihuahuan Desert. *J. Biogeography* 44, 322–334. doi: 10.1111/jbi.12851

Taberlet, P., Gielly, L., Pautou, G., and Bouvet, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Mol. Eco.* 17, 1105–1109. doi: 10.1007/BF00037152

Trejo, L., Limones, V., Peña, G., Scheinvar, E., Vargas-Ponce, O., Zizumbo-Villarreal, D., et al. (2018). Genetic variation and relationships among agaves to the production of Tequila and Mezcal in Jalisco. *Indust. Crops. Prod.* 125, 140–149. doi: 10.1016/j.indcrop.2018.08.072

Vargas-Ponce, O., Zizumbo-Villarreal, D., Martínez-Castillo, J., Coello-Coello, J., and Colunga-GarcíaMarín, P. (2007). *In situ* diversity and maintenance of Traditional *Agave* Landraces used in spirits production in West-Central Mexico. *Eco. Bot.* 61, 362–375. doi: 10.1663/0013-0001(2007)61[362:ISDAMO]2.0.CO;2

Ward, J. H.Jr (1963). Hierarchical grouping to optimize an objective function. *J. Amer. Statist. Assoc.* 58, 236–244. doi: 10.1080/01621459.1963.10500845

Wheeler, G. L., Dorman, H. E., Buchanan, A., Challagundla, L., and Wallace, L. E. (2014). A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. *Appl. Plant Sci.* 2:1400059. doi: 10.3732/apps.1400059

Yang, J. B., Tang, M., Li, H. T., Zhang, Z. R., and Li, D. Z. (2013). Complete chloroplast genome of the genus *Cymdidium*: lights into the species identification phylogenetic implications and population genetic analyses. *BMC Evol. Biol.* 13:84. doi: 10.1186/1471-2148-13-84

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Agronomy

REVIEW
published: 24 September 2020
doi: 10.3389/fagro.2020.00010

Pre-Columbian Rock Mulching as a Strategy for Modern *Agave* Cultivation in Arid Marginal Lands

Hector Ortiz-Cano[1], Jose Antonio Hernandez-Herrera[2], Neil C. Hansen[1], Steven L. Petersen[1], Michael T. Searcy[3], Ricardo Mata-Gonzalez[4], Teodoro Cervantes-Mendívil[5], Antonio Villanueva-Morales[6], Pil Man Park[7] and J. Ryan Stewart[1]*

[1] Department of Plant and Wildlife Sciences, Brigham Young University, Provo, UT, United States, [2] Departamento de Recursos Naturales, Universidad Autónoma Agraria Antonio Narro, Saltillo Coahuila, Mexico, [3] Department of Anthropology, Brigham Young University, Provo, UT, United States, [4] Department of Animal and Rangeland Sciences, College of Agricultural Sciences, Oregon State University, Corvallis, OR, United States, [5] Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), Campo Experimental Costa de Hermosillo, Sonora, Mexico, [6] División de Ciencias Forestales, Departamento de Estadística, Matemática y Cómputo, Universidad Autónoma Chapingo, Texcoco, Mexico, [7] Floriculture Research Division, National Institute of Horticultural and Herbal Sciences, Rural Development Administration, Wanju-gun, South Korea

OPEN ACCESS

Edited by:
Karolina Heyduk, University of Hawaii, United States

Reviewed by:
Daniel Sacristán Moraga, University of the Balearic Islands, Spain
Jessé Rodrigo Fink, Instituto Federal do Paraná Câmpus Palmas, Brazil

***Correspondence:**
Hector Ortiz-Cano
hectoroc@gmail.com

Specialty section:
This article was submitted to Plant-Soil Interactions, a section of the journal Frontiers in Agronomy

Received: *14 March 2020*
Accepted: *10 August 2020*
Published: *24 September 2020*

Citation:
Ortiz-Cano H, Hernandez-Herrera JA, Hansen NC, Petersen SL, Searcy MT, Mata-Gonzalez R, Cervantes-Mendívil T, Villanueva-Morales A, Park PM and Stewart JR (2020) Pre-Columbian Rock Mulching as a Strategy for Modern Agave Cultivation in Arid Marginal Lands. Front. Agron. 2:10. doi: 10.3389/fagro.2020.00010

Cultivation of C_3 and C_4 crops in semi-arid regions will be severely constrained as global temperatures rise. Consequently, alternative crops need to be sought out that adapt well to heat and drought and are productive despite limited access to water. Traits, such as crassulacean acid metabolism (CAM), enable economically important species such as those in the *Agave* genus adapt to drought and high temperatures. The succulence and high efficiency of agaves, which enables them to produce biomass with little water, underscores their feasibility as an alternative crop for semi-arid regions, such as the Sonoran Desert in the southwestern U.S. In this paper, we offer a review of the suitability for cultivation of agaves *via* dryland farming, particularly by rock mulching techniques used by pre-Columbian, Sonoran Desert farmers. This analysis dovetails with information also provided on the biological traits of *Agave* and its historical and present utilization. Pre-Columbian, Hohokam dryland farmers used rock mulching in the form of rock piles to cultivate agaves. Rock piles acted as a type of mulch to harvest rainfall and to retain soil moisture, which allowed the Hohokam to intensively cultivate agaves during multi-year droughts. Remains of Hohokam rock mulching for agave production can be found at archaeological sites in central Arizona, which provides evidence of the utility of dryland farming and ancient agricultural innovation to reconcile water scarcity in the region. Moreover, the use of rock piles likely bolstered *Agave* productivity in marginal lands. Although little is known of historic rock mulching to cultivate agaves and its biological implications on plant productivity we suggest its application as a dryland farming model could be a sustainable strategy in the U.S. Southwest.

Keywords: crassulacean, succulent, dryland farming, rock mulching, rock piles

INTRODUCTION

Increasingly limited access to water and elevating temperatures will continue to hamper productivity of conventional C_3 and C_4 crops in arid and semi-arid regions throughout the world in the coming decades (Porter and Semenov, 2005; De Micco and Aronne, 2012; Zandalinas et al., 2018). Contemporary agricultural challenges, limited water availability, and rising temperatures have constrained agriculture in dry regions throughout history (Ingram, 2010). Within the last 1,300 years, global warming has been associated with overexploitation of natural resources, increased urbanization, and accelerated agricultural development (Woodhouse et al., 2010). Models of changes in global temperature suggest that induced anthropogenic global warming became more frequent during the Medieval Warm Period, which occurred between 700 and 1300 CE (Galloway, 1986; Hughes and Diaz, 1994; Bradley et al., 2003; Stinchcomb et al., 2011). The Medieval Warm Period increased temperatures and reduced water levels of lakes and rivers throughout Europe, Asia, and North and South America (Van West and Dean, 2000; Chu et al., 2002; Sridhar et al., 2006; Helama et al., 2009; Woodhouse et al., 2010). In addition to the Medieval Warm Period, some have hypothesized that global temperatures rose and rainfall patterns progressively changed in arid regions during the pre-industrial period 800–1850 A.D. due to land-use changes for agriculture (e.g., conversion of forests and grasslands into cropland) (Galloway, 1986; Reick et al., 2010; Pongratz and Caldeira, 2012).

After the Industrial Revolution, the use of fossil fuels and concomitant increases in CO_2 emissions accelerated climate change (Callendar, 1938; Revelle and Suess, 1957; Neftel et al., 1985; Lemonnier and Ainsworth, 2018), increasing global temperatures and the occurrence of droughts during the twentieth century (Hansen et al., 1981; Solomon et al., 2010; Smith et al., 2019). At the end of the last century and beginning of the twenty-first century, globalized industrial development and creation of large urban centers resulted in cropland expansion in arid and semi-arid regions to meet increased food-production demands (Krausmann et al., 2013; Laurance et al., 2014). Conversely, relatively warmer temperatures in the early part of the twenty-first century, high evapotranspiration rates, erratic rainfall, and increasingly severe droughts have deleterious consequences to farmland by reducing crop yields, resulting in an increase of marginal lands (i.e., farmland and wildlands with limited access to irrigation water and depleted soil nutrients) (Schlaepfer et al., 2017). These increasingly warm and dry conditions in regions with limited resources suggest future edaphic, biological, and climatic constraints for cultivation of C_3 and C_4 crops. Such conditions increase the need for seeking, selecting, and cultivating drought-tolerant crops, such as those found in the succulent *Agave* genus, which cope with drought through nocturnal CO_2 fixation and CAM photosynthesis (Borland et al., 2009, 2015; Stewart, 2015).

Current challenges in dry regions to cultivate and produce food in hot and water-limited conditions bear similarity to those that native people faced long ago during severe droughts in what is now the U.S. Southwest (Ingram, 2010). Irrigation water has always been a naturally limited resource in arid regions (Troyo-Diéguez et al., 1990). For dry regions, there is a need for sustainable agricultural strategies to optimize crop yields and irrigation water (Troyo-Diéguez et al., 1990). To cope with scarce availability of water, innovative, indigenous dry-farming strategies were developed anciently to produce food during droughts (Lightfoot, 1996). These dryland farmers irrigated with rainfall runoff by optimizing rainwater catchment and rewetting the landscape using manmade stone features, such as rock terraces and rock mulch (Wilken, 1972; Lightfoot, 1994, 1996). The indelible signature left by the historic use of rock terraces and rock mulching can be seen in ancient and modern societies in dry regions throughout the world. For example, in the Negev Desert of the Middle East, the nomadic Nabateans, who settled in the region around 600-300 BCE, built and used rock terraces, check dams, and rock mulching to irrigate and catch rainfall water (Stager, 1976; Evenari et al., 1982; Lightfoot, 1994; Ashkenazi et al., 2012). At the apex of Nabatean civilization, such terraces became the main dry-farming technology to cultivate olives (*Olea europaea*), pomegranates (*Punica granathum*), and apples (*Malus domestica*) (Ynnilä, 2007). Similar examples can also be found in ancient civilizations throughout the deserts of Africa, Europe, and Asia (Wilken, 1972; Lightfoot, 1994; Biazin et al., 2012). In the ancient Americas, rock-farming techniques were used in a variety of cultures and time periods (Marcus, 2006; Kennett, 2012). In the Andean region, from the times of the Huarpa civilization to that of the Incan (200 BCE to 1400 CE), rock-wall terraces were heavily relied on to cultivate potatoes (*Solanum tuberosum*), quinoa (*Chenopodium quinoa*), and corn (*Zea mays*) (Denevan, 2003; Chapagain and Raizada, 2017). Mayans in southern Mesoamerica were very effective in cultivating corn using rock terraces (Turner, 1976; Fischbeck, 2001; Webb et al., 2004). In central and northern Mesoamerica, Aztecs cultivated marginal lands with corn and agaves in a system called milpas (Evans, 1990; Zizumbo-Villarreal et al., 2012; Trombold, 2017). Ancient Pacific Islanders used rock mulching to harvest rainwater and to cultivate perennial crops, such as taro (*Colocasia esculenta*), in land with limited access to water (Stevenson et al., 1999; Wozniak, 1999; Ladefoged et al., 2013). Pre-Columbian Hohokam people, which inhabited the deserts of the American Southwest, also cultivated drought-tolerant agaves in marginal lands using rock-mulching (Fish et al., 1985; Fish and Fish, 1990, 1992; Gasser and Kwiatkowski, 1991).

Among historic dry-farming examples in the U.S., Hohokam agricultural dryland systems in central and southern Arizona are key to understanding applications of dry farming for other arid regions affected by drought. The Hohokam mastered desert farming (Fish and Fish, 1992). Their dry-farming techniques were adapted and designed to produce food in extended droughts and in the harsh Sonoran Desert climate. They implemented rock-mulching to catch rainfall water and successfully cultivate agaves to feed thousands of desert dwellers during water scarcity periods (Fish and Fish, 1990). Rock mulching turned into the primary strategy to shore up food production during droughts. *Agave* was the main crop that allowed for unabated cultural, social, and economic development in the region (Fish, 2000). As

in the prehistoric past, modern central and southern Arizona is a region constrained by the harsh Sonoran Desert climate. Here the applications of indigenous dry-farming agriculture, principally Hohokam rock mulching, opens the possibility for cultivating agaves in current and future droughts. It is our intent to portray *Agave* as a drought-tolerant crop, which can be cultivated through the application of rock mulching to harvest rainwater as a feasible and sustainable dry-land agriculture system for arid regions. The purpose of this paper is to summarize literature available on (1) studies on the ecophysiology of agaves under drought conditions, (2) dry farming using rock mulching to cultivate agaves, (3) ancient and modern-day uses of agaves, and (4) the potential of rock mulching and *Agave* cultivation in future droughts.

BIOLOGICAL TRAITS OF AGAVES KEY FOR ITS CULTIVATION IN FUTURE DROUGHTS

Nearly 75% of the continental biological diversity of the *Agave* genus can be found in Mexico and 13% in U.S. deserts (Gentry, 1982; Garcia-Moya et al., 2011). The *Agave* genus evolved biological and morphological traits that enable species to adapt to erratic, hot, and drought-changing conditions of arid regions (Silva-Montellano and Eguiarte, 2003). Morphological traits of agaves, such as their shallow root systems, distinct rosette shape, and curved leaves to maximize rainfall interception, evolved to efficiently use small amounts of atmospheric and soil moisture in water-limited environments (Martorell and Ezcurra, 2007). Such limited water and heat conditions negatively affect the physiological performance of domesticated C_3 and C_4 crops (Nobel and Jordan, 1983). Crassulacean acid metabolism (CAM) photosynthesis is the main biological trait that drives productivity of these plants in hot and water-scarce conditions (Lüttge, 2004; Borland et al., 2011). Photosynthesis of agaves relies on nocturnal stomatal opening and CO_2 gas exchange as a strategy to avoid high evapotranspiration rates and leaf water loss during daylight hours (Lüttge, 2004). Nocturnal CO_2 fixation is the primary trait agaves use to survive dry climates and to adapt to warm temperatures (Borland et al., 2009). In addition, above and belowground morphological traits (North and Nobel, 1991), such as leaf succulence, rain-hair roots, and fibrous root architecture enable agaves to adjust physiological processes to available soil-moisture levels and heat in the different seasons of dry regions (De Micco and Aronne, 2012).

Agaves are monocarpic plants with a long life cycle to maturation (Nobel, 1977). Differences in plant maturation can be observed within and between species, regions, cultivation practices, and degree of domestication (e.g., domesticated agaves, hybrids of agaves, or wild agaves) (Zizumbo-Villarreal et al., 2013). Generally, cultivated agaves require a few years or up to a decade to mature to flower, and typically more than a decade to mature to flower in the wild (Cervantes et al., 2007; Núñez et al., 2008).

Aboveground morphological traits of agaves (e.g., shape, size of leaves, and succulence) enable these plants to survive and adapt to deserts by providing protection and storing water in the leaf parenchyma (Orians and Solbrig, 1977; Cervantes et al., 2007; Núñez et al., 2008). Additionally, the rosette arrangement of the curved *Agave* leaves funnel rainwater to the plant and soil during the summer monsoon season, re-wetting their rhizomes and the soil in the root zone (Gentry, 1982). Furthermore, Martorell and Ezcurra (2007) hypothesized that the rosette trait can also trap atmospheric moisture in the form of dew and fog between leaves. The thick succulent leaves of agaves function as plant water storage for periods of scarce rain and soil moisture. Even after a period of several months, when soil moisture has reached the permanent wilting point, agaves will remain physiologically functional (Nobel, 2003).

The *Agave* root system is composed of shallow roots (mean root length: 8 to 20 cm) and rhizomes (Arizaga and Ezcurra, 2002; Nobel, 2003; Bautista-Cruz et al., 2007). Offset growth from rhizomes and aerial bulbils act as the main asexual propagation strategy of agaves. Such offsets can extend several meters from the plant in search of soil moisture (Gibson, 1996; Nobel, 2003). Shallow roots allow rapid soil moisture absorption from the soil surface, particularly from small amounts of moisture deposited after light rain events (North and Nobel, 1991). Fibrous *Agave* root systems maximize soil water absorption, particularly in well-drained sandy soils with limited capacity to retain moisture (Cervantes et al., 2007). In addition, in very dry soils, dehydration of suberized peridermal cells of mature *Agave* roots prevents water loss and desiccation of the root vascular system (North and Nobel, 1991). Additionally, these lignified roots anchor *Agave* plants to the soil. When rainfall occurs, water pulses from rain rehydrate *Agave* roots, which promotes emergence of new root hairs, thereby increasing hydraulic conductance of *Agave* root systems (Palta and Nobel, 1989). Rainwater stimulates growth of ephemeral root hairs, which are vital for rapid water uptake and replenishing of water in *Agave* leaves (North and Nobel, 1991; Huang and Nobel, 1992).

Wild and cultivated agaves flourish in arid environments and poor soils in marginal lands (Gentry, 1972; Cervantes et al., 2007; Núñez et al., 2008). Edaphic requirements include sandy soils with good drainage, 60% gravel content, and deep water tables (Cervantes et al., 2007). Particularly sandy loam soils with low salinity contents are optimum for healthy establishment of agaves. Agaves can be found growing in rocky soils in which temperatures can reach 70°C (Gentry, 1972; Nobel, 1994). Agaves typically perform well in soils with low nutrient content. Nitrogen levels in soils range between 31 and 35 parts per million (ppm) and low P between 2.6 and 3.0 ppm for optimal growth of agaves (Cervantes et al., 2007). Ideal growing conditions for agaves can be found in low-elevation mesic areas on hillslopes. In the Sonoran Desert of northwestern Mexico and southern Arizona, optimum elevation ranges for *Agave* growth have been observed between 800 and 1200 m above sea level (Gentry, 1972; Nobel and Hartsock, 1986; Núñez et al., 2008; Parker et al., 2014; Hodgson et al., 2019). However, agaves can also be found in the coastal areas of the Sonoran Desert in Mexico (Gentry, 1972; Cervantes et al., 2007; Núñez et al., 2008). Ideal precipitation

levels for agaves vary from tropical to dry regions (Gentry, 1982; Nobel, 2003). In arid regions, such as Sonora, Mexico and Arizona, USA, agaves can survive rainless seasons for several years (Nobel, 2003). Some regions with wild populations of *Agave* receive as little as 7 mm of rain and other regions receive as much as 762 mm or more of annual precipitation (Gentry, 1972, 1982; Nobel, 1976).

CULTIVATION OF AGAVES USING ROCK-MULCHING BY THE HOHOKAM

Who Were the Hohokam?

The Hohokam were pre-Columbian dryland farmers that established a flourishing civilization in what is now central and southern Arizona between 450 and 1500 C.E (Fish and Fish, 2008). Hohokam agriculture was constrained by the hot, dry climate, and the wide expanse of marginal lands in the Sonoran Desert (Fish and Fish, 1990). Drought and water availability for agriculture acted as definitive factors that influenced innovation in the agricultural and cultural development of the Hohokam (Rice, 1998; Hunt et al., 2005). These two factors shaped Hohokam irrigation and dryland agriculture in the desert, leading to the use of extensive irrigation-canal networks and rock mulching to cultivate agaves during periods of drought with erratic rainfall (Woodbury, 1961; Fish and Fish, 2012). Such approaches can be compared with dryland-farming systems observed in other advanced, prehistoric indigenous societies of Mesoamerica and South America (Doolittle, 1995; Fish and Fish, 2008). The irrigation canals of the Hohokam were similar to highly engineered Andean and Aztec irrigation-canal systems in that they were a pivotal factor in their cultural development and were designed to efficiently irrigate large areas of farmland, which led to substantial food production (Armillas, 1948; Bennett, 1948; Mitchell, 1973).

Irrigated crops formed the basis of Hohokam civilization and its economy. Efficiently distributed irrigation water from rivers and canals allowed large settlements to develop along the main rivers in the Phoenix Basin (Doyel, 2007). However, recurrent droughts at the beginning of the second millennia CE triggered periods of unstable food production, which changed agricultural strategies and the crops they cultivated (Fish and Fish, 1990). The Hohokam shifted to using more dryland farming and reliance upon rainfall for crop irrigation, resulting in less canal irrigation in the region.

During this period of recurring droughts, agaves were adopted as a crop to compensate for yield deficits during water shortages and as a supplement to irrigated annual crops (Fish and Fish, 1992, 2008; Anderies et al., 2008). The Hohokam primarily used rock piles to cultivate agaves, which enhanced their productivity during drought periods in the Sonoran Desert (**Figure 1**; Fish et al., 1985; Fish and Fish, 2012).

FIGURE 1 | Hohokam rock pile remains at Tumamoc Hill reserve in Tucson, Arizona.

Dryland Farming Using Rock-Mulching to Cultivate Agaves

In the American Southwest, as in pre-Hispanic northern Mexico, the Hohokam implemented dryland agriculture strategies to cultivate *Agave* to ensure food security during droughts in the Tucson Basin (Fish et al., 1985; Anderies et al., 2008). A model of pre-Hispanic *Agave* cultivation by Anderies et al. (2008) suggests that *Agave* dryland farming was likely a strategy implemented by pre-Columbian groups to cope with climates with drought that reduced corn yields in the Sonoran Desert. The Hohokam adopted upland dryland farming on hillslopes using rock piles, terraces, and check dams to cope with low precipitation in the region (Fish and Fish, 1992). These structures allowed for the efficient use of rainwater to irrigate downhill floodplain crops and riparian vegetation (Fish and Fish, 1990). Rock piles and terraces were used to harvest rainfall runoff and to cultivate agaves (Fish and Fish, 1992).

Historical remains of Hohokam rock piles and evidence of *Agave* cultivation and processing in roasting pits can be found at archaeological sites outside the Tucson, Arizona area in the Tortolita Mountains; the Salt Gila Basin; the community of Marana; Tonto National Forest; San Pedro Valley; and Tumamoc Hill Reserve (**Figure 2**; Masse, 1979; Crown, 1987; Ciolek-Torrello et al., 1997; Adams and Adams, 1998). One of the most representative Hohokam rock-pile fields, which was found in Marana and characterized by Fish and Fish (1992), consisted of at least 42,000 rock piles nested with 120,000 m^2 of terraces and check dams within an area of 500 ha. They calculated that the rock pile fields could have annually produced 102,000 *Agave* plants with an average yield of 40.8 Mg ha^{-1}. However, comparing average planting density (i.e., 1,000–3,000 plants ha^{-1}) of agaves (Cervantes et al., 2007; Núñez et al., 2008) in modern plantations in the Sonoran Desert in Mexico, with the rock-pile fields found in Marana, Arizona suggests

FIGURE 2 | Archaeological site at Tumamoc Hill reserve in Tucson, Arizona.

FIGURE 3 | Agave growing in ancient rock pile at archaeological site in Casas Grandes, Chihuahua, Mexico (picture courtesy of M. Searcy, 2019).

that *Agave* productivity in Arizona was possibly higher than previously calculated by Fish et al. (1985) (i.e., 102,000 *Agave* plants in 500 hectares). For example, between 1000-3000 *Agave angustifolia* plants ha^{-1} can be annually cultivated in Sonora, Mexico in grasslands with no irrigation (Cervantes et al., 2007). Similarly, McDaniel (1985) suggested a planting density of 2,000–2,500 of *Agave americana* plants ha^{-1} cultivated in grassland in southern Arizona. If the calculations of Cervantes et al. (2007) and McDaniel (1985) regarding *Agave* cultivation area and planting density are applied to the largest Hohokam rock-pile field in Marana, Arizona, the minimum planting density would be 1,000 plants ha^{-1} for 500 ha of rock-pile fields. As such, the Hohokam potentially had the capacity to cultivate nearly 500,000 agaves, which is at least five times more plants than previously estimated by Fish et al. (1985). However, commercial modern *Agave* cultivation differs from the cultivation strategies of the Hohokam. This example is only used to highlight the productive potential of the land to cultivate agaves in the region.

Hohokam Agaves

Agave plant remains in Hohokam rock-pile fields underscore the importance of rock mulch for modern cultivation of agaves in the region (Fish et al., 1985; Fish and Fish, 1990; Fish, 2000). Though agaves were no longer cultivated prior to the arrival of Europeans to the region, agaves still can be found growing in some rock-pile fields and archaeological sites in Arizona (Hodgson and Salywon, 2013; Hodgson et al., 2019). Minnis and Plog (1976) observed a relationship between the occurrence of wild *Agave parryi* plants with proximity and distribution of agaves growing at archaeological sites in the Apache-Sitgreaves National Forest, suggesting putative historic cultivation of this *Agave* species in central Arizona. Similarly, Parker et al. (2010) observed genetic differences between putative cultigens of *A. parryi* and wild *A. parryi* plants at archaeological sites in central Arizona in the Mogollon Rim. Recently, a taxon, which was named *Agave sanpedroensis*, was discovered growing only in a rock-pile field west of Tucson, which is likely a relic of Hohokam cultivation (Hodgson et al., 2019). Living plants and dried tissue of *Agave* at rock piles and roasting pits have been found in Hohokam rock-pile fields at archaeological sites in southern and central Arizona (Fish et al., 1985; Adams and Adams, 1998; Fish, 2000; Parker et al., 2007; Fish and Fish, 2012).

Little is known about cultivation of agaves using rock piles outside of Arizona. Minnis et al. (2006) found little evidence of rock piles around the prehistoric archaeological site of Casas Grandes in Chihuahua, Mexico. However, ethnobotanists and archaeologists that visited archaeological sites near Casas Grandes in 2018 and 2019 found agaves growing in ancient rock piles and terraces, perhaps indicating ancient cultivation similar to that found in Arizona (W. Hodgson and M. Searcy, personal communication). In 2018, putative hybrids of *Agave palmeri* and *A. parryi* were found growing in a rock terrace in the Casas Grandes region in northern Mexico (W. Hodgson, personal communication). Likewise, in 2019, *Agave* hybrids, which were similar in appearance to those discovered in 2018 were growing in rock piles (M. Searcy, personal communication) (**Figure 3**). These recent findings suggest that cultivation of agaves using rock structures was likely more widespread than originally assumed. Such discoveries offer new avenues of research in the use of rock piles for prehistoric *Agave* cultivation.

AGAVE TRADITIONAL USES THROUGHOUT HISTORY, FOOD, DRINKS AND OTHER SUB-PRODUCTS

Uses in Ancient Times for Food and Fermented Drinks

Indigenous people in Mexico and the U.S. Southwest have used agaves as a source of carbohydrates and fiber for the past 10,000 years (Delgado-Lemus et al., 2014). Diversity in the *Agave* genus is largely concentrated in Mexico and the U.S. Southwest, and was cultivated mostly using dryland farming techniques, such as rock mulching using rainwater runoff (Lightfoot, 1994, 1996), widely throughout tropical and arid regions of Central and South America (Good-Avila et al., 2006). Beginning in pre-Columbian times, *Agave* was used for various purposes in what is now the U.S. Southwest, including beverages, ceremonial items, fiber-based products (e.g., clothing, footwear, containers, cordage, nets, etc.), food, medicine, and paint (Castetter et al., 1938). Over the span of several centuries, particularly during droughts, agaves were an important energy source that enriched the diets of indigenous people (Fish et al., 1985; Evans, 1990; Fish and Fish, 1990; Anderies et al., 2008).

The Hohokam were one of the few pre-Columbian indigenous groups in the U.S. Southwest that extensively used rock piles to cultivate agaves as a staple crop that ensured a reliable source of food, even during droughts (Fish et al., 1985; Dobyns, 1988; Fish and Fish, 1992). The Hohokam relied on the ability of these plants to concentrate sugars in stems and stalks through their long phenological cycle. Sugars in *Agave* are inulin-type polymers of fructose that concentrate in leaves, stems, and inflorescence stalks (Mancilla-Margalli and López, 2006; Urias-Silvas et al., 2008). However, removal of *Agave* stalks at the end of their life cycle induces sugar accumulation predominantly in the stem (Hodgson, 2001; Cervantes et al., 2007; Michel-Cuello et al., 2008). Inflorescence stalk emergence indicates plants have matured, and are ready to be harvested (Arizaga and Ezcurra, 2002).

Pre-Columbian indigenous people used roasting pits, also called earth ovens, to cook their food, but particularly to roast agaves (Walton, 1977; Fish et al., 1985; Cervantes et al., 2007; Perry and Flannery, 2007; Zizumbo-Villarreal et al., 2009). Roasting pits can reach temperatures between 150 and 200°C (Cervantes et al., 2007). Roasting *Agave* heads (or caudices) at this temperature enables thermal hydrolysis to break down carbohydrate polymers into sugar monomers, such as fructose and glucose, which are relatively easy to digest and ferment by yeast (Cervantes et al., 2007), making it possible to use agaves as a food source. Methods of cooking *Agave* heads using earth ovens share similarities (e.g., shape, diameter of 1.20–2.40 m, depth of 0.80–2.10 m) between indigenous groups across various regions, both in ancient and modern times (Cervantes et al., 2007; Towell and Lecón, 2010).

The practice of roasting agaves can be traced to its origins in pre-Columbian archaeological sites in central Arizona and northern Mexico (Fish et al., 1985). In Arizona, *Agave* roasting pits also can be found in rock pile-fields (Fish and Fish, 1992). These roasting pits attest to the ancient use of roasted agaves and their cultivation in rock piles in the region. In addition, several documents from the Spanish colonial period recorded historic uses of roasted agaves by natives. Early colonial Jesuits from Spain, in what is now northwestern Mexico, recorded that agaves were used for medicinal purposes and were roasted for food by the Opata people in the Sonoran Desert (Gutiérrez-Coronado et al., 2007; Flores and Araiza, 2012). Similarly, in what is now central Mexico in the mid-sixteenth century, colonial Spaniards documented medicinal uses of roasted *Agave* by the Aztec people in the ethnobotanical compendium Codex Florentino (Williams, 1990; Díaz et al., 1993). Uses of roasted and fermented agaves for food were also recorded in the Codex Azcatitlan and Codex Boturini (Morán, 2008).

Modern Use of *Agave* as a Food Source

As indicated above, out of all the organs of agaves, the stem head produces the most edible biomass (Nobel, 2003). Inflorescences, leaves, and stalks can also be used for food and to feed cattle (*Bos primigenius taurus*), sheep (*Ovis aries*), and goats (*Capra aegagrus hircus*) (Gentry, 1972; Pinos-Rodríguez et al., 2006, 2008, 2009; Hartung, 2016; Mellado, 2016). Gentry (1972) indicated that inflorescences of some *Agave* species are edible. Moreover, Fuentes-Rodriguez (1997) and Gentry (1972) suggest that the raw leaves remaining after clipping leaves from *Agave* stems, commonly called jimado in Spanish, can be used to feed cattle. Pinos-Rodríguez et al. (2006) found that leaves, flowering stalk, and bagasse of *Agave salmiana* can be used as food and to increase body weight of sheep. In popular Mexican cuisine throughout the country, *Agave* leaves are also used to cover goat or lamb stew while being cooked in underground roasting pits.

Alcoholic Spirits and Drinks From Agaves

Aguamiel, pulque, and mezcal constitute the main beverages produced from agaves (Stewart, 2015). In order to produce aguamiel, an emerging inflorescence is cut out of the stem head. *Agave* sap, which is rich in sugars, accumulates in the remaining basin. The sap juice is subsequently siphoned out of the basin and prepared as non-alcoholic drink known as aguamiel. Fermentation of aguamiel creates pulque (Rivas, 1991), a commonly consumed, mildly alcoholic beverage in rural areas of central and southern Mexico (Enríquez-Salazar et al., 2017).

It has been estimated that pulque made of *Agave mapisaga, A. americana, Agave atrovirens*, or *Agave salmiana*, was consumed in approximately 2000 B.C. (Escalante et al., 2016). Although distilled *Agave* spirits like Tequila or mezcals are very popular in modern times, artisanal crafting and consumption of pulque remains alive in some regions of central Mexico.

Although drinks, such as aguamiel and pulque, are still consumed in modern times, these beverages have been somewhat replaced by distilled alcoholic drinks made from distillation of fermented sap of agaves (Garay and Aurea, 2008). Distillation technologies, such as the use of copper alembic stills to distill alcohol, were adapted to produce *Agave* spirits in Mexico by the Spaniards in the 1500s (Gutiérrez-Coronado et al., 2007; Towell and Lecón, 2010). In combination with a wide array of

distillation methods, the diversity of *Agave* species in the different regions and the various cooking and fermentation methods of *Agave* heads employed by tribes across Mexico enabled a rich diversification of *Agave* spirits throughout Mexico from various species (Walton, 1977).

Tequila is the most popular *Agave* spirit crafted in Mexico, and differs from commercial mezcal in that it is made exclusively from *Agave tequilana* var. Azul, which is also known as blue agave (Colunga-GarcíaMarín and Zizumbo-Villarreal, 2006; Vargas-Ponce et al., 2007). In contrast, mezcals are made from a wide diversity of agaves across Mexico. In addition, similar to some French wines, tequila has an appellation of origin (denominación de origen), which requires that blue *Agave* plants only be grown in certain states of Mexico that are believed to enhance the quality of tequila (Bowen, 2015). Tequila is mainly produced at an industrial scale following quality-control regulations compliant with national and international standards established by the Tequila Regulatory Council (Macías, 2001).

Despite climate change and political and economic changes throughout Mexican history, the tequila industry has experienced continual growth. For example, Walton (1977) reported 1.67 million liters of tequila were produced in 1960. Based on statistical data of total production of tequila from the Tequila Regulatory Council (2019), peak tequila production occurred in 2018 which coincided the highest production level ever reached over the past 23 years. In 1995, 104.3 million liters of tequila were produced, but increased to 309.1 million liters in 2018. Similarly, the Tequila Regulatory Council (2019) recorded that global consumption of tequila increased from 279 thousand tons in 1995 to 1,139 thousand tons in 2018. The Secretariat of Agriculture Livestock, Rural Development, Fisheries and Food (SAGARPA) in Mexico reported in 2017 that tequila, relative to mezcal, is the major product from *Agave* in an expansive growth phase in Mexico. Outside of Mexico, tequila is consumed mainly in the United States, Germany, Spain, France, and the United Kingdom. In addition, SAGARPA (2017) reported that the tequila industry generated approximately $27 million dollars from export revenue, which is predicted to increase to $28 million dollars by 2024 and $29 million by 2030. The mezcal industry has also experienced sustained growth from about 2.5 million liters in 1950 to about 20 million liters in 2010 (Martínez Salvador et al., 2012).

Sweeteners and Syrups

Fructose sugars extracted from blue agave have also been used as alternative sweeteners (Heyer and Crawford, 2009; Stewart, 2015). Despite disadvantages of a relatively long life cycle and the monocarpic habit of agaves compared with other annual and perennial crops used in the sugar industry, *Agave* sugars are used as high-quality sweeteners. This emerging product can potentially work as a companion to the tequila industry. As with *Agave* spirits, the sweetener industry uses *Agave* juices as feedstock (Heyer and Crawford, 2009; Narváez-Zapata and Sánchez-Teyer, 2010). Sugars from *Agave* juice, particularly fructose, are extracted through acid or enzymatic hydrolysis (Garcia-Aguirre et al., 2009; Ávila-Fernández et al., 2011; Soto et al., 2011). The fructose sugars are used as additives in commercial *Agave* syrups, which are considered healthier sweeteners compared with sugar cane and high-fructose corn syrup (Hooshmand et al., 2014). The proportion of fructose in *Agave* syrup is significantly higher compared with the proportions in cane sugar and high-fructose corn syrup. Proportions of fructose to glucose are 50/50 in cane sugar (Glasziou, 1961); 55/45 in high fructose corn syrup (O'Brien-Nabors, 2001); and as high as 95/5 in *Agave* syrup (Garcia-Aguirre et al., 2009).

Modern Uses of Agave Fibers

Historically, *Agave sisalana, Agave fourcroydes* and *Agave lechuguilla* fibers have been used in the Mexican textile industry. Traditional uses of *Agave* fiber include ropes, twine, bags, mecapales, fabrics, brushes, and brooms (Colunga-GarcíaMarín and May-Pat, 1993; Kicińska-Jakubowska et al., 2012). Sisal is a hard fiber processed from the leaves of *A. sisalana*. Henequen fiber from *A. fourcroydes* and Tampico fiber (also called Mexican fiber) from *A. lechuguilla* have similar tensile and flexural properties as sisal fiber (Belmares et al., 1981; Kicińska-Jakubowska et al., 2012). More recently, *Agave* fibers have been used to reinforce industrial products, adding flexibility and strength to polymer-based composites (Joseph et al., 1999; Silva et al., 2010; Orue et al., 2016). Fiber-reinforced polymers have application in the aerospace, marine, automotive, military, and construction industries (Yilmaz and Arifuzzaman Khan, 2019). Compounds derived from *Agave* fibers, can also be used in synthetic drug manufacturing (Cushman et al., 2015). Steroidal saponins, tigogenin, and hecogenin are natural compounds extracted from *A. sisalana* leaves, which are used in the synthesis of steroidal hormones such as corticosteroids (Cripps and Blunden, 1978; Santos and Branco, 2014). Corticosteroids drugs like dexamethasone can be synthesized from tigogenin and hecogenin (Kongkathip et al., 1997; Santos and Branco, 2014) and may have application in treating respiratory-inflammatory conditions associated with COVID-19 produced by 2 SARS-CoV-2 (Al Saleh et al., 2020; McIntosh, 2020; Zhang et al., 2020).

Potential Uses of Agaves for Bioenergy

Agaves have also been proposed as a biofuel crop due to their relatively low lignin content (Somerville et al., 2010; Davis et al., 2011). The high lignin content of C_3 and C_4 biofuel crops reduces the efficiency of converting sugars into bioethanol (Somerville et al., 2010). Lignin percentages in agaves range between 3 and 15% (Iñiguez-Covarrubias et al., 2001; Li et al., 2012; Delfín-Ruíz et al., 2019), and can be more efficiently processed to produce sugars than C_4 crops used in the biofuel industry (Somerville et al., 2010).

Many crops used in the biofuel industry require high amounts of irrigation water, generating controversy related to their environmental footprint (Somerville, 2007; Moore et al., 2014). Agaves require much less water than C_4 crops, such as corn, to produce biomass, and the superior quality of ethanol derived from *Agave* compared with corn makes agaves an attractive alternative (Yan et al., 2011). In addition to the low lignin content

of *Agave*, CAM metabolism enables their growth in marginal lands and resilience to drought (Davis et al., 2011).

One challenge in the widespread use of agaves in the biofuel industry is the lack of agronomic knowledge for its cultivation, as well as the underlying ecology and climatic conditions of regions where this crop may be suitable (McDaniel, 1985). Lewis et al. (2015) suggested the U.S. Southwest, particularly Arizona, as one region for *Agave* cultivation for the biofuel industry, due to its suitable climate. Another constraint on *Agave* production involves the large amount of annual plant biomass needed to supply enough raw material to make it profitably sustainable as a bioenergy crop (Balan, 2014).

According to Escamilla-Treviño (2012), commercial cultivation of agaves for biofuel in the U.S. has been constrained mainly by the risk of low-temperature crop damage. For example, cultivation of species with a frost tolerance between −2 and −4°C, such as *A. tequilana*, *A. fourcroydes*, *A. angustifolia*, *A. salmiana*, and *A. sisalana* (Nobel, 2003), could be limited even in Arizona, where nocturnal low temperatures below 0°C occur throughout the winter season. Agaves could be genetically engineered to improve traits, which would allow for better adaptation from temperate to xeric environments, which would enable agaves to be widely cultivated in marginal environments in the U.S. (Yang et al., 2016). Another approach could be to use *Agave* species that have relatively wide cold tolerance, such as *A. americana* and *Agave utahensis*, which have been reported to adapt well to cold and hot temperatures in the region and can tolerate temperatures between −8 and −11°C (Nobel and Jordan, 1983; Escamilla-Treviño, 2012; Davis et al., 2017). Moreover, species putatively cultivated by the Hohokam in pre-Colombian times, including *A. palmeri*, *A. murpheyi*, *A. parryi*, and *A. sanpedroensis*, could potentially be used as crops in the future because they are endemic and well-adapted to hot summers, cold winters, and the dry climate of the Sonoran Desert (Adams and Adams, 1998; Parker et al., 2010; Hodgson and Salywon, 2013; Fish and Fish, 2014; Hodgson et al., 2019).

ROCK PILES TO CULTIVATE AGAVES IN MARGINAL LANDS

Rock Piles as Dryland Farming System for Agaves in Marginal Lands

Agave rock-pile fields are a cultural and agricultural legacy of the ancient Hohokam tribe in the Sonoran Desert (Hodgson et al., 2019). Rock piles are specialized features associated with the ancient practice of *Agave* cultivation in marginal lands (Dobyns, 1988; Fish and Fish, 1990, 1992, 2012, 2014; Sandor and Homburg, 2017). Because CAM metabolism enables agaves to grow well in water-limited environments and in nutrient-poor soils (Gentry, 1982; Nobel and Valenzuela, 1987; Nobel, 1991, 2003; Garcia-Moya et al., 2011), *Agave* was and is a well-suited crop for marginal lands. Prehistoric groups from central and southern Arizona lived in marginal lands with limited access to irrigation water (Fish and Fish, 1992). Rainfall was the primary source of water to irrigate agaves in rock piles. Rainwater in rock-pile fields was harvested in two ways: through the interception of rainfall hitting rocks and from rain-water runoff (Crown, 1987; Fish and Fish, 1992; Lightfoot, 1994, 1996).

Rock piles were built on downhill slopes such that the inclined angle mitigated downward runoff and enabled rock piles to reduce erosion and increase fertility of *Agave* fields (Fish and Fish, 1992; Sandor and Homburg, 2011). In such rock piles, rainwater generally flows along soil slopes, and upon intercepting rock piles, it slows down, leading to increased soil moisture beneath the rock piles (Fish and Fish, 1990, 1992; Homburg and Sandor, 2011). The reduced flow of rainwater leads to less gully formation. Likewise, sediments and minerals were mixed in the rainfall runoff and deposited underneath rocks. According to Homburg and Sandor (2011), the accumulation of minerals under rock piles improved the texture of the soil-surface horizons and increased soil moisture retention capacity. The minerals, sediments, and organic matter deposited below rock piles were a source of C, N, and P, which provided a source of soil fertility for agaves cultivated by ancient indigenous groups, such as the Hohokam.

Agave Species Likely Cultivated by the Hohokam in Rock Piles

Evidence of cultivation of different *Agave* species in rock piles includes plant tissue, such as spines and fibers at nearby roasting pits and artifacts found in rock-pile fields, which were likely used to process and harvest agaves, such as tabular knives and scrapers (Cantley, 1991; Fish and Fish, 1992; Ciolek-Torrello et al., 1997). Among *Agave* species native to Arizona, *A. murpheyi* and *A. sanpedroensis* have been recognized as species that were cultivated in rock piles by the Hohokam (Adams and Adams, 1998; Parker et al., 2007; Hodgson et al., 2019). Moreover, researchers have identified *Agave yavapaiensis*, *Agave verdensis*, and *Agave delamateri* as pre-Columbian *Agave* cultigens (Parker et al., 2007; Hodgson and Salywon, 2013). Similarly, *A. parryi* has been associated with archaeological sites and dryland farming in Arizona (Minnis and Plog, 1976; Parker et al., 2010, 2014). Evidence of other cultivated plants used by the Hohokam include pollen grains of corn and cotton, which were found in rock piles (Crown, 1987; Fish, 1988; Bohrer, 1991). Other native species, such as *Opuntia spp*., *Carnegia gigantea*, *Chenopodium* spp., *Amaranthus* spp., *Trianthema portulacastrum*, *Spharalcea ambigua, Boerhaavia spp*., and cholla (*Cylindropuntia fulgida*) have been found to populate rock-pile fields in archaeological sites (Fish et al., 1986; Crown, 1987; Bohrer, 1991; Hodgson et al., 2019).

POTENTIAL BENEFITS OF USING ROCK PILES TO CULTIVATE AGAVES

Rock Pile Fields for Agaves and Their Environments

While rock piles can be found at archaeological sites throughout the southwestern U.S. and northwestern Mexico, most are located in south-central Arizona (Fish et al., 1985; Fish and Fish, 1990, 1992), providing an ideal setting for modern *Agave* cultivation that could incorporate aspects of prehistoric rock-pile fields. Rainfall and temperatures at rock-pile fields in Marana,

Tucson, and San Pedro Valley, Arizona, whose elevations range between 600 and 900 m above sea level, suggest that Hohokam agaves were cultivated in an optimum environment that balanced temperature, rainfall, and soil moisture (Fish and Fish, 1990, 1992; Cantley, 1991; Hodgson et al., 2019). This unique balance likely maximized productivity of agaves, even in the dry and harsh conditions of the region.

Different studies using the environmental productivity index developed by Nobel for agaves (Nobel and Hartsock, 1986; Nobel and Quero, 1986; Nobel and Valenzuela, 1987; Garcia-Moya et al., 2011) found, in general, that mesic environments, such as archaeological sites with Hohokam rock piles, can lead to improved *Agave* productivity. The index indicates that CO_2 uptake and productivity of agaves is greater at elevations between 600 and 1200 m above sea level. In addition, Woodhouse et al. (1980) found that agaves are less productive when cultivated on steep slopes. Hohokam rock-pile fields occur more frequently on softly inclined slopes, which have higher rainfall moisture interception, and less negative soil water potentials than found on relatively steeper slopes (Cantley, 1991; Fish and Fish, 1992). Such conditions possibly promoted better interception of photosynthetic radiation and rainfall, which would have led to enhanced CAM photosynthesis and biomass of agaves. Understanding Hohokam rock pile field environments can help to identify potential locations to cultivate agaves, even during severe drought events. However, more information needs to be sought out to determine the agricultural limitations and future applications of rock piles in modern times.

Soil-Water Dynamics Under Rock Piles

Hohokam rock piles functioned as a type of mulching that reduced soil evapotranspiration and used rainwater to increase soil moisture content underneath rocks (Doelle, 1978; Fish et al., 1985; Fish and Fish, 1990, 1992; Lightfoot, 1996). The positive effects of available moisture in agaves have been observed in different experiments. In an experiment with *A. deserti*, Jordan and Nobel (1979) observed that rainfall and soil moisture act as the most important factors influencing plant mortality in their first year of establishment. They also found that rainfall stimulated increased succulence, increased leaf growth, and helped modulate nocturnal CO_2 gas exchange and water-use efficiency of first-year plants. Davis et al. (2017) found that irrigation increased efficiency of nocturnal CO_2 uptake of *A. americana*. Similarly, Nobel et al. (1989) observed that irrigation doubled CO_2 uptake of *Agave lechuguilla* and enhanced aerial and root biomass. Lightfoot (1996) and Sandor and Homburg (2011) hypothesized that the moisture harvested underneath rock piles from rainwater improved cultivation of crops, including *Agave*, in rock piles (Fish and Fish, 2014). In an experiment using agaves in rock piles conducted in different locations in central and southern Arizona, Fish and Fish (2014) found that seasonal rains replenished soil moisture below rock piles, which improved survival rates of *A. murpheyi* and *A. americana*. Nobel et al. (1992b) reported that after watering rocks with 10–30 mL of water, soil volumetric water content increased below rocks for a period ranging between 13 and 19 days, leading to increased nocturnal CO_2 uptake of *A. deserti*.

While moisture underneath rock piles likely enhanced *Agave* biomass productivity, Eickmeier and Adams (1978) indicated that available water and air temperature are the two most influential factors that affect *Agave* carbon assimilation. Although day-night temperature is an important factor in *Agave* nocturnal CO_2 uptake, as observed with *A. angustifolia* and *A. americana* (Holtum and Winter, 2014), soil moisture governs biomass productivity of agaves (Huang and Nobel, 1992). Nobel and Quero (1986) indicated that available soil moisture in summer and fall in the Sonoran and Chihuahuan Deserts acts as the driving factor that stimulates biomass of *Agave* plants by promoting emergence of new leaves, development of large aerial shoots, and enhancement of root hydraulic conductance. In addition, Nobel (1976) observed that leaf size and soil water content correlated with higher nocturnal CO_2 uptake of *A. deserti*. However, the benefits of increasing soil moisture by using rock piles needs to be further explored through additional lab and field experiments.

Soil Temperature Underneath Rock Piles

Cool temperatures below rock piles can reduce heat stress and desiccation of roots of *Agave* plants (Huang and Nobel, 1992). Since daily soil temperatures in the Sonoran Desert can reach 75°C (Nobel, 2003), rock piles can reduce soil-moisture evaporation rates (Sandor and Homburg, 2011), and also work as a barrier to reduce interception of solar radiation, which leads to cooler diurnal soil temperatures relative to that found in exposed soil (Wilken, 1972). Studies made on the thermal properties of rock piles in *A. deserti* and *A. americana* illustrate the advantages of rock piles as insulation to diurnal hot temperatures. Palta and Nobel (1989) observed that low soil temperatures underneath rocks positively affected *A. deserti* root respiration and reduced root dryness. Nobel et al. (1992b) measured *A. deserti* roots underneath boulders or rock fragments and compared roots of agaves growing in exposed soils and found that low temperatures and less-negative soil water potentials underneath rocks increased root number, thickness, and length.

Kaseke et al. (2012) found that convective heat transference of rock mulch can keep soils cooler during the day and increase nocturnal soil temperatures. However, since little is known regarding patterns of diurnal and nocturnal temperatures underneath and within Hohokam rock piles, characterization of such properties is necessary in similarly arranged rock piles. In addition, experimentation is needed to characterize nocturnal convective heat transference underneath rock piles, and the level of nocturnal CO_2 uptake of agaves in rock piles.

Insulative properties of rocks and their effect on temperatures below rock piles likely have a positive effect on symbiosis of microbes with agaves (Cui and Nobel, 1992). A study on the effects of soil temperatures on vesicular-arbuscular mycorrhizae infection found that soil temperatures around 25°C increased yields of *Sorghum bicolor* and *Triticum aestivum* due to high colonization of roots by mycorrhizae (Fabig et al., 1989). Little is known, however, about the effect of Hohokam rock-pile temperatures on the soil microbiome and their associated benefits. Nevertheless, the *Agave* rhizosphere is diverse in prokaryotic and fungal microorganisms, and is

correlated with the hosting capability of agaves and their adaptions to arid climates (Coleman-Derr et al., 2016). Symbiosis of *Agave* roots with soil microbes enhances root hydraulic conductance and nutrient uptake, particularly solubilizing P. Cui and Nobel (1992) observed that colonization of *A. deserti* with mycorrhizae improved hydraulic conductance, uptake of P in roots, and P allocation in leaves. In addition, colonization of mycorrhizae positively correlated with enhanced CO_2 uptake of *A. deserti*. Plant symbiosis with arbuscular mycorrhizae, ecto-mycorrhizae, ericoid mycorrhizae, and various bacteria contributes to increased uptake of N and P in the form of phosphates (Mensah et al., 2015). In a study where endophytic bacteria were isolated from the base of *A. tequilana* plants, Martínez-Rodríguez et al. (2014) identified the presence of 300 strains of bacteria with different capacities and benefits, such as N fixation, P solubilization, auxin production, and antagonism against *Fusarium oxysporum*.

Soil-Based Nutrients Underneath Rock Piles

Agaves in the wild are well-adapted to arid regions and generally perform adequately in rocky, nutrient-poor soils (Gentry, 1972, 1982). However, soil-based nutrients underneath rock piles (Homburg and Sandor, 2011; Sandor and Homburg, 2017) can bolster *Agave* primary productivity (Nobel et al., 1992a). Soil research by Homburg and Sandor (2011) suggests that the use of Hohokam rock piles enriched C, N, and available P to agaves due to organic-matter accumulation. Soil nutrient accumulation underneath Hohokam rock piles possibly occurred due to runoff, microbial decomposition of organic matter, or soil bioturbation. The available nutrients below rock piles likely enhanced the physiological response of agaves to drought and extreme temperatures, and improved growth and productivity. However, it is necessary to assess the dynamics between soil-based nutrients underneath rock piles with *Agave* nutrient assimilation to more fully understand the benefits afforded by rock piles to *Agave* productivity. In addition, agricultural parameters, such as soil pH levels, soil electric conductivity, and nutrient cycling in the soil beneath rock piles, need future research.

Nutrients in the soil under rock piles, as observed by Homburg and Sandor (2011), contributed to the productivity of cultivated agaves. Available nutrients in the soil assist in the productivity of agaves (Nobel et al., 1992a). Nobel et al. (1988) observed that fertilization with N, P, K, and B enhanced growth and nocturnal CO_2 uptake of *A. lechuguilla*. Similarly, irrigation, in combination with fertilization with N, P, and K, increased foliar leaf area, leaf number, and concentration of sugars, particularly fructose and glucose in *A. tequilana* and *Agave potatorum* (Martínez et al., 2012; Zúñiga-Estrada et al., 2018). Valenzuela and Gonzalez (1995) found that fertilization of *A. lechuguilla* and *A. tequilana* with P and N increased leaf area. Similarly, for *A. deserti*, Nobel et al. (1989) observed that fertilization promoted leaf growth and high rates of CO_2 uptake, which led to high biomass accumulation.

Opportunity for Researching Agaves in Rock Piles

Relict rock piles at archaeological sites represent a valuable agricultural example of how the Hohokam made marginal lands productive by cultivating agaves during severe droughts. While a number of archaeologists reported that rock piles were the main dryland-farming strategy used by the Hohokam to cultivate agaves (Masse, 1979; Fish et al., 1985; Crown, 1987; Dobyns, 1988; Fish and Fish, 1990, 1992; Cantley, 1991; Lightfoot, 1996; Sandor and Homburg, 2011), little is known about the agronomic potential and applications of rock piles in modern *Agave* cultivation. To bring to light possible uses of rock piles, it is necessary to sort through what has been published regarding the environmental details of rock-pile fields in order to experimentally replicate these ancient agroecosystems.

Experiments that assess cultivation, pest management, and the physiological responses of agaves using rock piles are needed, particularly to observe plant productivity, CO_2 uptake, temperature requirements, and soil-plant water relations in rock piles. In addition, characterizing the hydrothermal properties of rock piles requires examining how they can preserve soil moisture and modulate soil temperatures. The microbiome and fauna of rock piles are additional factors that could potentially enhance nutrition, improve water status, and contribute to general plant health. Microbiomes in rock piles can positively impact plant health of agaves in future droughts, particularly in preventing pests and disease. Future research is needed on the environmental, social, and economic impacts of using rock piles to cultivate agave, particularly in the continually changing and fragile agroecosystems of dry regions.

CONCLUSIONS

Further experimentation and development of innovative agricultural strategies is crucial to the use of agaves as a crop under scenarios of severe drought and global warming. Throughout history, agaves have been used as commodities for food (Anderies et al., 2008; Delgado-Lemus et al., 2014), beverages (Walton, 1977; Stewart, 2015; Escalante et al., 2016), and fiber (Colunga-GarcíaMarín and May-Pat, 1993). In modern times, agaves continue to be used for such purposes, but are now also used as substrates for sweeteners (Heyer and Crawford, 2009; Stewart, 2015), biofuels (Somerville et al., 2010; Davis et al., 2011), synthetic drugs (Santos and Branco, 2014; Cushman et al., 2015) and industrial materials (Silva et al., 2010; Orue et al., 2016). We suggest dryland farming of *Agave* as a means to minimize the use of irrigation water and sustainably maintain crop productivity in arid regions.

Current challenges to successful cultivation of agaves, such as low rainfall and excessive heat in arid regions and marginal lands, are similar to those that prehistoric, indigenous farmers faced during droughts in the Sonoran Desert. The

pre-Columbian Hohokam were skilled in the use of rock piles to cultivate agaves during droughts (Dobyns, 1988; Fish and Fish, 1992). These rock piles acted as a mulch that harvested rainwater moisture, preserved soil moisture, reduced soil evapotranspiration, and insulated soil in their immediate environs. Despite the lack of empirical data, moisture harvested during the monsoon season beneath and around rock piles likely decreased drought stress, stimulating biomass productivity of agaves.

The use of rock piles for *Agave* cultivation promises ecological benefits, such as minimizing soil erosion and maximizing crop productivity in marginal lands with minimal input of chemical fertilization and pesticides

Rock pile cultivation of agaves is promising, but it requires field-based research to characterize their productive potential. More research is also needed to understand how the Hohokam rock-pile system could be used to cultivate crops other than agaves. However, with even from the little that is known, rock piles provide a sustainable crop-production-technology alternative for efficient use of water in dry areas and to revive cultivation of agaves in limited-resource environments in the region.

AUTHOR CONTRIBUTIONS

This manuscript is a collaborative effort of different individuals. The manuscript was written by HO-C and JS. The following collaborators JH-H, NH, SP, MS, RM-G, TC-M, AV-M, and PP provided valuable feedback and reviewed the final version of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

Funding was provided through in-house university funds. Funding was provided through, Wildlife and Wildlans Conservations, PhD program, Department of Wildlife and Wildlife Sciences, Brigham Young University.

ACKNOWLEDGMENTS

We express our sincere gratitude to all the collaborators and reviewers that made this paper possible. We would like to extend a special thank you to Paul and Suzanne Fish and Wendy Hodgson for their encouragement and for sharing their knowledge with us about Hohokam Agave cultivation.

REFERENCES

Adams, K. R., and Adams, R. K. (1998). *How Does Our Agave Grow? Reproductive Biology of a Suspected Ancient Arizona Cultivar, Agave murpheyi Gibson. CALS Publ. Arch. Univ. Ariz.* Available online at: https://repository.arizona.edu/handle/10150/554320 (accessed October 11, 2019).

Al Saleh, A. S., Sher, T., and Gertz, M. A. (2020). Multiple myeloma in the time of COVID-19. *Acta Haematol.* 143, 1–7. doi: 10.1159/000507690

Anderies, J. M., Nelson, B. A., and Kinzig, A. P. (2008). Analyzing the impact of agave cultivation on famine risk in arid pre-Hispanic northern Mexico. *Hum. Ecol.* 36, 409–422. doi: 10.1007/s10745-008-9162-9

Arizaga, S., and Ezcurra, E. (2002). Propagation mechanisms in *Agave macroacantha* (Agavaceae), a tropical arid-land succulent rosette. *Am. J. Bot.* 89, 632–641. doi: 10.3732/ajb.89.4.632

Armillas, P. (1948). A sequence of cultural development in Meso-America. *Mem. Soc. Am. Archaeol.* 4, 105–111. doi: 10.1017/S0081130000000423

Ashkenazi, E., Avni, Y., and Avni, G. (2012). A comprehensive characterization of ancient desert agricultural systems in the Negev Highlands of Israel. *J. Arid Environ.* 86, 55–64. doi: 10.1016/j.jaridenv.2012.02.020

Ávila-Fernández, Á., Galicia-Lagunas, N., Rodríguez-Alegría, M. E., Olvera, C., and López-Munguía, A. (2011). Production of functional oligosaccharides through limited acid hydrolysis of agave fructans. *Food Chem.* 129, 380–386. doi: 10.1016/j.foodchem.2011.04.088

Balan, V. (2014). Current challenges in commercially producing biofuels from lignocellulosic biomass. *ISRN Biotechnol.* 2014, 463074–463074. doi: 10.1155/2014/463074

Bautista-Cruz, A., Carrillo-González, R., Arnaud-Viñas, M. R., Robles, C., and de León-González, F. (2007). Soil fertility properties on *Agave angustifolia* Haw. *plantations. Soil Tillage Res.* 96, 342–349. doi: 10.1016/j.still.2007.08.001

Belmares, H., Barrera, A., Castillo, E., Verheugen, E., Monjaras, M., Patfoort, G. A., et al. (1981). New composite materials from natural hard fibers. *Ind. Eng. Chem. Prod. Res. Dev.* 20, 555–561. doi: 10.1021/i300003a026

Bennett, W. C. (1948). The Peruvian co-tradition. *Mem. Soc. Am. Archaeol.* 4, 1–7. doi: 10.1017/S0081130000000289

Biazin, B., Sterk, G., Temesgen, M., Abdulkedir, A., and Stroosnijder, L. (2012). Rainwater harvesting and management in rainfed agricultural systems in sub-Saharan Africa–a review. *Phys. Chem. Earth Parts ABC* 47, 139–151. doi: 10.1016/j.pce.2011.08.015

Bohrer, V. L. (1991). Recently recognized cultivated and encouraged plants among the Hohokam. *Kiva* 56, 227–235. doi: 10.1080/00231940.1991.11758169

Borland, A. M., Barrera Zambrano, V. A., Ceusters, J., and Shorrock, K. (2011). The photosynthetic plasticity of crassulacean acid metabolism: an evolutionary innovation for sustainable productivity in a changing world. *New Phytol.* 191, 619–633. doi: 10.1111/j.1469-8137.2011.03781.x

Borland, A. M., Griffiths, H., Hartwell, J., and Smith, J. A. C. (2009). Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. *J. Exp. Bot.* 60, 2879–2896. doi: 10.1093/jxb/erp118

Borland, A. M., Wullschleger, S. D., Weston, D. J., Hartwell, J., Tuskan, G. A., Yang, X., et al. (2015). Climate-resilient agroforestry: physiological responses to climate change and engineering of crassulacean acid metabolism (CAM) as a mitigation strategy. *Plant Cell Environ.* 38, 1833–1849. doi: 10.1111/pce.12479

Bowen, S. (2015). *Divided Spirits: Tequila, Mezcal, and the Politics of Production*. Oakland, CA: Univ of California Press. doi: 10.1525/california/9780520281042.001.0001

Bradley, R. S., Hughes, M. K., and Diaz, H. F. (2003). Climate in medieval time. *Science* 302, 404–405. doi: 10.1126/science.1090372

Callendar, G. S. (1938). The artificial production of carbon dioxide and its influence on temperature. *Q. J. R. Meteorol. Soc.* 64, 223–240. doi: 10.1002/qj.49706427503

Cantley, G. J. (1991). *Archaeology of a Rockpile Field in the Santan Mountains, Arizona*. MA thesis, Phoenix, AZ, Arizona State University.

Castetter, E. F., Bell, W. H., and Grove, A. R. (1938). *The Early Utilization and the Distribution of Agave in the American Southwest*. Albuquerque: University of New Mexico Press.

Cervantes, M., T., Armenta-Calderón, A. D., and Sánchez-Arellano, J. G. (2007). *El Cultivo del Maguey Bacanora (Agave angustifolia Haw.) en la Sierra de Sonora*. Hermosillo: Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP).

Chapagain, T., and Raizada, M. N. (2017). Agronomic challenges and opportunities for smallholder terrace agriculture in developing countries. *Front. Plant Sci.* 8:331. doi: 10.3389/fpls.2017.00331

Chu, G., Liu, J., Sun, Q., Lu, H., Gu, Z., Wang, W., et al. (2002). The 'Medieval Warm Period' drought recorded in Lake Huguangyan, tropical South China. *Holocene* 12, 511–516. doi: 10.1191/0959683602hl566ft

Ciolek-Torrello, R., Homburg, J. A., and Sandor, J. (1997). "Vanishing River Volume 2: Agricultural, Subsistence, and Environmental Studies: Part 2: Chapters 4-7," in *Vanishing River: Landscapes and Lives of the Lower Verde Valley: The Lower Verde Archaeological Project: Volume 2: Agricultural, Subsistence, and Environmental Studies*. (Tucson, AZ: Statistical Research, Inc. Press), 57–147.

Coleman-Derr, D., Desgarennes, D., Fonseca-Garcia, C., Gross, S., Clingenpeel, S., Woyke, T., et al. (2016). Plant compartment and biogeography affect microbiome composition in cultivated and native *Agave* species. *New Phytol.* 209, 798–811. doi: 10.1111/nph.13697

Colunga-GarcíaMarín, P., and May-Pat, F. (1993). Agave studies in Yucatan, Mexico. I. Past and present germplasm diversity and uses. *Econ. Bot.* 47, 312–327. doi: 10.1007/BF02862301

Colunga-GarcíaMarín, P., and Zizumbo-Villarreal, D. (2006). "Tequila and other agave spirits from west-central Mexico: current germplasm diversity, conservation and origin," in *Plant Conservation and Biodiversity* (Dordrecht: Springer), 79–93. doi: 10.1007/978-1-4020-6444-9_6

Cripps, A. L., and Blunden, G. (1978). A quantitative gas-liquid chromatographic method for the estimation of hecogenin and tigogenin in the leaves, juice and sapogenin concentrates of agave sisalana. *Steroids* 31, 661–669. doi: 10.1016/S0039-128X(78)80006-3

Crown, P. L. (1987). Classic period Hohokam settlement and land use in the Casa Grande Ruins area, Arizona. *J. Field Archaeol.* 14, 147–162. doi: 10.1179/009346987792208466

Cui, M., and Nobel, P. S. (1992). Nutrient status, water uptake and gas exchange for three desert succulents infected with mycorrhizal fungi. *New Phytol.* 122, 643–649. doi: 10.1111/j.1469-8137.1992.tb00092.x

Cushman, J. C., Davis, S. C., Yang, X., and Borland, A. M. (2015). Development and use of bioenergy feedstocks for semi-arid and arid lands. *J. Exp. Bot.* 66, 4177–4193. doi: 10.1093/jxb/erv087

Davis, S. C., Dohleman, F. G., and Long, S. P. (2011). The global potential for agave as a biofuel feedstock. *GCB Bioenergy* 3, 68–78. doi: 10.1111/j.1757-1707.2010.01077.x

Davis, S. C., Kuzmick, E. R., Niechayev, N., and Hunsaker, D. J. (2017). Productivity and water use efficiency of *Agave americana* in the first field trial as bioenergy feedstock on arid lands. *GCB Bioenergy* 9, 314–325. doi: 10.1111/gcbb.12324

De Micco, V., and Aronne, G. (2012). "Morpho-anatomical traits for plant adaptation to drought," in *Plant Responses to Drought Stress* (Springer), 37–61. doi: 10.1007/978-3-642-32653-0

Delfín-Ruíz, M. E., Calderón-Santoyo, M., Ragazzo-Sánchez, J. A., Gómez-Rodríguez, J., López-Zamora, L., and Aguilar-Uscanga, M. G. (2019). Acid pretreatment optimization for xylose production from *Agave tequilana* Weber var. *azul, Agave americana var. oaxacensis, Agave karwinskii,* and *Agave potatorum bagasses* using a Box-Behnken design. *Biomass Convers. Biorefinery.* 1–10. doi: 10.1007/s13399-019-00497-z

Delgado-Lemus, A., Casas, A., and Téllez, O. (2014). Distribution, abundance and traditional management of *Agave potatorum* in the Tehuacán Valley, Mexico: bases for sustainable use of non-timber forest products. *J. Ethnobiol. Ethnomed.* 10:63. doi: 10.1186/1746-4269-10-63

Denevan, W. M. (2003). *Cultivated Landscapes of Native Amazonia and the Andes*. New York, NY: Oxford University Press.

Díaz, G., Rodgers, A., and Byland, B. E. (1993). *The Codex Borgia: A Full-Color Restoration of the Ancient Mexican Manuscript*. Mineola, NY: Courier Corporation.

Dobyns, H. F. (1988). *Piman Indian Historic Agave Cultivation*. Tucson, AZ: Desert Plants USA.

Doelle, W. H. (1978). "Hohokam use of nonriverine resources", in *Experiments in the Archaeology of the American Southwest*, ed P. Grebinger, 245–274.

Doolittle, W. E. (1995). Indigenous development of Mesoamerican irrigation. *Geogr. Rev.* 85:301. doi: 10.2307/215275

Doyel, D. E. (2007). "Irrigation, production, and power in Phoenix basin Hohokam society," in *Hohokam Millenn* (Santa Fe: School of Advanced Research Press), 82–89.

Eickmeier, W. G., and Adams, M. S. (1978). Gas exchange in *Agave lecheguilla* Torr. (Agavaceae) and its ecological implications. *Southwest. Nat.* 22, 473–485. doi: 10.2307/3670254

Enríquez-Salazar, M. I., Veana, F., Aguilar, C. N., Iliana, M., López, M. G., Rutiaga-Quinones, O. M., et al. (2017). Microbial diversity and biochemical profile of aguamiel collected from *Agave salmiana* and *A. atrovirens* during different seasons of year. *Food Sci. Biotechnol.* 26, 1003–1011. doi: 10.1007/s10068-017-0141-z

Escalante, A., López, D. S., Velázquez, J. G., Giles-Gómez, M., Bolívar, F., and López-Munguía, A. (2016). Pulque, a traditional mexican alcoholic fermented beverage: historical, microbiological, and technical aspects. *Front. Microbiol.* 7:1026. doi: 10.3389/fmicb.2016.01026

Escamilla-Treviño, L. L. (2012). Potential of plants from the genus agave as bioenergy crops. *BioEnergy Res.* 5, 1–9. doi: 10.1007/s12155-011-9159-x

Evans, S. T. (1990). The productivity of maguey terrace agriculture in central Mexico during the Aztec period. *Lat. Am. Antiq.* 1, 117–132. doi: 10.2307/971983

Evenari, M., Shanan, L., Tadmor, N., and Shkolnik, A. (1982). *The Negev: The Challenge of a Desert*. Cambridge: Harvard University Press. doi: 10.4159/harvard.9780674419254

Fabig, B., Moawad, A. M., and Achtnich, W. (1989). Effect of VA Mycorrhiza on dry weight and phosphorus content in shoots of cereal crops fertilized with rock phosphates at different soil pH and temperature levels. *Z. Für Pflanzenernähr. Bodenkd.* 152, 255–259. doi: 10.1002/jpln.19891520218

Fischbeck, S. L. (2001). *Agricultural Terrace Productivity in the Maya Lowlands of Belize*. Ph.D., dissertation, University of Wisconsin.

Fish, P. R., Fish, S. K., Long, A., and Miksicek, C. (1986). Early corn remains from Tumamoc Hill, southern Arizona. *Am. Antiq.* 51, 563–572. doi: 10.2307/281752

Fish, S., and Fish, P. (1990). "An archaeological assessment of ecosystem in the Tucson Basin of southern Arizona", in *The Ecosystem Approach in Anthropology: From Concept to Practice*, ed E. F. Moran, 159–160.

Fish, S. K. (1988). "Environment and subsistence: The pollen evidence," in *Recent Research on Tucson Basin Prehistory Proceedings of the Second Tucson Basin Conference, Institute for American Research Anthropological Papers* (Tucson AZ: Institute for American Research), 31–38.

Fish, S. K. (2000). "Hohokam impacts on Sonoran Desert environment", in *Imperfect Balance: Landscape Transformations in the Precolumbian Americas*, ed D. L. Lentz, 251–280. doi: 10.7312/lent11156-013

Fish, S. K., and Fish, P. R. (1992). Prehistoric landscapes of the Sonoran desert Hohokam. *Popul. Environ.* 13, 269–283. doi: 10.1007/BF01271027

Fish, S. K., and Fish, P. R. (2008). *The Hohokam Millennium*. Santa Fe: School for Advanced Research Press.

Fish, S. K., and Fish, P. R. (2012). "Hohokam Society and Water *Management*," in *The Oxford Handbook of North American Archaeology* (New York).

Fish, S. K., and Fish, P. R. (2014). "Agave (*Agave spp*.): a crop lost and found in the US-Mexico borderlands," in *New Lives for Ancient and Extinct Crops*, ed P. E. Minnis (Tucson, AZ: University of Arizona Press), 102–138.

Fish, S. K., Fish, P. R., Miksicek, C., and Madsen, J. (1985). *Prehistoric Cultivation in Southern Arizona*. Tucson, AZ: Desert plants USA.

Flores, N. B., and Araiza, P. L. S. (2012). El mezcal en Sonora, México, más que una bebida espirituosa. Etnobotánica de Agave angustifolia Haw. *Estud. Soc. Rev. Aliment. Contemp. Desarro. Reg.* 2, 173–197.

Fuentes-Rodriguez, J. (1997). A comparison of the nutritional value of opuntia and agave plants for ruminants. *J. Prof. Assoc. Cactus Dev.* 2, 20–22.

Galloway, P. R. (1986). Long-term fluctuations in climate and population in the preindustrial era. *Popul. Dev. Rev.* 12, 1–24. doi: 10.2307/1973349

Garay, T., and Aurea, M. (2008). *"El Recreo de los Amigos." Mexico City's Pulquerias During the Liberal Republic* (1856-1911). Available online at: https://repository.arizona.edu/handle/10150/194973 (accessed November 17, 2019).

Garcia-Aguirre, M., Saenz-Alvaro, V. A., Rodriguez-Soto, M. A., Vicente-Magueyal, F. J., Botello-Alvarez, E., Jimenez-Islas, H., et al. (2009). Strategy for biotechnological process design applied to the enzymatic hydrolysis of agave fructo-oligosaccharides to obtain fructose-rich syrups. *J. Agric. Food Chem.* 57, 10205–10210. doi: 10.1021/jf902855q

Garcia-Moya, E., Romero-Manzanares, A., and Nobel, P. S. (2011). Highlights for agave productivity. *Gcb Bioenergy* 3, 4–14. doi: 10.1111/j.1757-1707.2010.01078.x

Gasser, R. E., and Kwiatkowski, S. M. (1991). Regional signatures of Hohokam plant use. *Kiva* 56, 207–226. doi: 10.1080/00231940.1991.11758168

Gentry, H. S. (1972). *The Agave Family in Sonora*. Washington, DC: US Agricultural Research Service, US Department of Agriculture.

Gentry, H. S. (1982). *Agaves of Continental North America*. Tucson, AZ: University of Arizona Press.

Gibson, P. T. (1996). Correcting for inbreeding in parent-offspring regression estimates of heritability with non-additive and genotype ^{x} environment effects present. *Crop Sci.* 36, 594–600. doi: 10.2135/cropsci1996.0011183X003600030012x

Glasziou, K. T. (1961). Accumulation & transformation of sugars in stalks of sugar cane. Origin of glucose & fructose in the inner space ^{1}. *Plant Physiol.* 36, 175–179. doi: 10.1104/pp.36.2.175

Good-Avila, S. V., Souza, V., Gaut, B. S., and Eguiarte, L. E. (2006). Timing and rate of speciation in *Agave* (Agavaceae). *Proc. Natl. Acad. Sci. U.S.A.* 103, 9124–9129. doi: 10.1073/pnas.0603312103

Gutiérrez-Coronado, M. L., Acedo-Félix, E., and Valenzuela-Quintanar, A. I. (2007). Industria del bacanora y su proceso de elaboración. *CYTA-J. Food* 5, 394–404. doi: 10.1080/11358120709487718

Hansen, J., Johnson, D., Lacis, A., Lebedeff, S., Lee, P., Rind, D., et al. (1981). Climate impact of increasing atmospheric carbon dioxide. *Science* 213, 957–966. doi: 10.1126/science.213.4511.957

Hartung, T. (2016). *Cattail Moonshine & Milkweed Medicine: The Curious Stories of 43 Amazing North American Native Plants*. North Adams, MA: Storey Publishing.

Helama, S., Timonen, M., Holopainen, J., Ogurtsov, M. G., Mielikäinen, K., Eronen, M., et al. (2009). Summer temperature variations in Lapland during the Medieval Warm Period and the little Ice Age relative to natural instability of thermohaline circulation on multi-decadal and multi-centennial scales. *J. Quat. Sci. Publ. Quat. Res. Assoc.* 24, 450–456. doi: 10.1002/jqs.1291

Heyer, J. A., and Crawford, J. R. (2009). *Use of Natural Agave Extract as a Natural Sweetener Replacing Other Added Sweeteners in Food Products and Medicines*. U.S. Patent Application 11/999, 719.

Hodgson, W. C. (2001). Taxonomic novelties in American *Agave* (Agavaceae). *Novon* 11, 410–416. doi: 10.2307/3393152

Hodgson, W. C., and Salywon, A. M. (2013). Two new *Agave* species (Agavaceae) from central Arizona and their putative pre-Columbian domesticated origins. *Brittonia* 65, 5–15. doi: 10.1007/s12228-012-9255-z

Hodgson, W. C., Salywon, A. M., and Doelle, W. H. (2019). Hohokam lost crop found: A new Agave (Agavaceae) species only known from large-scale pre-Columbian agricultural fields in southern Arizona. *Syst. Bot.* 43, 734–740. doi: 10.1600/036364418X697445

Holtum, J. A. M., and Winter, K. (2014). Limited photosynthetic plasticity in the leaf-succulent CAM plant *Agave angustifolia* grown at different temperatures. *Funct. Plant Biol.* 41, 843–849. doi: 10.1071/FP13284

Homburg, J. A., and Sandor, J. A. (2011). Anthropogenic effects on soil quality of ancient agricultural systems of the American Southwest. *Catena* 85, 144–154. doi: 10.1016/j.catena.2010.08.005

Hooshmand, S., Holloway, B., Nemoseck, T., Cole, S., Petrisko, Y., Hong, M. Y., et al. (2014). Effects of agave nectar versus sucrose on weight gain, adiposity, blood glucose, insulin, and lipid responses in mice. *J. Med. Food* 17, 1017–1021. doi: 10.1089/jmf.2013.0162

Huang, B., and Nobel, P. S. (1992). Hydraulic conductivity and anatomy for lateral roots of *Agave deserti* during root growth and drought-induced abscission. *J. Exp. Bot.* 43, 1441–1449. doi: 10.1093/jxb/43.11.1441

Hughes, M. K., and Diaz, H. F. (1994). Was there a Medieval Warm Period, and if so, where and when? *Clim. Change* 26, 109–142. doi: 10.1007/978-94-011-1186-7_1

Hunt, R. C., Guillet, D., Abbott, D. R., Bayman, J., Fish, P., Fish, S., et al. (2005). Plausible ethnographic analogies for the social organization of Hohokam canal irrigation. *Am. Antiq.* 70, 433–456. doi: 10.2307/40035308

Ingram, S. E. (2010). *Human Vulnerability to Climatic Dry Periods in the Prehistoric US Southwest*. PhD dissertation, Arizona State University.

Iñiguez-Covarrubias, G., Diaz-Teres, R., Sanjuan-Dueñas, R., Anzaldo-Hernández, J., and Rowell, R. M. (2001). Utilization of by-products from the tequila industry. Part 2: potential value of *Agave tequilana* Weber azul leaves. *Bioresour. Technol.* 77, 101–108. doi: 10.1016/S0960-8524(00)00167-X

Jordan, P. W., and Nobel, P. S. (1979). Infrequent establishment of seedlings of *Agave deserti* (Agavaceae) in the northwestern Sonoran Desert. *Am. J. Bot.* 66, 1079–1084. doi: 10.1002/j.1537-2197.1979.tb06325.x

Joseph, P. V., Joseph, K., and Thomas, S. (1999). Effect of processing variables on the mechanical properties of sisal-fiber-reinforced polypropylene composites. *Compos. Sci. Technol.* 59, 1625–1640. doi: 10.1016/S0266-3538(99)00024-X

Kaseke, K. F., Mills, A. J., Henschel, J., Seely, M. K., Esler, K., and Brown, R. (2012). The effects of desert pavements (gravel mulch) on soil micro-hydrology. *Pure Appl. Geophys.* 169, 873–880. doi: 10.1007/s00024-011-0367-2

Kennett, D. J. (2012). *Archaic-Period Foragers and Farmers in Mesoamerica. The Oxford Handbook of Mesoamerican Archaeology*. New York, NY: Oxford University Press. 141–150. doi: 10.1093/oxfordhb/9780195390933.013.0010

Kicińska-Jakubowska, A., Bogacz, E., and Zimniewska, M. (2012). Review of natural fibers. Part I—vegetable fibers. *J. Nat. Fibers* 9, 150–167. doi: 10.1080/15440478.2012.703370

Kongkathip, N., Kongkathip, B., Noimai, N., and Suvipa San-oon. (1997). *Synthesis of Antiinflammatory Steroids Intermediates From Steroids Extracted From Industrial Waste of Agave Sisalana*. Warasan Witthayasat MoKo. Available online at: https://agris.fao.org/agris-search/search.do?recordID=TH2001002855 (accessed June 18, 2020).

Krausmann, F., Erb, K.-H., Gingrich, S., Haberl, H., Bondeau, A., Gaube, V., et al. (2013). Global human appropriation of net primary production doubled in the 20th century. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10324–10329. doi: 10.1073/pnas.1211349110

Ladefoged, T. N., Flaws, A., and Stevenson, C. M. (2013). The distribution of rock gardens on Rapa Nui (Easter Island) as determined from satellite imagery. *J. Archaeol. Sci.* 40, 1203–1212. doi: 10.1016/j.jas.2012.09.006

Laurance, W. F., Sayer, J., and Cassman, K. G. (2014). Agricultural expansion and its impacts on tropical nature. *Trends Ecol. Evol.* 29, 107–116. doi: 10.1016/j.tree.2013.12.001

Lemonnier, P., and Ainsworth, E. A. (2018). "Crop Responses to Rising Atmospheric [CO2] and Global Climate Change," in *Food Security and Climate Change* (Chichester: John Wiley & Sons, Ltd), 51–69. doi: 10.1002/9781119180661.ch3

Lewis, S. M., Gross, S., Visel, A., Kelly, M., and Morrow, W. (2015). Fuzzy gis-based multi-criteria evaluation for us agave production as a bioenergy feedstock. *Gcb Bioenergy* 7, 84–99. doi: 10.1111/gcbb.12116

Li, H., B. Foston, M., Kumar, R., Samuel, R., Gao, X., Hu, F., et al. (2012). Chemical composition and characterization of cellulose for Agave as a fast-growing, drought-tolerant biofuels feedstock. *RSC Adv.* 2, 4951–4958. doi: 10.1039/c2ra20557b

Lightfoot, D. R. (1994). Morphology and ecology of lithic-mulch agriculture. *Geogr. Rev.* 172–185. doi: 10.2307/215329

Lightfoot, D. R. (1996). The nature, history, and distribution of lithic mulch agriculture: an ancient technique of dryland agriculture. *Agric. Hist. Rev.* 44, 206–222.

Lüttge, U. (2004). Ecophysiology of crassulacean acid metabolism (CAM). *Ann. Bot.* 93, 629–652. doi: 10.1093/aob/mch087

Macías, M. A. (2001). El cluster en la industria del tequila en Jalisco, México[1]. *Agroalimentaria* 6, 55–72.

Mancilla-Margalli, N. A., and López, M. G. (2006). Water-soluble carbohydrates and fructan structure patterns from *Agave* and *Dasylirion* species. *J. Agric. Food Chem.* 54, 7832–7839. doi: 10.1021/jf060354v

Marcus, J. (2006). The roles of ritual and technology in Mesoamerican water management. *Agric. Strateg.* 221–254. doi: 10.2307/j.ctvdjrr1w.14

Martínez Salvador, M., Mata-González, R., Morales Nieto, C., and Valdez-Cepeda, R. (2012). Agave salmiana plant communities in central mexico as affected by commercial use. *Environ. Manage.* 49, 55–63. doi: 10.1007/s00267-011-9759-4

Martínez, R., S., Trinidad, S., A., Robles, C., Galvis, S., et al. (2012). Crecimiento y sólidos solubles de *Agave potatorum* Zucc. inducidos por riego y fertilización. *Rev. Fitotec. Mex.* 35, 61–68.

Martínez-Rodríguez, J., del, C., Mora-Amutio, M. D. la, Plascencia-Correa, L. A., Audelo-Regalado, E., Guardado, F. R., Hernández-Sánchez, E., et al. (2014). Cultivable endophytic bacteria from leaf bases of *Agave tequilana* and their role as plant growth promoters. *Braz. J. Microbiol.* 45, 1333–1339. doi: 10.1590/S1517-83822014000400025

Martorell, C., and Ezcurra, E. (2007). The narrow-leaf syndrome: a functional and evolutionary approach to the form of fog-harvesting rosette plants. *Oecologia* 151, 561–573. doi: 10.1007/s00442-006-0614-x

Masse, W. B. (1979). An intensive survey of prehistoric dry farming systems near Tumamoc Hill in Tucson, Arizona. *Kiva* 45, 141–186. doi: 10.1080/00231940.1979.11757933

McDaniel, R. G. (1985). *Field Evaluations of Agave in Arizona*. Tucson, AZ: Desert Plants USA.

McIntosh, J. J. (2020). Corticosteroid guidance for pregnancy during COVID-19 pandemic. *Am. J. Perinatol.* 37, 809–812. doi: 10.1055/s-0040-1709684

Mellado, M. (2016). Dietary selection by goats and the implications for range management in the Chihuahuan Desert: a review. *Rangel. J.* 38, 331–341. doi: 10.1071/RJ16002

Mensah, J. A., Koch, A. M., Antunes, P. M., Kiers, E. T., Hart, M., and Bücking, H. (2015). High functional diversity within species of arbuscular mycorrhizal fungi is associated with differences in phosphate and nitrogen uptake and fungal phosphate. *Metabolism*. 25, 533-546. doi: 10.1007/s00572-015-0631-x

Michel-Cuello, C., Juárez-Flores, B. I., Aguirre-Rivera, J. R., and Pinos-Rodríguez, J. M. (2008). Quantitative characterization of nonstructural carbohydrates of mezcal Agave (*Agave salmiana* Otto ex Salm-Dick). *J. Agric. Food Chem.* 56, 5753–5757. doi: 10.1021/jf800158p

Minnis, P. E., and Plog, S. E. (1976). A study of the site specific distribution of *Agave parryi* in east central Arizona. *Kiva* 41, 299–308. doi: 10.1080/00231940.1976.11757854

Minnis, P. E., Whalen, M. E., and Howell, R. E. (2006). Fields of power: upland farming in the prehispanic casas grandes polity, Chihuahua, Mexico. *Am. Antiq.* 71, 707–722. doi: 10.2307/40035885

Mitchell, W. P. (1973). The hydraulic hypothesis: a reappraisal. *Curr. Anthropol.* 14, 532–534. doi: 10.1086/201379

Moore, K. J., Karlen, D. L., and Lamkey, K. R. (2014). "Future prospects for corn as a biofuel crop", in *Compendium of Bioenergy Plants: Corn*, Chapter 13, eds S. L. Goldman and C. Kole, 5, 331–352.

Morán, E. (2008). Constructing identity: the role of food in mexica migration and creation accounts. *Lat. Am.* 52, 15–27 doi: 10.1111/j.1557-203X.2008.00003.x

Narváez-Zapata, J. A., and Sánchez-Teyer, L. F. (2010). Agaves as a raw material: recent technologies and applications. *Recent Pat. Biotechnol.* 3, 185–191. doi: 10.2174/187220809789389144

Neftel, A., Moor, E., Oeschger, H., and Stauffer, B. (1985). Evidence from polar ice cores for the increase in atmospheric CO2 in the past two centuries. *Nature* 315:45. doi: 10.1038/315045a0

Nobel, P. S. (1976). Water relations and photosynthesis of a desert CAM plant, *Agave deserti*. *Plant Physiol.* 58, 576–582. doi: 10.1104/pp.58.4.576

Nobel, P. S. (1977). Water relations of flowering of *Agave deserti*. *Bot. Gaz.* 138, 1–6. doi: 10.1086/336888

Nobel, P. S. (1991). Achievable productivities of certain CAM plants: basis for high values compared with C3 and C4 plants. *New Phytol.* 119, 183–205. doi: 10.1111/j.1469-8137.1991.tb01022.x

Nobel, P. S. (1994). *Remarkable Agaves and Cacti*. New York, NY: Oxford University Press.

Nobel, P. S. (2003). *Environmental Biology of Agaves and Cacti*. Cambridge, NY: Cambridge University Press.

Nobel, P. S., García-Moya, E., and Quero, E. (1992a). High annual productivity of certain agaves and cacti under cultivation. *Plant Cell Environ.* 15, 329–335. doi: 10.1111/j.1365-3040.1992.tb00981.x

Nobel, P. S., and Hartsock, T. L. (1986). Temperature, water, and PAR influences on predicted and measured productivity of *Agave deserti* at various elevations. *Oecologia* 68, 181–185. doi: 10.1007/BF00384785

Nobel, P. S., and Jordan, P. W. (1983). Transpiration stream of desert species: resistances and capacitances for a C3, a C4, and a CAM plant. *J. Exp. Bot.* 34, 1379–1391. doi: 10.1093/jxb/34.10.1379

Nobel, P. S., Miller, P. M., and Graham, E. A. (1992b). Influence of rocks on soil temperature, soil water potential, and rooting patterns for desert succulents. *Oecologia* 92, 90–96. doi: 10.1007/BF00317267

Nobel, P. S., and Quero, E. (1986). Environmental productivity indices for a Chihuahuan desert CAM plant, Agave lechuguilla. *Ecology* 67, 1–11. doi: 10.2307/1938497

Nobel, P. S., Quero, E., and Linares, H. (1988). Differential growth responses of agaves to nitrogen, phosphorus, potassium, and boron applications. *J. Plant Nutr.* 11, 1683–1700. doi: 10.1080/01904168809363925

Nobel, P. S., Quero, E., and Linares, H. (1989). Root versus shoot biomass: responses to water, nitrogen, and phosphorus applications for *Agave lechuguilla*. *Bot. Gaz.* 150, 411–416. doi: 10.1086/337787

Nobel, P. S., and Valenzuela, A. G. (1987). Environmental responses and productivity of the CAM plant, *Agave tequilana*. *Agric. For. Meteorol.* 39, 319–334. doi: 10.1016/0168-1923(87)90024-4

North, G. B., and Nobel, P. S. (1991). Changes in hydraulic conductivity and anatomy caused by drying and rewetting roots of *Agave deserti* (Agavaceae). *Am. J. Bot.* 78, 906–915. doi: 10.1002/j.1537-2197.1991.tb14494.x

Núñez, N., L., Salazar-Solano, V., and Acedo-Felix, E. (2008). *El Bacanora: Cultivo, Regulación y Mercados*. Hermosillo: Centro de Investigación en Alimentos y Desarrollo (CIAD).

O'Brien-Nabors, L. (2001). *Alternative Sweeteners, Third Edition, Revised and Expanded*. Boca Raton, FL: CRC Press.

Orians, G. H., and Solbrig, O. T. (1977). A cost-income model of leaves and roots with special reference to arid and semiarid areas. *Am. Nat.* 111, 677–690. doi: 10.1086/283199

Orue, A., Jauregi, A., Unsuain, U., Labidi, J., Eceiza, A., and Arbelaiz, A. (2016). The effect of alkaline and silane treatments on mechanical properties and breakage of sisal fibers and poly(lactic acid)/sisal fiber composites. *Compos. Part Appl. Sci. Manuf.* 84, 186–195. doi: 10.1016/j.compositesa.2016.01.021

Palta, J. A., and Nobel, P. S. (1989). Influence of soil O2 and CO2 on root respiration for *Agave deserti*. *Physiol. Plant.* 76, 187–192. doi: 10.1111/j.1399-3054.1989.tb05630.x

Parker, K. C., Hamrick, J. L., Hodgson, W. C., Trapnell, D. W., Parker, A. J., and Kuzoff, R. K. (2007). Genetic consequences of pre-Columbian cultivation for Agave murpheyi and A. delamateri (Agavaceae). *Am. J. Bot.* 94, 1479–1490. doi: 10.3732/ajb.94.9.1479

Parker, K. C., Trapnell, D. W., Hamrick, J. L., and Hodgson, W. C. (2014). Genetic and morphological contrasts between wild and anthropogenic populations of *Agave parryi* var. huachucensis in south-eastern Arizona. *Ann. Bot.* 113, 939–952. doi: 10.1093/aob/mcu016

Parker, K. C., Trapnell, D. W., Hamrick, J. L., Hodgson, W. C., and Parker, A. J. (2010). Inferring ancient agave cultivation practices from contemporary genetic patterns. *Mol. Ecol.* 19, 1622–1637. doi: 10.1111/j.1365-294X.2010.04593.x

Perry, L., and Flannery, K. V. (2007). Precolumbian use of chili peppers in the valley of Oaxaca, Mexico. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11905–11909. doi: 10.1073/pnas.0704936104

Pinos-Rodríguez, J. M., Aguirre-Rivera, J. R., García-López, J. C., Rivera-Miranda, M. T., González-Muñoz, S., López-Aguirre, S., et al. (2006). Use of "Maguey" (*Agave salmiana* Otto ex. Salm-Dick) as forage for ewes. *J. Appl. Anim. Res.* 30, 101–107. doi: 10.1080/09712119.2006.9706596

Pinos-Rodríguez, J. M., Zamudio, M., and González, S. S. (2008). The effect of plant age on the chemical composition of fresh and ensiled Agave salmiana leaves. *South Afr. J. Anim. Sci.* 38, 43–50. doi: 10.4314/sajas.v38i1.4108

Pinos-Rodríguez, J. M., Zamudio, M., González, S. S., Mendoza, G. D., Bárcena, R., Ortega, M. E., et al. (2009). Effects of maturity and ensiling of Agave salmiana on nutritional quality for lambs. *Anim. Feed Sci. Technol.* 152, 298–306. doi: 10.1016/j.anifeedsci.2009.05.002

Pongratz, J., and Caldeira, K. (2012). Attribution of atmospheric CO2 and temperature increases to regions: importance of preindustrial land use change. *Environ. Res. Lett.* 7:034001. doi: 10.1088/1748-9326/7/3/034001

Porter, J. R., and Semenov, M. A. (2005). Crop responses to climatic variation. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 2021–2035. doi: 10.1098/rstb.2005.1752

Reick, C., Raddatz, T., Pongratz, J., and Claussen, M. (2010). Contribution of anthropogenic land cover change emissions to pre-industrial atmospheric CO2. *Tellus B Chem. Phys. Meteorol.* 62, 329–336. doi: 10.1111/j.1600-0889.2010.00479.x

Revelle, R., and Suess, H. E. (1957). Carbon dioxide exchange between atmosphere and ocean and the question of an increase of atmospheric CO2 during the past decades. *Tellus* 9, 18–27. doi: 10.3402/tellusa.v9i1.9075

Rice, G. (1998). War and water: an ecological perspective on Hohokam irrigation. *Kiva* 63, 263–301. doi: 10.1080/00231940.1998.117 58357

Rivas, H. G. (1991). *Cocina Prehispánica Mexicana: la Comida de los Antiguos Mexicanos*. Mexico, D.F: Panorama Editorial.

SAGARPA (2017). *Planeación Agrícola Nacional 2017-2030, Agave tequilero y mezcalero Mexicano*. Available online at: https://www.gob.mx/cms/uploads/attachment/file/257066/Potencial-Agave_Tequilero_y_Mezcalero.pdf (accessed September 11, 2019).

Sandor, J. A., and Homburg, J. A. (2011). "Soil and landscape responses to American Indian agriculture in the Southwest," in *Movement, Connectivity, and Landscape Change in the Ancient Southwest: the 20th Anniversary Southwest Symposium*. (Boulder, CO: University Press of Colorado), 141–59.

Sandor, J. A., and Homburg, J. A. (2017). Anthropogenic soil change in ancient and traditional agricultural fields in arid to semiarid regions of the Americas. *J. Ethnobiol.* 37, 196–218. doi: 10.2993/0278-0771-37.2.196

Santos, J. D. G., and Branco, A. (2014). GC-MS characterisation of sapogenins from sisal waste and a method to isolate pure hecogenin. *BioResources* 9, 1325–1333. doi: 10.15376/biores.9.1.1325-1333

Schlaepfer, D. R., Bradford, J. B., Lauenroth, W. K., Munson, S. M., Tietjen, B., Hall, S. A., et al. (2017). Climate change reduces extent of temperate drylands and intensifies drought in deep soils. *Nat. Commun.* 8:14196. doi: 10.1038/ncomms14196

Silva, F. de A., Filho, R. D. T., Filho, J. de A. M., and Fairbairn, E. de M. R. (2010). Physical and mechanical properties of durable sisal fiber?cement composites. *Constr. Build. Mater.* 24, 777–785. doi: 10.1016/j.conbuildmat.2009.10.030

Silva-Montellano, A., and Eguiarte, L. E. (2003). Geographic patterns in the reproductive ecology of *Agave lechuguilla* (Agavaceae) in the Chihuahuan desert. I. Floral characteristics, visitors, and fecundity. *Am. J. Bot.* 90, 377–387. doi: 10.3732/ajb.90.3.377

Smith, M. N., Stark, S. C., Taylor, T. C., Ferreira, M. L., de Oliveira, E., Restrepo-Coupe, N., et al. (2019). Seasonal and drought-related changes in leaf area profiles depend on height and light environment in an Amazon forest. *New Phytol.* 222, 1284–1297. doi: 10.1111/nph.15726

Solomon, S., Rosenlof, K. H., Portmann, R. W., Daniel, J. S., Davis, S. M., Sanford, T. J., et al. (2010). Contributions of stratospheric water vapor to decadal changes in the rate of global warming. *Science* 327, 1219–1223. doi: 10.1126/science.1182488

Somerville, C. (2007). Biofuels. *Curr. Biol.* 17, R115–R119. doi: 10.1016/j.cub.2007.01.010

Somerville, C., Youngs, H., Taylor, C., Davis, S. C., and Long, S. P. (2010). Feedstocks for lignocellulosic biofuels. *Science* 329, 790–792. doi: 10.1126/science.1189268

Soto, J. L. M., González, J. V., Nicanor, A. B., and Ramírez, E. G. R. (2011). Enzymatic production of high fructose syrup from *Agave tequilana* fructans and its physicochemical characterization. *Afr. J. Biotechnol.* 10, 19137–19143. doi: 10.5897/AJB11.2704

Sridhar, V., Loope, D. B., Swinehart, J. B., Mason, J. A., Oglesby, R. J., and Rowe, C. M. (2006). Large wind shift on the Great Plains during the Medieval Warm Period. *Science* 313, 345–347. doi: 10.1126/science.1128941

Stager, L. E. (1976). Farming in the Judean desert during the Iron Age. *Bull. Am. Sch. Orient. Res.* 221, 145–158. doi: 10.2307/1356097

Stevenson, C. M., Wozniak, J., and Haoa, S. (1999). Prehistoric agricultural production on Easter Island (Rapa Nui), Chile. *Antiquity* 73, 801–812. doi: 10.1017/S0003598X00065546

Stewart, J. R. (2015). Agave as a model CAM crop system for a warming and drying world. *Front. Plant Sci.* 6:684. doi: 10.3389/fpls.2015.00684

Stinchcomb, G. E., Messner, T. C., Driese, S. G., Nordt, L. C., and Stewart, R. M. (2011). Pre-colonial (AD 1100–1600) sedimentation related to prehistoric maize agriculture and climate change in eastern North America. *Geology* 39, 363–366. doi: 10.1130/G31596.1

Tequila Regulatory Council (2019). *Estadísticas de Producción Total: Tequila y Tequila 100%, Consumo de agave para Tequila y Tequila 100% de agave*. Available online at: https://www.crt.org.mx (accessed November 15, 2019).

Towell, J. L., and Lecón, A. A. (2010). *Caminos y Mercados de México*. Mexico D.F: Universidad Nacional Autónoma de México/Instituto Nacional de Antropología e Historia.

Trombold, C. D. (2017). Agricultural Intensification on the epiClassic northern Mesoamerican frontier: the La Quemada Terraces. *J. Anthropol. Archaeol.* 48, 309–319. doi: 10.1016/j.jaa.2017.09.005

Troyo-Diéguez, E., de Lachica-Bonilla, F., and Fernàndez-Zayas, J. L. (1990). A simple aridity equation for agricultural purposes in marginal zones. *J. Arid Environ.* 19, 353–362. doi: 10.1016/S0140-1963(18)30799-7

Turner, B. L. (1976). Population density in the classic Maya lowlands: new evidence for old approaches. *Geogr. Rev.* 66, 73–82. doi: 10.2307/213316

Urias-Silvas, J. E., Cani, P. D., Delmée, E., Neyrinck, A., López, M. G., and Delzenne, N. M. (2008). Physiological effects of dietary fructans extracted from *Agave tequilana* Gto. and *Dasylirion* spp. *Br. J. Nutr.* 99, 254–261. doi: 10.1017/S0007114507795338

Valenzuela, Z., A. G., and Gonzalez, E., D. R. (1995). *Fertilization of (Agave tequilana Weber) of Tequila*. Jalisco: TERRA Mex.

Van West, C. R., and Dean, J. S. (2000). Environmental characteristics of the AD 900–1300 period in the Central Mesa Verde region. *Kiva* 66, 19–44. doi: 10.1080/00231940.2000.11758420

Vargas-Ponce, O., Zizumbo-Villarreal, D., and Marin, P. C.-G. (2007). *In situ* diversity and maintanenace of traditional Agave landraces used in spirits production in West-Central Mexico. *Econ. Bot.* 61:362. doi: 10.1663/0013-0001(2007)61[362:ISDAMO]2.0.CO;2

Walton, M. K. (1977). The evolution and localization of mezcal and tequila in Mexico. *Rev. Geográfica* 85, 113–132.

Webb, E. A., Schwarcz, H. P., and Healy, P. F. (2004). Detection of ancient maize in lowland Maya soils using stable carbon isotopes: evidence from Caracol, Belize. *J. Archaeol. Sci.* 31, 1039–1052. doi: 10.1016/j.jas.2004.01.001

Wilken, G. C. (1972). Microclimate management by traditional farmers. *Geogr. Rev.* 62, 544–560. doi: 10.2307/213267

Williams, D. E. (1990). A review of sources for the study of Náhuatl plant classification. *Adv. Econ. Bot.* 8, 249–270.

Woodbury, R. B. (1961). A reappraisal of Hohokam irrigation. *Am. Anthropol.* 63, 550–560. doi: 10.1525/aa.1961.63.3.02a00070

Woodhouse, C. A., Meko, D. M., MacDonald, G. M., Stahle, D. W., and Cook, E. R. (2010). A 1,200-year perspective of 21st century drought in southwestern North America. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21283–21288. doi: 10.1073/pnas.0911197107

Woodhouse, R. M., Williams, J. G., and Nobel, P. S. (1980). Leaf orientation, radiation interception, and nocturnal acidity Increases by the CAM plant *Agave Deserti* (Agavaceae). *Am. J. Bot.* 67, 1179–1185. doi: 10.1002/j.1537-2197.1980.tb07751.x

Wozniak, J. A. (1999). Prehistoric horticultural practices on Easter Island: lithic mulched gardens and field systems. *Rapa Nui J.* 13, 95–99.

Yan, X., Tan, D. K., Inderwildi, O. R., Smith, J. A. C., and King, D. A. (2011). Life cycle energy and greenhouse gas analysis for agave-derived bioethanol. *Energy Environ. Sci.* 4, 3110–3121. doi: 10.1039/c1ee01107c

Yang, X., Cushman, J. C., Borland, A. M., Edwards, E. J., Wullschleger, S. D., Tuskan, G. A., et al. (2016). A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytol.* 207, 491–504.

Yilmaz, N. D., and Arifuzzaman Khan, G. M. (2019). "2 - Flexural behavior of textile-reinforced polymer composites," in *Mechanical and Physical Testing of Biocomposites, Fibre-Reinforced Composites and Hybrid Composites Woodhead Publishing Series in Composites Science and Engineering*. eds. M. Jawaid, M. Thariq, and N. Saba (Woodhead Publishing), 13–42. doi: 10.1016/B978-0-08-102292-4.00002-3

Ynnilä, H. (2007). *To Petra via Jabal Haroun: Nabataean-Roman road remains in the Finnish Jabal Haroun project survey area*. MA thesis. University of Helsinki. Available online at: https://helda.helsinki.fi/handle/10138/19544 (accessed August 30, 2020).

Zandalinas, S. I., Mittler, R., Balfagón, D., Arbona, V., and Gómez-Cadenas, A. (2018). Plant adaptations to the combination of drought and high temperatures. *Physiol. Plant.* 162, 2–12. doi: 10.1111/ppl.12540

Zhang, X., Song, K., Tong, F., Fei, M., Guo, H., Lu, Z., et al. (2020). First case of COVID-19 in a patient with multiple myeloma successfully treated with tocilizumab. *Blood Adv.* 4, 1307–1310. doi: 10.1182/bloodadvances.2020001907

Zizumbo-Villarreal, D., Flores-Silva, A., and Marín, P. C.-G. (2012). The archaic diet in Mesoamerica: incentive for milpa development and species domestication. *Econ. Bot.* 66, 328–343. doi: 10.1007/s12231-012-9212-5

Zizumbo-Villarreal, D., González-Zozaya, F., Olay-Barrientos, A., Platas-Ruíz, R., Cuevas-Sagardí, M., Almendros-López, L., et al. (2009). Archaeological

evidence of the cultural importance of *Agave spp.* in pre-Hispanic Colima, Mexico. *Econ. Bot.* 63, 288–302. doi: 10.1007/s12231-009-9092-5

Zizumbo-Villarreal, D., Vargas-Ponce, O., Rosales-Adame, J. J., and Colunga-García Marín, P. (2013). Sustainability of the traditional management of Agave genetic resources in the elaboration of mezcal and tequila spirits in western Mexico. *Genet. Resour. Crop Evol.* 60, 33–47. doi: 10.1007/s10722-012-9812-z

Zúñiga-Estrada, L., Rosales, R., E., Yáñez-Morales, M., de, J., and Jacques-Hernández, C. (2018). Características y productividad de una planta MAC, *Agave tequilana* desarrollada con fertigación en Tamaulipas, México. *Rev. Mex. Cienc. Agríc.* 9, 553–564. doi: 10.29312/remexca.v9i3.1214

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 09 November 2020
doi: 10.3389/fpls.2020.536135

Phylogeny, Diversification Rate, and Divergence Time of *Agave sensu lato* (Asparagaceae), a Group of Recent Origin in the Process of Diversification

*Ofelia Jiménez-Barron[1], Ricardo García-Sandoval[2], Susana Magallón[3], Abisaí García-Mendoza[4], Jorge Nieto-Sotelo[4], Erika Aguirre-Planter[1] and Luis E. Eguiarte[1]**

[1] Laboratorio de Evolución Molecular y Experimental, Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad Universitaria, Mexico City, Mexico, [2] Departamento de Biología, Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad Universitaria, Mexico City, Mexico, [3] Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad Universitaria, Mexico City, Mexico, [4] Jardín Botánico, Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad Universitaria, Mexico City, Mexico

OPEN ACCESS

Edited by:
Charles Bell,
University of New Orleans,
United States

Reviewed by:
Robin Van Velzen,
Wageningen University and Research,
Netherlands
Matthew T. Lavin,
Montana State University,
United States

***Correspondence:**
Luis E. Eguiarte
fruns@unam.mx

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: 18 February 2020
Accepted: 06 October 2020
Published: 09 November 2020

Citation:
Jiménez-Barron O, García-Sandoval R, Magallón S, García-Mendoza A, Nieto-Sotelo J, Aguirre-Planter E and Eguiarte LE (2020) Phylogeny, Diversification Rate, and Divergence Time of Agave sensu lato (Asparagaceae), a Group of Recent Origin in the Process of Diversification. Front. Plant Sci. 11:536135. doi: 10.3389/fpls.2020.536135

Agave sensu lato is one of the most diverse and complex genera of Asparagaceae, with more than 250 species. The morphological, ecological, and evolutionary diversity of the group has complicated its taxonomical study. We conducted phylogenetic analyses of DNA sequence data to reconstruct the phylogenetic relationships of the *Agave* genus. We included 107 species of the Asparagaceae family from which 83 correspond to the *Agave sensu lato* clade (*Agave sensu stricto* + *Polianthes* + *Manfreda* and *Prochnyanthes*, which together represent 30% of the genus) and as outgroups the genera *Dasylirion*, *Hesperoyucca*, *Chlorogalum*, *Camassia*, *Hesperaloe*, *Yucca*, *Beschorneria*, and *Furcraea*, in order to estimate the age and propose the history of their diversification. Previous studies postulated the relevance of the Miocene in the speciation rates of the agaves, as well as the relevance of the type of inflorescence in its diversification. However, these assertions have not been well supported. The analysis of chloroplast regions resulted in low resolution, which could be the consequence of the few variable sites. On the other hand, the internal transcribed spacer (ITS) implemented in our analysis ensued in higher resolution and better support values. Our phylogenetic analyses recovered five groups; one is the Striatae group, which is the sister group to *Agave sensu stricto* clade. Within this clade, we found three main groups with high support; these groups are not related with previous morphological proposals. We also analyzed the dates of origin and diversification rates. A Bayesian analysis of macroevolutionary mixtures indicated two significant shifts; the first was identified at 6.18 Ma, where the speciation rate increased to 4.10 species/Mya, this shift occurred during the late Miocene period, characterized by the emergence of arid biomes in North America. The second was identified at a stem age of 2.68 Ma

where the speciation rate increased to 6.04 species/Mya. Concerning the ancestral reconstruction state of the inflorescence type in the *Agave sensu stricto* clade, the spike inflorescence character was predominant in the early-diverging groups, whereas the late-diverging groups present panicle inflorescences as the predominant character and higher speciation rates.

Keywords: evolutionary radiation, comparative method, extinction, speciation, inflorescence, ancestral state reconstruction, Bayesian inference, ITS

INTRODUCTION

The process of evolutionary radiation has been considered as one of the most important sources of biological diversity, through relatively rapid differentiation from a single ancestor into new species that inhabit a variety of environments and may differ in the characters they use to exploit them (Schluter, 1996, 2000; Olson and Arroyo-Santos, 2009). An evolutionary radiation is a complex process that may involve phenotypic and physiological differentiation, adaptation, speciation, and extinction (Schluter, 2001; Coyne and Orr, 2004; De Queiroz, 2007; Givnish, 2010; Nosil, 2012).

The factors that influence the net species diversification rate–that is to say, the net result of speciation and extinction for each taxon–are multiple and complex (Scott and Arnold, 1995; Magallón and Sanderson, 2001; Magallón and Castillo, 2009; Arakaki et al., 2011; Schlumpberger, 2012; Schlumpberger and Renner, 2012; Van der Niet and Johnson, 2012; Hernández-Hernández et al., 2014). These include extrinsic factors, such as physical space, climate, other organisms, or available habitats, or intrinsic factors, such as morphological or physiological traits, characters that affect the body's adequacy through its growth, survival, and/or reproduction (Glor, 2010; Losos, 2010; Bouchenak-Khelladi et al, 2015). The comparative study of lineages suspected of having radiation events with those that apparently have not experienced radiation events may help to identify the factors that influenced such changes in their diversification rates.

One of the challenges when studying the diversification of a group is understanding the factors influencing speciation and extinction rates, since the possible factors are many and it is difficult to disentangle which one or if several factors are affecting diversification (Zamora-Abrego et al., 2013). The comparative method allows us to make formal and statistical comparisons between species and thus analyze information, such as morphological, physiological, or ecological characters, by incorporating information on the phylogenetic relationships of the group of interest (Harvey and Pagel, 1991; Garamszegi, 2014). In this way, the comparative method has allowed the evaluation of the extent at which variation of a character is due to its evolutionary history or to adaptive pressures (Morales, 2000; Rezende and Garland, 2003).

Two diversification events have been proposed for the *Agave sensu lato* clade since its origin in the Miocene 7.9–10.2 Ma (Good-Avila et al., 2006; Scheinvar et al., 2017). The first occurred ∼6–8 Ma and correlates with the emergence of arid biomes in Mexico. The second was ∼3–2.5 Ma, and it was apparently associated to the evolution of reproductive characters and pollination syndromes (Eguiarte et al., 2000; Good-Avila et al., 2006; Rocha et al., 2006). The genus *Agave* has experienced constant taxonomic revisions, in part due to their high morphological variation and species diversity (Gentry, 1982; Álvarez de Zayas, 1995; García-Mendoza, 1995, 2002; Hernández-Sandoval, 1995; García-Mendoza et al., 2019) in addition to their low molecular variation, which suggests rapid diversification immediately following the origin of the group (Bogler and Simpson, 1995, 1996; Eguiarte et al., 2000; Bogler et al., 2006; Good-Avila et al., 2006; Smith et al., 2008; Archibald et al., 2015; Heyduk et al., 2016; McKain et al., 2016; Flores-Abreu et al., 2019). Previous studies have failed to recover *Agave* as a monophyletic genus, since it usually nests within the clade *Agave sensu lato*, along with *Manfreda*, *Polianthes*, and *Prochnyanthes* (Bogler and Simpson, 1995, 1996; Eguiarte et al., 2000; Bogler et al., 2006; Good-Avila et al., 2006; Flores-Abreu et al., 2019).

As mentioned above, the second diversification of *Agave sensu lato* is considered to be the result of the pressures imposed by pollinators (Good-Avila et al., 2006; Rocha et al., 2006; Flores-Abreu et al., 2019). It has been shown that pollinators influence the reproductive and phenological traits of various groups of plants, for instance selecting for synchronization of the flowering time of individuals of a given plant species (Percival and Morgan, 1965; Van der Niet and Johnson, 2012; Lagomarsino et al., 2016). Gentry (1982) divided the agaves based on their type of inflorescence into two subgenera, following Berger (1915), *Agave* subgenus, with paniculate inflorescences and *Littaea* subgenus, with spike inflorescences. Chiropterophily syndrome was attributed to species presenting paniculate inflorescences, whereas species presenting spike inflorescences were considered to be pollinated exclusively by insects (Schaffer and Schaffer, 1977). However, subsequent studies showed that, regardless of the shape of the inflorescence, the pool of *Agave* pollinators can be broad (Arizaga et al., 2000; Silva-Montellano and Eguiarte, 2003; Rocha et al., 2005, 2006; Trejo-Salazar et al., 2015). One of the proposed scenarios to explain the evolutionary history of the *Agave* pollination syndrome is that they evolved from a group of Asparagaceae with moth pollination, which later specialized in a bat pollination syndrome (Smith et al., 2008; McKain et al., 2016). It is likely that, in the specific case of the agaves, they went from being pollinated by insects to become specialized for bat pollination, as bats might have exerted pressure on the agaves and selected for larger inflorescences and individuals producing a greater amount of nectar (Schaffer and Schaffer, 1977, 1979; Eguiarte et al., 2000; Rocha et al., 2005). Coevolution studies

report a similar time of origin for both Asparagaceae and the Phyllostomidae (Chiroptera) family, to which the genus *Leptonycteris* belongs, that is the primary pollinator of many *Agave* species (Flores-Abreu et al., 2019).

The aim of the present study is to carry out a phylogenetic reconstruction of the *Agave* genus, by substantially increasing the taxonomic sampling and selecting the appropriate molecular markers, relative to previous studies, in order to obtain a higher resolution level and support values. Our taxonomic sample included 83 species of *Agave sensu lato* (*Agave sensu stricto* clade + *Manfreda* + *Polianthes* and *Prochnyanthes*); from this, 74 species correspond to *Agave sensu stricto*, this sampling includes at least one member of each of the morphological groups proposed by Gentry (1982). **Table 1** describes the subgenus and group to which each species analyzed in this study belongs. For this, two types of markers were used: chloroplast and nuclear, the latter had only been used at the intergeneric level or in the limited number of species within *Agave* (Bogler and Simpson, 1996; Lledías et al., 2020). Furthermore, Bayesian approaches were used to estimate the time of divergence of the main groups that conform *Agave sensu lato* and to reconstruct the ancestral character states for the type of inflorescence in order to trace the evolutionary history of this character and to assess its potential importance in the diversification of the group.

MATERIALS AND METHODS

Taxon Sampling, DNA Isolation, and Amplification

A total of 107 species of the Asparagaceae family were sampled for the phylogenetic analysis from which 83 correspond to the *Agave sensu lato* clade (74 *Agave sensu stricto* + 4 *Polianthes* + 4 *Manfreda* and 1 *Prochnyanthes*, which together represent 30% of the species that conform the genus) and as outgroups the genera *Dasylirion* (3), *Hesperoyucca* (1), *Chlorogalum* (2), *Camassia* (2), *Hesperaloe* (3), *Yucca* (5), *Beschorneria* (4), and *Furcraea* (4).

The concatenated matrix used for the chloroplast phylogeny included sequences of mat*K* (1,436 bp), *rps*16 (835 bp), *trn*H-*psb*A (561 bp), and *rpl*32-*trn*L (838 bp) from 43 *Agave sensu stricto* species + 2 *Manfreda* + 1 *Polianthes* and *Furcraea* (2) + *Beschorneria* (1) and *Yucca* (1) species as outgroups. For the internal transcribed spacer (ITS) data set, we analyzed 577 bp from 72 *Agave sensu stricto* species + 4 *Manfreda* + 4 *Polianthes* + 1 *Prochnyanthes* and *Dasylirion* (3), *Hesperoyucca* (1), *Chlorogalum* (2), *Camassia* (2), *Hesperaloe* (3), *Yucca* (5), *Beschorneria* (4), and *Furcraea* (4) as outgroups. All newly generated nucleotide sequences for this study were deposited in the NCBI GenBank.

Total genomic DNA was isolated from silica-dried leaf materials and herbarium specimens, using a modification of the CTAB method (Doyle and Doyle, 1987). We used polymerase chain reactions (PCR) to amplify five gene regions, including four plastid DNA regions: mat*K*, *rps*16, *trn*H-*psb*A (Shaw et al., 2005), *rpl*32-*trn*L (Shaw et al., 2007), and the nuclear *ITS*1–*ITS*2 region (Bogler and Simpson, 1996). Amplified products were purified and sequenced by Macrogen, United States, and

TABLE 1 | List of taxa included in the phylogenetic inference analyses.

Subgenus *Littaea*		
Littaea group	**No. species sampled**	**Sampled species names**
Amolae (8)	4	*attenuata*, *nizandensis*, *ocahui*, *vilmoriniana*
Marginatae (27)	16	*convallis*, *chazaroi*, *doctorensis*, *ghiesbreghtii*, *glomeruliflora*, *horrida*, *kerchovei*, *lechuguilla*, *montium*, *peacockii*, *pelona*, *pintilla*, *titanota*, *triangularis*, *univittata*, *victoria-reginae*
Filiferae (7)	3	*felgeri*, *multifilifera*, *schidigera*
Parviflorae (4)	3	*parviflora*, *polianthiflora*, *schottii*
Polycephalae (5)	2	*chiapensis*, *pendula*
Urceolatae (2)	1	*arizonica*
Striatae (5)	4	*dasylirioides*, *striata*, *rzedowskiana*, *petrophila*
Choripetalae (3)	3	*bracteosa*, *ellemeetiana*, *guiengola*
Total 61	36	59% of total
Subgenus *Agave*		
Agave group	**No. species sampled**	**Sampled species names**
Americanae (6)	4	*americana*, *lurida*, *scabra* (*asperrima*), *scaposa*
Antillanae (14)	1	*antillarum*
Campaniflorae (3)	2	*aurea*, *capensis*
Ditepalae (12)	5	*applanata*, *colorata*, *delamateri*, *phillipsiana*, *wocomahi*
Hiemiflorae (13)	4	*atrovirens*, *isthmensis*, *potatorum*, *seemanniana*
Crenatae (7)	3	*cupreata*, *inaequidens*, *maximiliana*
Deserticolae (10)	4	*cerulata*, *deserti*, *mckelveyana*, *sobria*
Marmoratae (4)	3	*grijalvensis*, *marmorata*, *zebra*
Parryanae (8)	3	*gentryi*, *ovatifolia*, *parryi*
Rigidae (13)	4	*angustifolia*, *datylio*, *rhodacantha*, *tequilana*
Salmianae (5)	1	*salmiana*
Sisalanae (5)	2	*desmettiana*, *sisalana*
Umbeliflorae (2)	1	*shawii*
Total 102	37	36% of total
Agave sensu lato	**No. species sampled**	**Sampled species names**
Manfreda (37)	4	*scabra*, *virginica*, *umbrophila*, *hauniensis*
Polianthes (14)	4	*bicolor*, *densiflora*, *geminiflora*, *longiflora*
Prochnyanthes (1)	1	*mexicana*
Total 52	9	22% of total
Outgroup	**No. species sampled**	**Sampled species names**
Dasylirion	3	*wheeleri*, *texanum*, *longissimum*
Chlorogalum	2	*parviflorum*, *purpureum*
Hesperoyucca	1	*whipplei*
Hesperaloe	3	*changii*, *parviflora*, *nocturna*

(Continued)

TABLE 1 | Continued

Outgroup	No. species sampled	Sampled species names
Yucca	5	*filifera*, *thompsoniana*, *linearifolia*, *madrensis*, *brevifolia*
Camassia	2	*quamash*, *leichtlinii*
Furcraea	4	*longaeva*, *martinezii*, *guatemalensis*, *pubescens*
Beschorneria	4	*calcicola*, *rigida*, *albiflora*, *yuccoides*
Total	24	

We followed the infrageneric classifications for Agave proposed by Gentry (1982) and Álvarez de Zayas (1995). The numbers between parentheses represent the number of currently known species that conform each one of the subgeneric groups analyzed. Prochnyanthes, Polianthes, and Manfreda species numbers are in agreement with Solano and Feria (2007),Castro-Castro et al. (2010), and Castro-Castro et al. (2018), respectively.

the complementary chains were visualized and assembled using the DNA Baser version 2.9.97 program (HeracleSoftware). The resulting sequences were aligned with MAFFT (Katoh and Standley, 2013), followed by manual adjustment in PhyDE (Müller et al., 2006). The accession numbers of the sequences obtained in this study and the ones downloaded from the GenBank data base are available as **Supplementary Table 1**.

Phylogenetic Reconstructions

The best maximum likelihood (ML) tree for the concatenated matrix from plastid regions was constructed using RAxML (Stamatakis, 2014). We conducted an exhaustive search using PartitionFinder2 (Lanfear et al., 2016) to select the appropriate partitioning scheme for our chloroplast matrix. We provided PartitionFinder2 with subsets for each region, and for the two coding regions, we provided subsets for each nucleotide position. Under the Bayesian Information Criterion (BIC), the "greedy" algorithm, and models = all, PartitionFinder2 identified two partitions: the first partition corresponded to the intron *rps16*, and its best substitution model was the GTR + I model. The second partition included the intergenic spacers *matK* + *trnH-psbA* + *rpl32-trnL*, and the best substitution model was the TrN + I model. The best substitution model for each partition was corroborated in jModelTest (Darriba et al., 2012) under a BIC and then used in the phylogenetic analyses. The analyses were run for 10,000 generations with 1,000 bootstrap replicates, and *Yucca filifera* was specified as outgroup.

For the ITS matrix, the best model was selected based on BIC implemented in jModelTest (Darriba et al., 2012). We ran the ML analyses implementing a GTR + G model for 10,000 generations with 1,000 bootstrap replicates and specifying *Dasylirion* clade as the outgroup.

A Bayesian phylogenetic tree was reconstructed using MrBayes 3.2.2 (Ronquist and Huelsenbeck, 2003). The best substitution model for each partition set was selected using a reversible-jump strategy (Huelsenbeck et al., 2003), and rate heterogeny was modeled with a gamma distribution (Huelsenbeck and Rannala, 2004). Two independent runs with four chains (three heated and one cold) were conducted concurrently for 20,000,000 generations and sampling every 1,000 generations. When the estimated sample size (ESS) value exceeded 200 and the potential scale reduction factor (PSRF) was close to 1.0, it was considered that convergence of the chains occurred. The 25% samples were discarded as burn-in.

Estimation of Divergence Times and Ancestral State Reconstruction

Bayesian age estimation for the divergence of internal nodes was conducted under an "uncorrelated relaxed clock" model with a lognormal distribution and the tree Birth–Death model in BEAST v2 (Bouckaert et al., 2014). The root node was calibrated under a lognormal distribution, with a mean of 62.49 Ma, which corresponds to the age of the order Asparagales estimated by Magallón et al. (2015). A second point of calibration was the stem age of *Yucca*, with a lognormal distribution and a mean of 14.2 Ma; this includes the age of the strata of the fossil *Protoyucca shadishii* from the middle Miocene, which is considered as being closely related to the *Yucca* genus (Tidwell and Parker, 1990; Wikstrom et al., 2001), and corresponds with previous molecular estimates for the divergence of the *Yucca* clade (Good-Avila et al., 2006; McKain et al., 2016; Flores-Abreu et al., 2019). The analysis was run for 200,000,000 generations sampling every 20,000 from which 25% was discarded as burn-in. The molecular clock analyses were conducted in the CIPRES Science Gateway (Miller et al., 2010). Log outputs of the BEAST analyses were evaluated with tracer v1.5 (Rambaut et al., 2018). Files containing the sampled trees of each MCMC run were combined using LogCombiner v1.7.5, annotated using TreeAnnotator v1.7.5 (Helfrich et al., 2018), and visualized using FigTree v1.4.0.

Inference of the ancestral states for the inflorescence type in the species included in the analysis was based on descriptions and morphological studies (Gentry, 1982; Carrillo-Reyes et al., 2003). The discrete trait inflorescence was coded as a binary character, in which spike inflorescence = 0 and panicle inflorescence = 1. Reconstruction was based on our Bayesian posterior random sample of 500 post burn-in topologies obtained with BEAST v2.0. The ancestral inflorescence type of key nodes from the *Agave sensu lato* clade was reconstructed using the BayesMultistate model as implemented in BayesTraits 3.0.2 (Pagel et al., 2004). Initially, a ML analysis was run to derive empirical priors. After setting these priors (uniform distribution 0–10), a Bayesian inference (BI) analysis was performed using a reversible-jump Markov Chain Monte Carlo (rjMCMC) for 5 million generations, sampling every 10,000 generations and discarding the first 25% as burn-in. The convergence of the chains was verified in trace plots and ESS values. The results of BayesTraits were processed using the same script as in Harrington and Reeder (2017), in which we can graph the probability of each character state for that node and the probability of no node existence.

Diversification Rate Analyses

To analyze the diversification among agaves, we used a Bayesian analysis of macroevolutionary mixtures (BAMM) v2.5.0 software (Rabosky et al., 2014) for (R Studio Team, 2020). Priors were obtained with BAMMtools by providing the BEAST maximum

clade credibility tree and total species number across the *Agave sensu lato* clade. It is well-known that incomplete taxon sampling can bias analyses of speciation and extinction from phylogenetic trees. BAMM accounts for incomplete sampling by analyzing the proportion of tips sampled for a given clade under the assumption that species are missing at random from the tree; species number was obtained from published sources.

Diversification rates were inferred using the function "speciation–extinction" of BAMM, which allows detecting rate shifts (assumed a compound Poisson process in the phylogeny) along tree branches. The evolutionary rate parameters used were: expected number of shift = 1.0, lamdaIntPrior = 1.0, lambdaShiftPrior = 0.05, and muInitPrior = 1.0. BAMM uses rjMCMC to explore the distinct evolutionary models that best explain the whole diversification of the clade. The analysis was conducted by concurrently running two independent chains for 20,000,000 generations and assuming convergence of the chains when the ESS value exceeded 200. For diversification analyses, we retrieved the configuration of rate shifts with the highest posterior probability through the "getBestShiftConfiguration" function of BAMMtools. These configurations were depicted as phylorate plots, which represent the analyzed phylogeny with its branches colored to reflect the instantaneous diversification rate. Rates-through-time plots were generated for speciation (λ), extinction (μ), and diversification (r) for both *Agave sensu lato* clade and other groups identified as having significant rate shifts in speciation. We used the functions getCladeRates to obtain estimates of the speciation rate (λ) and an extinction rate (μ) for a specific clade.

RESULTS

Phylogenetic Analysis

For the chloroplast phylogeny, a total of 3,670 bp from 46 *Agave sensu lato* species were analyzed. The data set contained a total of 29 variable sites, from which 19 were informative. The BI and ML reconstruction were congruent and recovered some of the morphological delimited genera. The *Furcraea* and *Beschorneria* species included in the analyses were grouped together. The *Agave sensu lato* group included *Manfreda*, *Polianthes*, and *Agave sensu stricto* and appeared as monophyletic (0.87 PP/95.3% BS), whereas *Agave sensu stricto* was not monophyletic. ML and BI trees are shown in **Supplementary Figures S1**, **S2**.

For the ITS data set, we analyzed 577 bp for the total 105 species described in "Materials and Methods" section, and the matrix contained a total of 168 variable sites of which 155 were informative. Both analyses (ML and BI) resulted in congruent topologies (**Figures 1A,B**). BI best resolved the earliest-diverging clades: one composed by the *Hesperaloe*/*Hesperoyucca* + *Chlorogalum*/*Camassia* + *Yucca* groups and a second one formed by the *Furcraea*/*Beschorneria* + *Agave sensu lato* groups. However, the *Hesperaloe*/*Hesperoyucca*, *Chlorogalum*/*Camassia*, and *Yucca* groups were not resolved. The ML tree was also unresolved for the *Hesperaloe*/*Hesperoyucca*, *Chlorogalum*/*Camassia*, and *Yucca* + *Furcraea*/*Beschorneria* + *Agave sensu lato* clades. Nonetheless, in both analyses, BI and ML trees were in agreement that the *Yucca* group is independent to the lineage leading to the *Furcraea*/*Beschorneria* + *Agave sensu lato* groups. The *Furcraea–Beschorneria* clade (0.95 PP/100% BS) came out as a sister group to *Agave sensu lato* in both analyses (0.99 PP/99.8% BS). However, the *Agave sensu stricto* group is paraphyletic with respect to *Manfreda*, *Prochnyanthes* plus *Polianthes*, that together constitute a clade with high support value (1 PP/100% BS) (**Figures 1A,B**).

Two species of the *Agave sensu lato* clade, *Agave ellemeetiana* (1 PP/98.4 BS) and *Agave bracteosa* (0.99 PP/98.4 BS) belonging to the Choripetalae group, consistently emerged early in the evolution of the group using either BI or ML methods forming a paraphyletic grade (**Figures 1A,B**).

The clade that conformed by *Agave dasylirioides*, *Agave striata*, *Agave rzedowskiana*, and *Agave petrophila* was well supported by our analyses (1.0 PP/100% BS) and was clearly separated from the clade that we will name here as *Agave sensu stricto* and from Group II containing *Manfreda*, *Polianthes*, and *Prochnyanthes* (1.0 PP/100 BS%) (**Figures 1A,B**).

Inside the *Agave sensu stricto* clade (1.0 PP/98.4% BS), *Agave pelona* (1.0 PP/98.4% BS) was positioned as the sister of the *Agave sensu stricto* clade (**Figures 1A,B**). *Agave sensu stricto* was conformed in our analyses by three distinctive groups: III, IV, and V (**Figures 1A,B**).

Group III (1 PP/86.4% BS) comprises *Agave parryi*, *Agave arizonica*, *Agave felgeri*, *Agave multifilifera*, *Agave parviflora*, *Agave schidigera*, *Agave polianthiflora*, and *Agave schottii*.

Group IV (0.85 PP/100% BS) included *Agave angustifolia* and a subgroup formed by *Agave antillarum*, *Agave attenuata*, *Agave chiapensis*, *Agave colorata*, *Agave convallis*, *Agave ghiesbreghtii*, *Agave guiengola*, *Agave horrida*, *Agave kerchovei*, *Agave marmorata*, *Agave pendula*, *Agave titanota*, and *Agave triangularis*.

Group V (0.99 PP/89.2% BS) was composed of *Agave americana*, *Agave datylio*, *Agave atrovirens*, *Agave aurea*, *Agave capensis*, *Agave cerulata*, *Agave cupreata*, *Agave delamateri*, *Agave deserti*, *Agave desmettiana*, *Agave maximiliana*, *Agave mckelveyana*, *Agave phillipsiana*, *Agave potatorum*, *Agave scaposa*, *Agave seemanniana*, *Agave shawii*, *Agave sobria*, *Agave vilmoriniana*, *Agave wocomahi*, *Agave zebra*, *Agave grijalvensis*, *Agave isthmensis*, *Agave lurida*, *Agave rhodacantha*, *Agave sisalana*, and *Agave tequilana* (**Figures 1A,B**).

Other agave species that also belong to *Agave sensu stricto* were not part of the above-mentioned strongly supported clades and were found paraphyletic to these clades, both by BI and ML (**Figures 1A,B**), including *Agave applanata*, *Agave asperrima*, *Agave doctorensis*, *Agave lechuguilla*, *Agave montium*, *Agave nizandensis*, *Agave ocahui*, *Agave peacockii*, *Agave salmiana*, *Agave univittata*, *Agave victoria-reginae*, *Agave gentryi*, *Agave ovatifolia*, *Agave chazaroi*, *Agave glomeruliflora*, and *Agave inaequidens*. The only discrepancy was *A. ocahui* that according to ML is a Group V member, but not in BI.

Divergence Times

Our analyses based on ITS sequences estimated the divergence of the *Yucca* group from the *Hesperaloe*/*Hesperoyucca* and

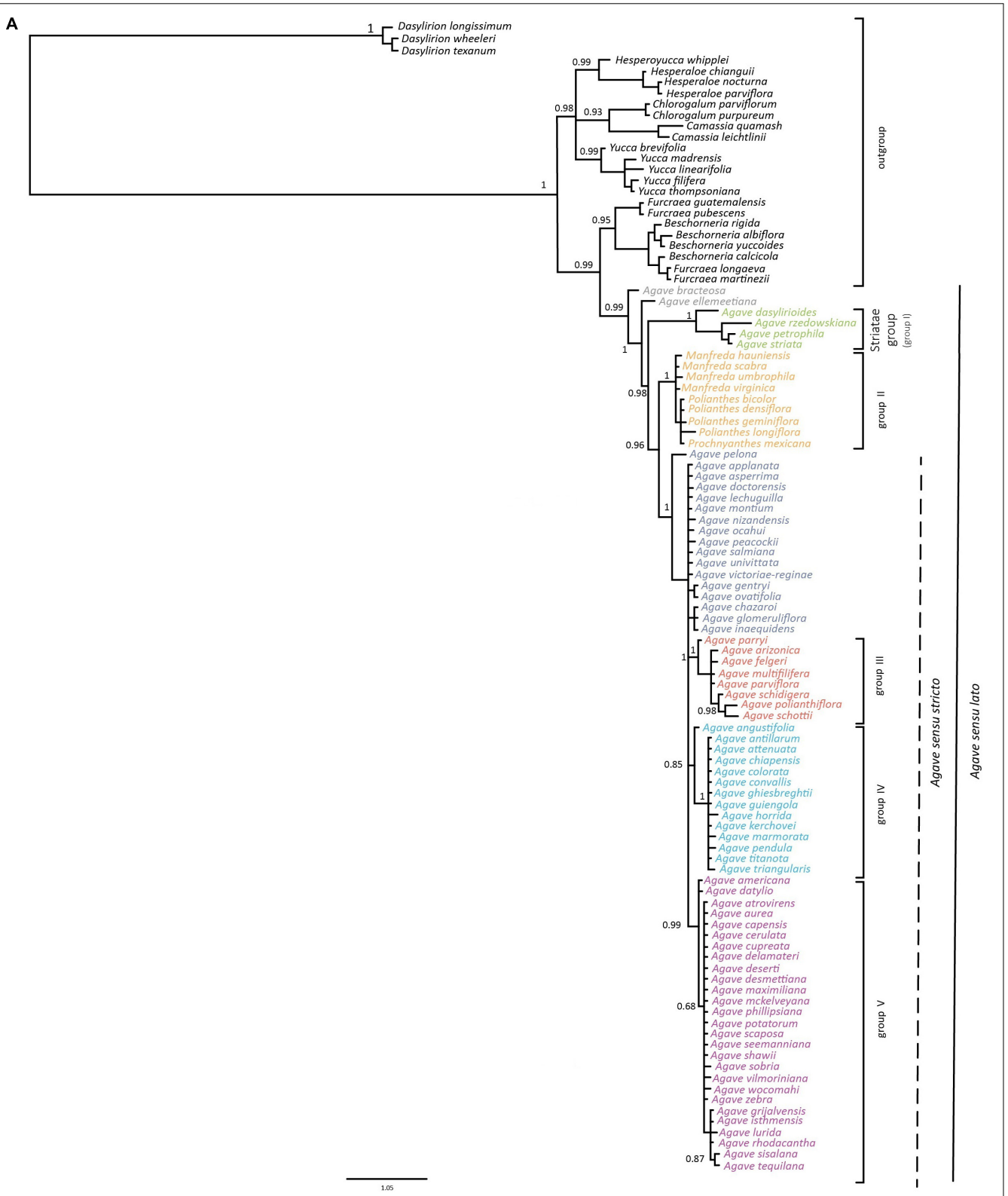


FIGURE 1 | Continued

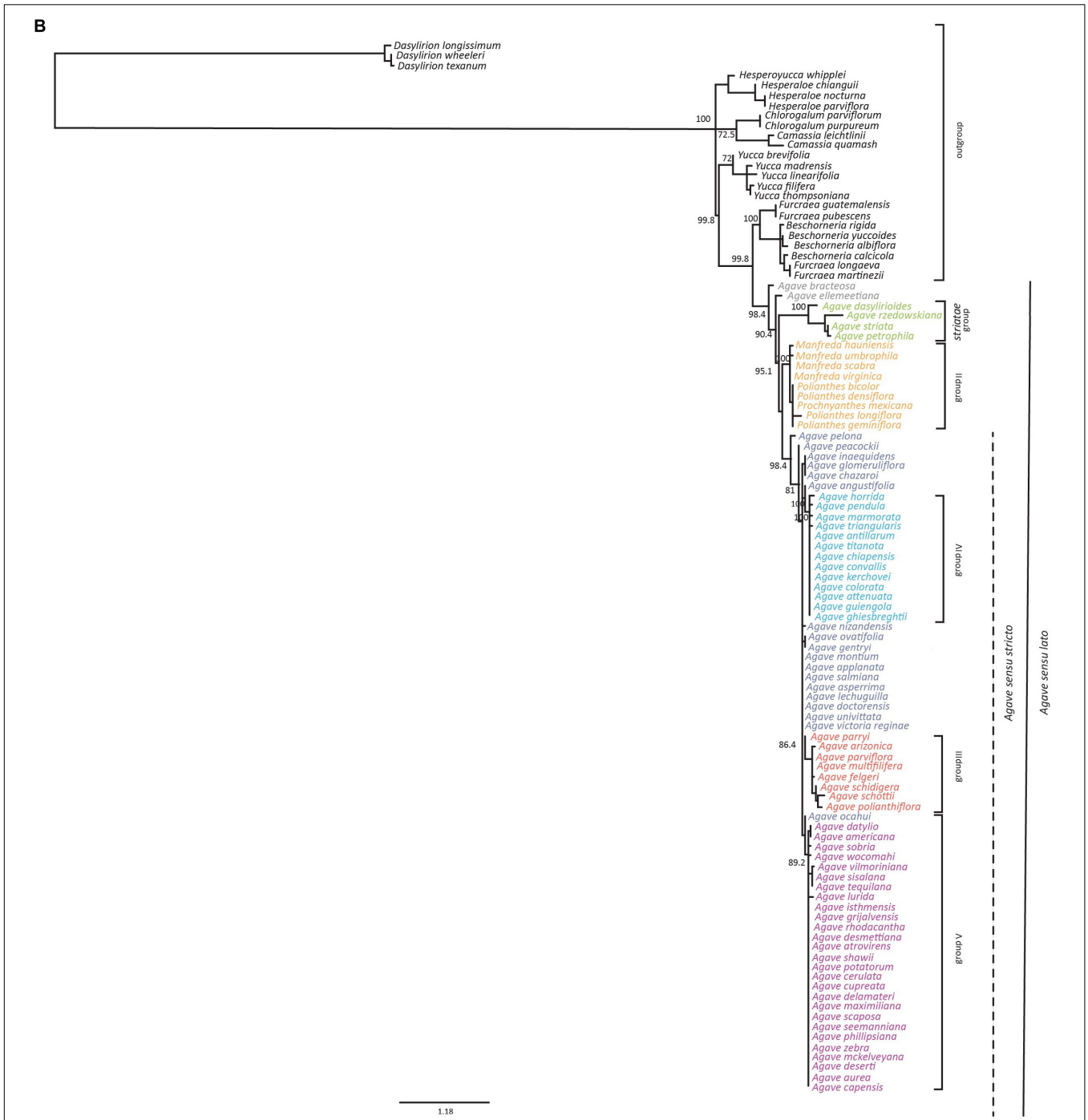


FIGURE 1 | (A) Molecular phylogeny of *Agave sensu lato* using *Dasylirion* as an outgroup reconstructed from the ITS sequences and using a Bayesian inference analysis. The numbers next to the nodes indicate the posterior probability values above 0.70. **(B)** Molecular phylogeny of *Agave sensu lato* using *Dasylirion* as an outgroup reconstructed from the ITS sequences and using a maximum likelihood analysis. Numbers next to the nodes are bootstrap values generated from running 1,000 replicates of the trees; only values above 70% are shown.

Camassia/*Chlorogalum* groups with a stem age at 14.2, but a far more recent crown age for *Yucca* at 8.52 Ma; this should be considered with caution, given that only five species of this genus were analyzed. The paraphyletic group that includes *Hesperaloe*/*Hesperoyucca* and *Camassia*/*Chlorogalum* had a more recent stem age of 7.09 Ma. Finally, the *Beschorneria*/*Furcraea* clade presented a stem age of 9 Ma and a crown age of 4.8 Ma (**Figure 2**).

For *Agave sensu lato*, the stem age was 9 Ma and a crown age of 6.18 Ma. Noticeably, within the *Agave sensu lato* clade, the Striatae

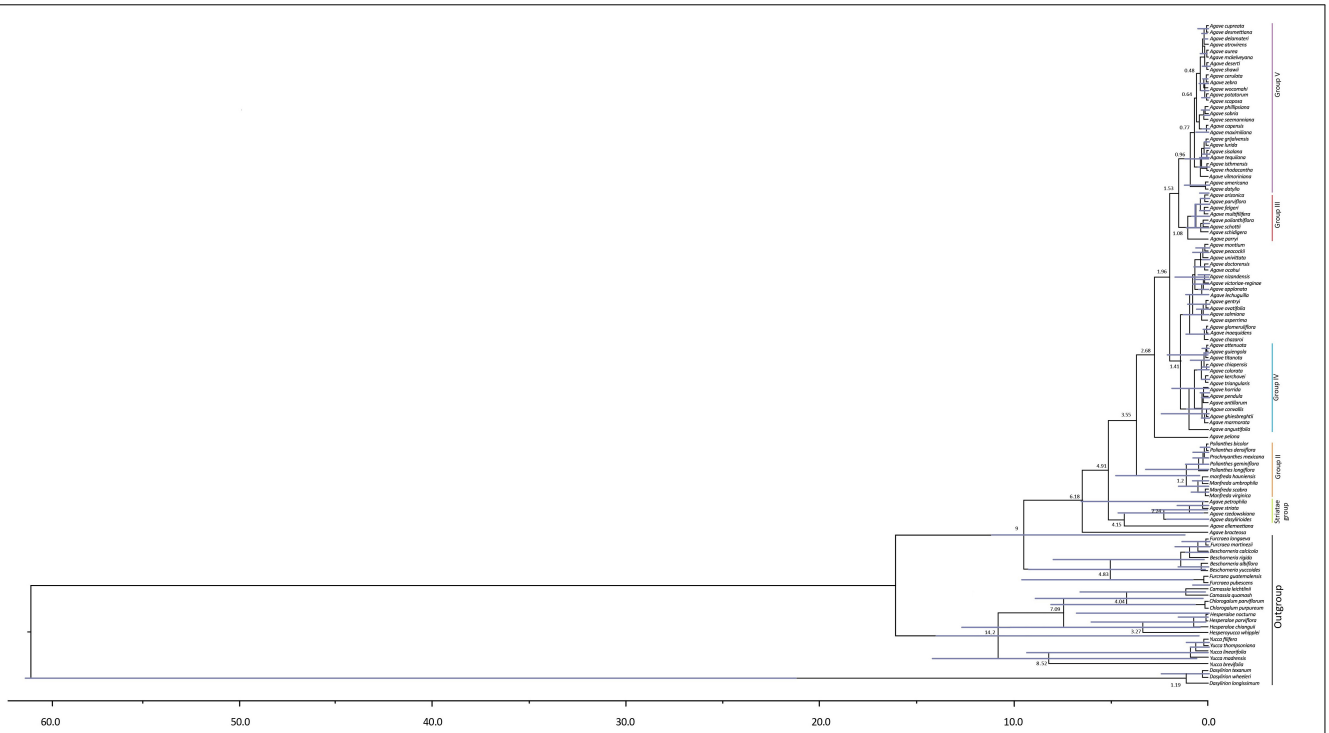


FIGURE 2 | Timing of *Agave sensu lato* diversification. Chronogram derived from the maximum clade credibility tree estimated with BEAST. Numbers indicate mean divergence times and mean ages, and the 95% highest posterior densities (HPD) are represented with bars.

group diverged earlier than the other clades, with a stem age of 4.15 Ma and a crown age of 2.24 Ma (**Figure 2**). The paraphyletic group including *Manfreda*, *Polianthes*, and *Prochnyanthes* had a stem age of 3.55 Ma and a crown age of 1.2 Ma. For the *Agave sensu stricto* clade, we estimated a stem age of 3.55 Ma and a crown age of 2.68 Ma. For Group III, we obtained a stem age of 1.53 Ma and a crown age of 1.08 Ma; for Group IV, a stem age of 1.41 Ma and a crown age of 1.04 Ma; and finally, for Group V, a stem age of 1.53 Ma and a crown age of 0.96 Ma (**Table 2**).

Inflorescence Reconstruction Analysis

The reconstruction of the inflorescence types showed ambiguous results in the sense that it was not clear if the common ancestor of the *Furcraea–Beschorneria* and *Agave sensu lato* clades had a paniculated inflorescence or not. In contrast, for the common ancestor of the *Agave sensu stricto* clade, we found a higher probability (75% of the reconstructions) for the presence of a spike inflorescence (**Figure 3**).

The reconstructions for the common ancestor of two species with early divergence from the *Agave sensu lato* clade (*A. bracteosa* and *A. ellemeetiana*), as well as Striatae clade (**Figure 3**), indicated that they may have presented a spike inflorescence (reconstruction probabilities of 80 and 60%, respectively, for each group), which was also the case for the ancestor of the paraphyletic group containing *Manfreda*, *Polianthes*, and *Prochnyanthes* (93% of the reconstructions).

Within the *Agave sensu stricto* clade, *A. pelona* is the sister taxa to the rest of the group and presents a spike inflorescence, in agreement with 75% of the reconstructions for the common ancestor of the *Agave sensu stricto* clade having spike inflorescences. The common ancestor for Group III most likely showed a spike inflorescence (98%). In contrast, for Group V, the one with the most recent origin, the common ancestor, and the extant species presented a panicle inflorescence (98%), with the exception of *A. vilmoriniana*, which reverted to a spike inflorescence (**Figure 3**).

Diversification Rates

Our estimate of diversification rate (r) using the ITS data estimated for *Agave sensu lato* was 1.50 species/Myr, with a speciation rate of $\lambda = 3.66$ species/Myr and an extinction rate of $\mu = 2.16$ species/Myr (more information in **Table 2**). In contrast, the diversification rate in the *Yucca* clade was an order of magnitude lower, $r = 0.12$ species/Myr, with $\lambda = 1.64$ species/Myr and $\mu = 1.52$ species/Myr, similar to the estimated rates for the *Furcraea–Beschorneria* clade $r = 0.11$ species/Myr, $\lambda = 1.65$ species/Myr, $\mu = 1.54$ species/Myr (**Table 2**).

BAMM identifies configurations of the rate shifts, that is, the sets of shifts that are identified together and enables to compute the relative probability of those configurations. The rate shift configuration analyses exhibited two main changes (**Figure 4A**). The first shift detected an increase in speciation rate, $r = 1.80$ species/Myr, $\lambda = 4.10$ species/Myr; this branch corresponds to the stem age of *A. bracteosa* and its sister group the *Agave sensu lato* clade at 6.18 Mya.

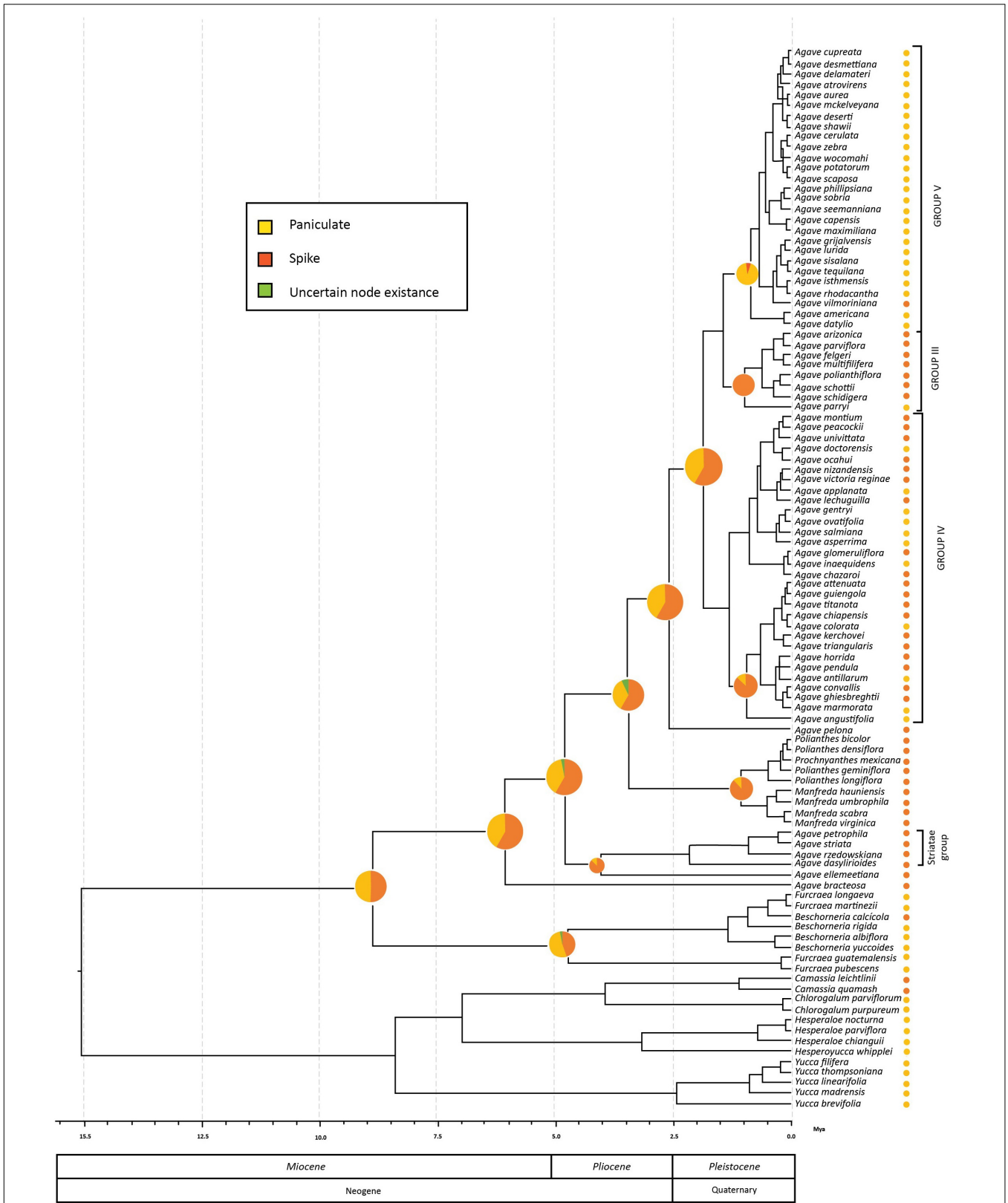


FIGURE 3 | Ancestral reconstruction of inflorescence type mapped onto *Agave sensu lato* chronogram resulting from BEAST analyses. Circles next to the tips taxa are coded to represent the actual inflorescence type. The pie charts show the probability values generated from BayesTraits for the ancestral inflorescence type reconstructed at each node.

TABLE 2 | Significant shifts in diversification rate where *r* is the diversification rate = ($\lambda - \mu$).

Clade	Stem age (Ma)	Speciation rate (λ) (species/Mya)	Extinction rate (μ) (species/Mya)	Diversification rate (*r*) (species/Mya)	Event shift rate
Yucca	14.2	1.64	1.52	0.12	
Furcraea–Beschorneria	9	1.65	1.54	0.11	
Striatae	4.15	1.76	1.58	0.18	
Manfreda–Polianthes–Prochnyanthes	3.55	3.22	2.39	0.82	
Agave sensu lato	9	3.66	2.16	1.50	
Agave sensu stricto	3.55	5.67	2.69	2.98	
Group III	1.53	6.05	2.74	3.31	
Group IV	1.41	6.15	2.74	3.41	
Group V	1.53	6.15	2.73	3.41	
Shift 1: *Agave bracteosa* + *Agave sensu lato*	6.18	Start 1.65 End 4.10	Start 1.54 End 2.30	Start 0.11 End 1.8	Increase speciation rate
Shift 2: *Agave sensu stricto*	2.68	Start 3.22 End 6.04	Start 2.39 End 2.73	Start 0.82 End 3.31	Increase speciation rate

Shift 1 and Shift 2 indicate the reconstructed start rate versus end rate.

The second shift also detected an increase in speciation rate (r = 3.31 species/My, λ = 6.04 species/Myr); this shift is located at the branch of the *Agave sensu stricto* clade at 4.91 Mya (see **Table 2**). We obtained the rate-through-time plots of speciation, extinction, and net diversification rates for all taxa included, as well as for *Agave sensu lato* and *Agave sensu stricto* clades, with BAMM in order to examine rate variation through time (**Figure 4B**).

DISCUSSION

The number of variable sites found for the plastid data set was low (29 variable from a total of 3,670 sites), which is congruent with previous analyses that included plastid markers for the *Agave sensu lato* clade (Flores-Abreu et al., 2019). The phylograms obtained previously in different studies detected clades in which *Manfreda*, *Polianthes*, and *Prochnyanthes* species are placed within the *Agave sensu stricto* group (Good-Avila et al., 2006; Flores-Abreu et al., 2019). Indeed, *Agave sensu lato* is a group that has been difficult to taxonomically classify because of overlapping variation of morphological characters between species (Gentry, 1982; García-Mendoza, 2002), which could be due to the recent origin of the group, the recent diversification events, as well as permissive hybridization between species and a long generation time (McKain et al., 2016).

The nuclear data set has more variable sites (169 from a total of 577 bp) than the plastid data set and showed higher resolution and support values. The first phylogenetic study from ITS markers for the clade by Bogler and Simpson (1996) only included nine species, whereas Eguiarte et al. (2000) included 10 species for the *Agave sensu lato* clade. In the aligned *ITS2* sequence, there is a large deletion shared by *Yucca* species, first noticed by Bogler and Simpson (1996). The groups within *Agave sensu lato* detected in our study are consistent with these previous studies (Bogler and Simpson, 1996; Eguiarte et al., 2000, 2006). In our analysis (BI and ML), the species *A. bracteosa* and *A. ellemeetiana* of the Choripetalae group can be observed, as well as the Striatae group as paraphyletic with respect to the clade *Agave sensu stricto*. It is possible that the marker used in our study, the nuclear marker ITS, allowed us to trace a different history to the one obtained using chloroplast markers. Moreover, it is known that a greater taxonomic sampling can influence the results of phylogenetic analyses.

The phylogenetic results of certain groups, such as the herbaceous, polycarpic (iteroparous) clade *Manfreda–Polianthes–Prochnyanthes*, are interesting, given that the *Polianth*es and *Prochnyanthes* are paraphyletic with respect to *Manfreda*. This nesting was also reported in previous studies (see Bogler and Simpson, 1996; Eguiarte et al., 2000; Bogler et al., 2006). The Striatae group proposed by Gentry (1982) originally consisted of only three species: *A. striata*, *A. dasylirioides* (both included in this study), and *A. stricta* (closely related to *A. striata*, see Scheinvar et al., 2017 and **Figure 2** therein). Subsequently, other species that belong to this group have been described: *A. petrophila* (García-Mendoza and Salas, 1998), *A. gracielae* (Galván and Zamudio, 2013), *A. cremnophila* (Starr et al., 2018), *A. lexii* (García-Morales and García-Jiménez, 2019), *A. rzedowskiana* (Carrillo-Reyes et al., 2003), *A. tenuifolia* (Galván and Zamudio, 2013), *A. albopilosa* (Cabral Cordero et al., 2007), and *Agave kavandivi* (García-Mendoza and Chávez-Rendón, 2013), which conform a total of 11 species, are all endemic to Mexico. *A. petrophila* and *A. rzedowskiana* were analyzed in our study. One of the most important characters of the Striatae group is the presence of finely denticulated leaf margins. It is also relevant to mention that this group seems to be formed by polycarpic (iteroparous) species, in contrast with most *A. sensu stricto* species, that are usually monocarpic (semelparous). The Striatae group seems to be less frequently pollinated by bats than the other species in the *A. sensu stricto* group (Rocha

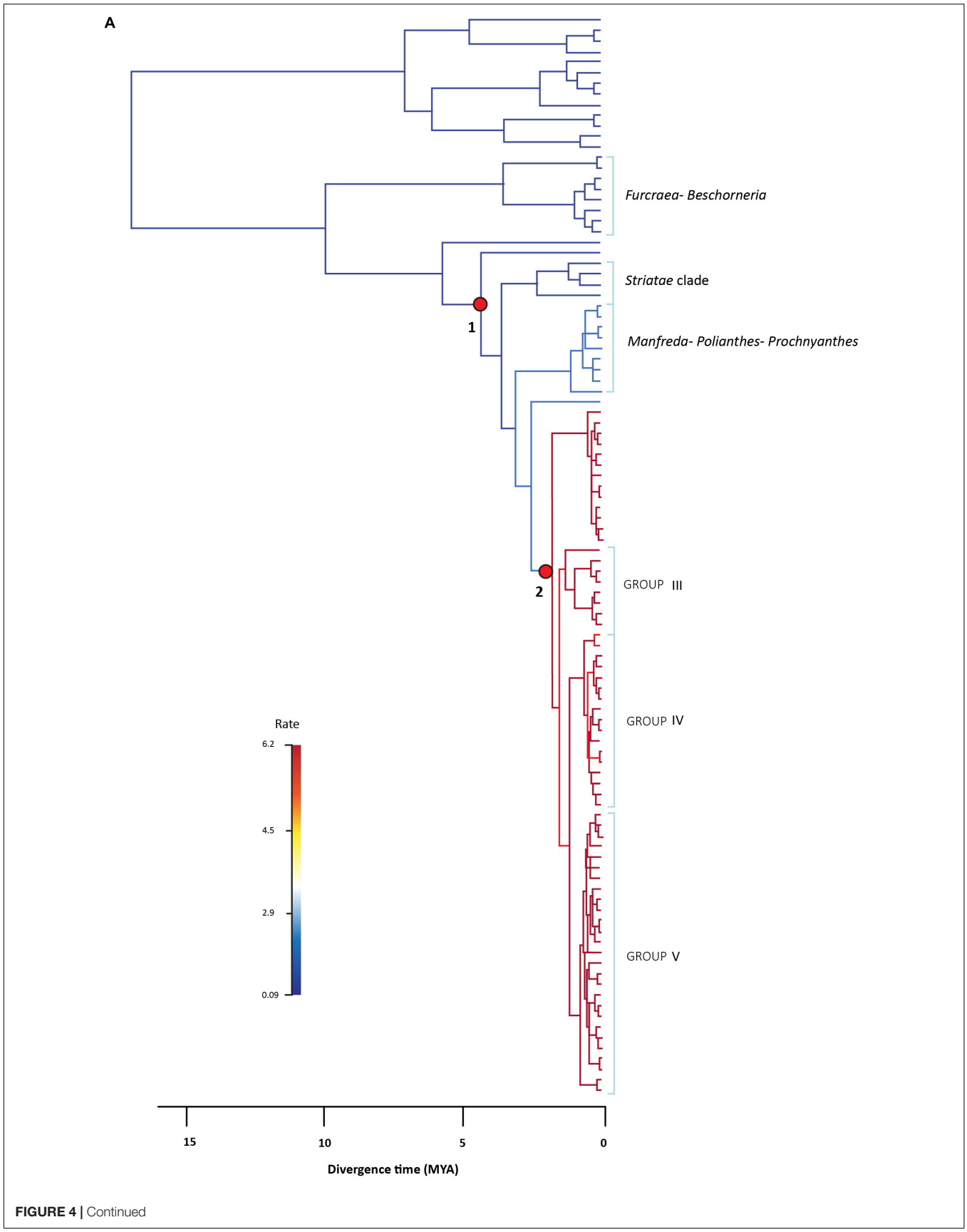


FIGURE 4 | Continued

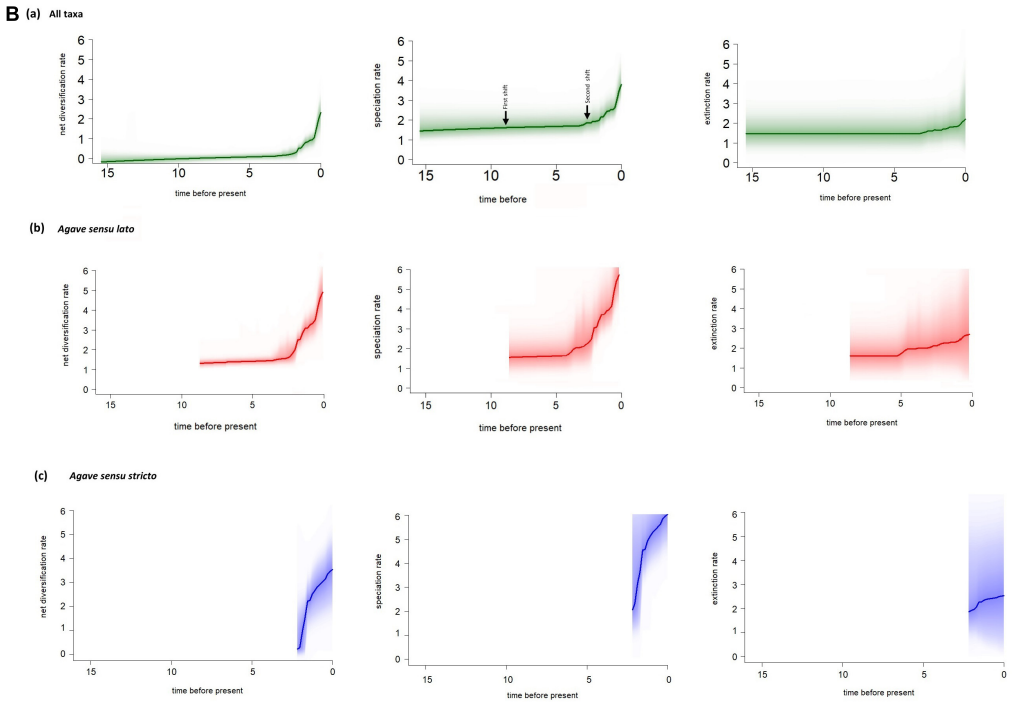


FIGURE 4 | BAMM analysis of rate shift configurations and diversification within *Agave sensu lato*. **(A)** Rate shift configurations with the two highest posterior probabilities from the 95% credible set are indicated with red circles, and branches are colored according to median net diversification rates (cool colors = slow, warm = fast). The first shift corresponds to the stem of *A. bracteosa* branch and the sister group *Agave sensu lato* clade, and the second shift corresponds to the *Agave sensu stricto* clade. Group III corresponds to the group including *A. parryi*, Group IV corresponds to the group including *A. angustifolia*, and Group V corresponds to the group including *A. americana* and *A. datylio*. **(B)** Evolutionary rates through time plots for (a) all taxa sampled, (b) *Agave sensu lato*, and (c) *Agave sensu stricto*; solid lines denote the mean of each rate-through-time curve across all agaves, and the shading intensity of the colored line for each species reflects the relative probability of a given diversification trajectory, with upper and lower bounds representing the 90% Bayesian credible interval on the distribution of rates through time.

et al., 2005, 2006). It will be important to include all the species of the Striatae group in the future and corroborate the coherence of this clade.

The two species at the base of the *Agave sensu lato* clade are *A. ellemeetiana* and *A. bracteosa*, which Gentry (1982) defined as part of his Choripetalae group. Gentry (1982) already recognized the uniqueness of this group, stating that their unarmed leaves and discoid floral receptacle are the principal characters that separate these species into another group, noticing that (page 89): "This distinctive flower structure together with the unarmed leaves without terminal spine, could justify removal from *Agave* to a separate genus," but future formal trait analyses are needed. Gentry (1982) also included *A. guiengola*, because of its virtually tubeless flower and the insertion of the filaments at the base

TABLE 3 | Age estimation in million years of major divergence events and compared with those found in other studies.

Clade	Stem group mean age	Crown group mean age	Good-Avila et al. (2006)	McKain et al. (2016)	Flores-Abreu et al. (2019)
Asparagaceae			34.2–31.7	41.27	
Yucca	14.2	8.52	19.5–11.1	20.10	7.37–16.48
Furcraea–Beschorneria	9	4.8		11.41	4.62–12.34
Striatae	4.15	2.24			
Manfreda–Polianthes–Prochnyanthes	3.55	1.2			
Agave sensu lato	9	6.18	13.1–5.9	3.09	4.62–13.34
Agave sensu stricto	4.91	3.55			
Clade III	1.53	1.08			
Clade IV	1.41	1.04			
Clade V	1.53	0.96			

of the flowers, a species that in our analysis is positioned in Group IV of *Agave sensu lato*. This last placement is not surprising, as Gentry (1982, p. 97) himself noted that "Its broad, white, ovate leaves, with their conspicuous coarse teeth, and its monocarpic rather than polycarpic habit, set off *Agave guiengola* from either of the species mentioned (i.e., *A. ellemeetiana* and *A. bracteosa*)." In particular, *A. ellemeetiana* and *A. bracteosa* are interesting species, since they have margins without any type of teeth, whereas the Striatae group presents serrulate margins; the presence of margins without teeth or serrulate is a character present in many of the early divergent species in the phylogeny of the *Agave sensu lato*. For instance, *A. pelona*, which is paraphyletic with respect to the *Agave sensu stricto* clade, was named this way by Gentry (1982) because of the absence of marginal teeth. It would be relevant to make an analysis including the total species within the Striatae group proposed by Gentry (1982), in order to observe if there is a modification in the phylogenetic relationships within this current clade.

Agave sensu lato started diversifying at 6.18 Ma according to our crown age estimate. This is congruent with the first significant diversification shift obtained from BAMM, which is at the base of the *Agave sensu lato* clade. This original shift in the diversification rate was previously reported by Good-Avila et al. (2006) and Flores-Abreu et al. (2019). At this point, the mean speciation rate increased (λ = 4.10 species/Myr). This shift rate occurred in the late Miocene period, which is characterized by the emergence of arid biomes in America, resulting in the rise of new mountains, such as the Trans-Mexican Neovolcanic Belt and the Sierra Madre Occidental (Mastretta-Yanes et al., 2015). This emergence caused changes in humidity and wind currents, originating new habitats, which generated new ecological opportunities for several lineages that inhabit these arid areas today (Morán-Zenteno and Wilson, 1994; Good-Avila et al., 2006; Arakaki et al., 2011; De-Nova et al., 2012; Hernández-Hernández et al., 2014). The first rate shift is linked with the origin of the *Agave sensu lato* clade, and it could be the starting point for the diversification of agaves. In *Agave sensu lato*, the earlier divergent groups [i.e., *A. ellemeetiana* and *A. bracteosa* (Choripetalae group) and clades Striatae and *Manfreda–Polianthes–Prochnyanthes*] have predominant ancestors with spike inflorescences according to our ancestral character state reconstruction, which are commonly pollinated by bees and hawk moths (Eguiarte, 1995; Rocha et al., 2005, 2006). This trend was maintained in the *Agave sensu stricto* clade, where the earlier divergent groups still have spike inflorescences. For instance, Groups III and IV, two of the most recent groups, at 1.2 Ma usually display spike inflorescences, although several conversions toward paniculate inflorescences also occur. This time could have represented a period of transition, when the agaves went from having mainly spike inflorescences to evolving paniculate inflorescences, until reaching the origin of the recent Group V (0.96 Ma), where the predominant character is a paniculate inflorescence.

Are these reconstructed chains of events consistent with what we know about inflorescence development and evolution? Inflorescence architecture is the consequence of developmental programs that dictate inflorescence meristem activity and determine organ topology, geometry, and phenology by means of the regulatory processes affecting meristem identity, size, and maintenance, as well as axillary meristem initiation and organogenesis (Zhang and Zheng, 2014). There is ample evidence that these programs are hormonally and genetically controlled, and the rich diversity in inflorescence architecture in angiosperms is evidence of its enormous plasticity (Harder and Prusinkiewicz, 2013; Zhang and Zheng, 2014). Inflorescence architecture can influence pollination and seed yield, playing important roles in natural selection. Complex, simple, or small architectures solve the problem of attracting specific kinds of pollinators or promote self-pollination (Harder and Prusinkiewicz, 2013). Agavoideae displays varying inflorescence architectures, panicles being more common in *Agave*, *Manfreda*, *Beschorneria*, *Furcraea*, *Hesperaloe*, and *Yucca*, whereas spike or raceme inflorescences are found in *Hesperaloe*, *Polianthes*, *Prochnyanthes*, and also in *Agave* (Aker, 1982; Gentry, 1982; Starr, 1997; García-Mendoza, 2000; Castro-Castro et al., 2010, 2018; Solano et al., 2013; Cházaro-Basáñez and Vázquez-Ramírez, 2015). High plasticity of inflorescence architecture has been more clearly demonstrated in grasses, where molecular switches can significantly increase secondary and tertiary branching, thus changing inflorescence morphology (Zhang and Zheng, 2014). Therefore, inflorescence architecture in Agavoideae can be reasonably considered as homoplastic, given its plastic nature. The underlying natural forces that kept quite stable spike inflorescences in the early-diverging *Agave* groups (Choripetalae and Groups I and II), panicle inflorescences in the late-diverging groups (Group V), and frequent reversions between the two forms (Groups III and IV) remain to be studied.

During the Pliocene and Pleistocene, the agaves had the greatest amount of diversification events. This corresponds to the second rate shift within the stem of *A. pelona* and its sister group *Agave sensu stricto* at 2.68 Ma when we observed an increase in the speciation rate (λ = 6.04 species/Myr). This is concordant with Scheinvar et al. (2017),Scheinvar (2018), and Aguirre-Planter et al. (2020) that suggested that current *Agave sensu lato* distribution and species richness could be related to glaciation and interglacial events during the Pleistocene that caused the expansion and contraction of the species distribution, thus influencing the evolution of agave populations. Three localities of interest during this period are the southern portion of Sierra Sur de Chihuahua, which served as refuge during the last interglacial period, the Sierra Madre del Sur, which is considered as a refuge during the Last Glacial Maximum (21,000–17,000 years), and the California Sierra during both periods. This second increase in the diversification rate is in accordance with a second diversification proposed by Good-Avila et al. (2006), but that was not detected by Flores-Abreu et al. (2019), even though the latter study had a large sample. Good-Avila et al. (2006) suggested that this second shift was related to changes in the pollinators (to bat pollination), but the analyses of Flores-Abreu et al. (2019) falsified this idea, as bat pollination in *Agave* seems to be older. The second shift could represent a secondary adaptation to different climates that permitted the lineages to diversify, including adaptation to more mesic conditions in Central and West Mexico, giving rise, for instance, to the radiation of the *Manfreda–Polianthes–Prochnyanthes* herbaceous group and

of some agave groups that live in less arid environments in central Mexico.

It is clear that agaves keep a close relationship with their pollinators, as exemplified by the large number of agave species that are distributed in the so called "nectar corridors" found in the migratory routes of several bat species (Moreno-Valdez et al., 2004; Trejo-Salazar et al., 2016), in the phenologies of pollen and nectar production (Schaffer and Schaffer, 1977; Nassar et al., 2003), and in the relationship between bat visits and agave reproduction rate and genetic variation, as well as in their function as primary pollinators, although the total list of floral visitors and potential pollinators is wide (Howell and Roth, 1981; Eguiarte and Búrquez, 1987; Arizaga et al., 2000; Slauson, 2000, 2001; Molina-Freaner and Eguiarte, 2003). However, it does not appear to be a strict and tight coevolution process, as in other groups of the Asparagaceae (Flores-Abreu et al., 2019), such as in the *Yucca* family, because few bat species, in particular *Leptonycteris yerbabuenae*, visit and pollinate many species of agave, whereas each species of *Yucca* seems to coevolve with a particular *Tegeticula* moth (Pellmyr, 2003).

On the other hand, the divergence time for clades reported in our study is more recent than the values estimated in previous studies. We consider that this is a consequence of sampling more taxa, as the inferred branch lengths become also shorter. Nonetheless, the periods in which we found increases in speciation rate are congruent with previous studies, as commented above (Good-Avila et al., 2006; Flores-Abreu et al., 2019). It is likely that pollinators, especially bats, have influenced the *Agave* diversification processes and have had a relevant role in selecting the type of inflorescence that agaves currently present.

The diversification rate of *Agave sensu lato* ($r = 1.50$ species/Myr) is clearly higher than that of related groups, such as *Yucca* ($r = 0.12$ species/Myr) and *Furcraea–Beschorneria* ($r = 0.11$ species/Myr), values similar to those reported by Flores-Abreu et al. (2019). The mean speciation rate showed that speciation was low during the stem divergence of the group and increased at the base of the *Agave sensu lato* clade. *Agave sensu lato* has a considerably higher diversification rate than average estimates reported for other flowering plants, which range from 0.078 to 0.09 species/Mya (Magallón and Castillo, 2009). The speciation rate seems to increase rapidly at the *Agave sensu stricto* branch ($\lambda = 5.67$ species/Myr), continuing to increase in the three main groups that conform the *Agave sensu stricto* clade: Group III with a $\lambda = 6.05$ species/Myr, Group IV with a $\lambda = 6.15$ species/Myr, and Group V with a $\lambda = 6.15$ species/Myr. On the basis of the elevated speciation rate values estimated for the *Agave sensu lato* clades, we can conclude that as a group of recent origin, it is experiencing an intense process of diversification (**Table 3**).

DATA AVAILABILITY STATEMENT

All sequences generated for this study were deposited in the NCBI GenBank under accession numbers shown in **Supplementary Table S1**.

AUTHOR CONTRIBUTIONS

OJ-B contributed to the laboratory work, data analysis, and drafting of the manuscript. RG-S contributed to the phylogenetic analysis, reconstruction of ancestral state analyses, and design of some figures. SM helped in drafting and correcting sections of the manuscript. AG-M helped in drafting and correcting sections of the manuscript and with the species collection for the analyses. JN-S contributed with ideas for the design of the project, DNA material, and correcting sections of the manuscript. EA-P helped in drafting and correcting sections of the manuscript. LE, as the project leader, designed and coordinated the project and logistics and drafted and corrected the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Instituto de Ecología, Universidad Nacional Autónoma de México (operative funding to LE) and CONACYT Investigación Científica Básica 2011/167826 to LE.

ACKNOWLEDGMENTS

This manuscript forms part of the doctoral research conducted by Ofelia Jimenez, who thanks the Doctorado en Ciencias Biológicas and especially the Universidad Nacional Autonoma de México and acknowledges the scholarship provided by the Consejo Nacional de Ciencia y Tecnologia (Grant No. 508586). We thank Enrique Scheinvar and Laura Espinosa-Asuar for providing laboratory help during this study and the members of the Laboratorio de Evolución Molecular y Experimental for their support and comments during the realization of this study. We thank Jesús Gutiérrez, Idalia Rojas, Yajima Osorno, and Luz María Rangel from the Instituto de Biología of the UNAM for isolating genomic DNA from several *Agave* species of this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.536135/full#supplementary-material

Supplementary Figure 1 | Molecular phylogeny of *Agave sensu lato* using *Yucca* as an outgroup reconstructed from the chloroplast data set sequences and using a maximum likelihood analysis. The numbers next to the nodes indicate the posterior probability values above 70%.

Supplementary Figure 2 | Molecular phylogeny of *Agave sensu lato* using *Yucca* as an outgroup reconstructed from the chloroplast data set sequences and using a Bayesian inference analysis. The numbers next to the nodes indicate the posterior probability values above 0.70.

REFERENCES

Aguirre-Planter, E., Parra-Leyva, J. G., Ramírez-Barahona, S., Scheinvar, E., Lira-Saade, R., and Eguiarte, L. E. (2020). Phylogeography and genetic diversity in a southern North American desert: *Agave kerchovei* from the Tehuacán-Cuicatlán Valley, Mexico. *Front. Plant Sci.* 11:863.

Aker, C. L. (1982).). Regulation of flower, fruit and seed production by monocarpic perennial *Yucca whipplei*. *J. Ecol.* 70, 357–372. doi: 10.2307/2259884

Álvarez de Zayas, A. (1995). Los agaves de las Antillas. *Boletín Sociedad Botánica México* 57, 37–48. doi: 10.17129/botsci.1475

Arakaki, M., Pascal-Atonie, C., Nyffeler, R., Lendel, A., Eggli, U., Ogburn, R. M., et al. (2011). Contemporaneous and recent radiations of the world's major succulent plant linages. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8379–8384.

Archibald, J. K., Kephart, S. R., Theiss, K. E., Petrosky, A. L., and Culley, T. M. (2015). Multilocus phylogenetic inference in subfamily Chlorogaloideae and related genera of Agavaceae- Informing questions in taxonomy at multiple ranks. *Mol. Phylogenet. Evol.* 84, 266–283. doi: 10.1016/j.ympev.2014.12.014

Arizaga, S., Ezcurra, E., Peters, E., Ramírez, F., and Vega, E. (2000). Pollination ecology of *Agave macroacantha* (Agavaceae) in a Mexican tropical desert. II. The role of pollinators. *Am. J. Bot.* 87, 1011–1017. doi: 10.2307/2657001

Berger, A. (1915). *Die Agaven: Beiträge zu einer Monographie*. Stuttgart: G. Fischer.

Bogler, D. J., Pires, C., and Francisco-Ortega, J. (2006). Phylogeny of Agavaceae base on ndhF, rbcL, and ITS sequences: implications of molecular data for classification. *Aliso* 22, 313–328. doi: 10.5642/aliso.20062201.26

Bogler, D. J., and Simpson, B. B. (1995). A chloroplast DNA study of the Agavaceae. *Syst. Bot.* 20, 191–205. doi: 10.2307/2419449

Bogler, D. J., and Simpson, B. B. (1996). Phylogeny of Agavaceae based on ITS DNA sequence variation. *Am. J. Bot.* 9, 1225–1235. doi: 10.1002/j.1537-2197.1996.tb13903.x

Bouchenak-Khelladi, Y., Onstein, R. E., Xing, Y., Schwery, O., and Linder, H. P. (2015). On the complexity of triggering evolutionary radiations. *New Phytol.* 207, 313–326. doi: 10.1111/nph.13331

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., et al. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537.

Cabral Cordero, I., Villarreal Quintanilla, J. Á, and Estrada Castillón, E. A. (2007). *Agave albopilosa* (Agavaceae, subgénero Littaea, grupo Striatae), una especie nueva de la Sierra Madre Oriental en el noreste de México. *Acta Botánica Mexicana* 80, 51–57. doi: 10.21829/abm80.2007.1046

Carrillo-Reyes, P., Vega Aviña, R., and Ramírez-Delgadillo, R. (2003). *Agave rzedowskiana*, a new species in subgenus Littaea (Agavaceae) from western Mexico. *Brittonia* 3, 240–244. doi: 10.1663/0007-196x(2003)055[0240:aransi]2.0.co;2

Castro-Castro, A., Rodríguez, A., Vargas-Amado, G., and Ramírez-Delgadillo, R. (2010). Variación morfológica del género Prochnyanthes (Agavaceae). *Acta Bot. Mexicana* 92, 29–49. doi: 10.21829/abm92.2010.282

Castro-Castro, A., Zamora-Tavares, P., Carrillo-Reyes, P., and Rodríguez, A. (2018). Manfreda santana-michelii (Asparagaceae subfamily Agavoideae), a striking new species from Sierra Madre del Sur in Western Mexico. *Syst. Bot.* 43, 497–501. doi: 10.1600/036364418x697229

Cházaro-Basáñez, M., and Vázquez-Ramírez, J. (2015). Introducing the succulent flora of Mexico: *Beschorneria yuccoides* (Agavaceae). *Cactus Succulent J.* 87, 197–198.

Coyne, J. A., and Orr, H. A. (2004). *Speciation*. Sunderland: Sinauer Associates.

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109

De Queiroz, K. (2007). Species concepts and species delimitation. *Syst. Biol.* 6, 879–886. doi: 10.1080/10635150701701083

De-Nova, J. A., Medina, R., Montero, J. C., Weeks, A., Rosell, J. A., Olson, M. E., et al. (2012). Insights into the historical construction of species-rich Mesoamerican seasonally dry tropical forests: the diversification of *Bursera* (Burseraceae, Sapindales). *New Phytol.* 193, 276–287. doi: 10.1111/j.1469-8137.2011.03909.x

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 9, 11–15.

Eguiarte, E. L., Souza, V., and Silva-Montellano, A. (2000). Evolución de la Familia Agavaceae: Filogenia, biología reproductiva y genética de poblaciones. *Boletín Sociedad Botánica México* 66, 131–150. doi: 10.17129/botsci.1618

Eguiarte, L., and Búrquez, A. (1987). Reproductive ecology of *Manfreda brachystachya*, an iteroparous species of Agavaceae. *Southwestern Nat.* 32, 169–178. doi: 10.2307/3671560

Eguiarte, L. E. (1995). Hutchinson (Agavales) vs. Huber y Dahlgren (Asparagales): análisis moleculares sobre la filogenia y evolución de las familias Agavaceae sensu Hutchinson dentro de las monocotiledóneas. *Boletín Sociedad Botánica México* 56, 45–56. doi: 10.17129/botsci.1463

Eguiarte, L. E., Tambutti, M., Silva-Montellano, A., Golubov, J. K., Rocha, M., and Souza, V. (2006). Zur Naturgeschichte der Gattung *Agave* (Agavaceae): Taxonomie, Ökologie und Schutz (Teil 1). *Avonia* 24, 8–19.

Flores-Abreu, I. N., Trejo-Salazar, R. E., Sánchez-Reyes, L. L., Good, S. V., Magallón, S., García-Mendoza, A., et al. (2019). Tempo and mode in coevolution of *Agave sensu lato* (Agavoideae, Asparagaceae) and its bat pollinators, Glossophaginae (Phyllostomidae). *Mol. Phylogenet. Evol.* 133, 176–188. doi: 10.1016/j.ympev.2019.01.004

Galván, R., and Zamudio, S. (2013). Una nueva especie de *Agave* subgénero Littaea (Agavaceae) del estado de Querétaro, México. *Acta Bot. Mexicana* 105, 1–10. doi: 10.21829/abm105.2013.228

Garamszegi, L. Z. (2014). *Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology*. Berlin: Springer.

García-Mendoza, A. (1995). "Riqueza y endemismos de la familia Agavaceae en México," in *Conservación de Plantas en Peligro de Extinción: Diferentes Enfoques*, eds E. Linares, P. Dávila, F. Chiang, R. Bye, and T. Elias (México city: Universidad Nacional Autónoma de México), 51–75.

García-Mendoza, A. (2000). Revisión taxonómica de las especies arborescentes de *Furcraea* (Agavaceae) en México y Guatemala. *Boletín Sociedad Botanica México* 66, 113–129. doi: 10.17129/botsci.1617

García-Mendoza, A. (2002). Distribution of agave (Agavaceae) in México. *Cactus Succulent J.* 74, 177–188.

García-Mendoza, A., and Salas, E. M. (1998). Una nueva especie de *Agave*, subgénero littaea (Agavaceae) de Guerrero y Oaxaca, México. *JSTOR* 18, 227–230.

García-Mendoza, A. J., and Chávez-Rendón, C. (2013). *Agave kavandivi* (Agavaceae: grupo Striatae), una especie nueva de Oaxaca, México. *Rev. Mexicana Biodivers.* 84, 1070–1076. doi: 10.7550/rmb.35241

García-Mendoza, A. J., Franco Martínez, I. S., and Sandoval Gutiérrez, D. (2019). Cuatro especies nuevas de *Agave* (Asparagaceae, Agavoideae) del sur de México. *Acta Bot. Mexicana* 126, 1–18.

García-Morales, L. J., and García-Jiménez, J. (2019). *Agave lexii* (Asparagaceae: Agavoideae), a new species from Mexico. *Novon* 27, 201–204. doi: 10.3417/2019402

Gentry, H. S. (1982). *Agaves of Continental North America*. Tucson: University of Arizona.

Givnish, T. J. (2010). Ecology of plant speciation. *Taxon* 59, 1326–1366. doi: 10.1002/tax.595003

Glor, R. E. (2010). Phylogenetic insights on adaptive radiation. *Annu. Rev Ecol. Evol. Syst.* 41, 251–270. doi: 10.1146/annurev.ecolsys.39.110707.173447

Good-Avila, S. V., Souza, V., Gaut, B. S., and Eguiarte, L. E. (2006). Timing and rate of speciation in *Agave* (Agavaceae). *Proc. Natl. Acad. Sci. U.S.A.* 103, 9124–9129. doi: 10.1073/pnas.0603312103

Harder, L. D., and Prusinkiewicz, P. (2013). The interplay between inflorescence development and function as the crucible of architectural diversity. *Ann. Bot.* 112, 1477–1493. doi: 10.1093/aob/mcs252

Harrington, S. M., and Reeder, T. W. (2017). Phylogenetic inference and divergence dating of snakes using molecules, morphology and fossils: new insights into convergent evolution of feeding morphology and limb reduction. *Biol. J. Linnean Soc.* 121, 379–394. doi: 10.1093/biolinnean/blw039

Harvey, P. H., and Pagel, M. D. (1991). *The Comparative Method in Evolutionary Biology*. Oxford: Oxford University Press.

Helfrich, P., Rieb, E., Abrami, G., Lücking, A., and Mehler, A. (2018). "TreeAnnotator: versatile visual annotation of hierarchical text relations," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (LREC 2018).

Hernández-Sandoval, L. (1995). Análisis cladístico de la familia Agavaceae. *Boletín Sociedad Botánica México* 56, 57–68. doi: 10.17129/botsci.1464

Hernández-Hernández et al., T., Brown, J. W., Schlumpberger, B. O., Eguiarte, L. E., and Magallón, S. (2014). Beyond aridification: multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New Phytol.* 202, 1382–1397. doi: 10.1111/nph.12752

Heyduk, K., McKain, M. R., Lalani, F., and Leebens-Mack, J. (2016). Evolution of a CAM anatomy predates the origins of Crassulacean acid metabolism in the Agavoideae (Asparagaceae). *Mol. Phylogenet. Evol.* 105, 102–113. doi: 10.1016/j.ympev.2016.08.018

Howell, D. J., and Roth, B. S. (1981). Sexual reproduction in agaves: the benefits of bats; the cost of semelparous advertising. *Ecology* 62, 1–7. doi: 10.2307/1936660

Huelsenbeck, J. P., Larget, B., and Alfaro, M. E. (2003). Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol. Biol. Evol.* 21, 1123–1133. doi: 10.1093/molbev/msh123

Huelsenbeck, J. P., and Rannala, B. (2004). Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst. Biol.* 53, 904–913. doi: 10.1080/10635150490522629

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Lagomarsino, L. P., Condomine, F. L., Antonelli, A., Mulch, A., and Charles, C. D. (2016). The abiotic and abiotic drivers of rapid diversifications in Andean bellflowers (Campanulaceae). *New Phytol.* 210, 1430–1442. doi: 10.1111/nph.13920

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2016). Partition Finder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773.

Lledías, F., Gutiérrez, J., Martínez-Hernández, A., García-Mendoza, A., Sosa, E., Hernández-Bermúdez, F., et al. (2020). Mayahuelin, a type I Ribosome Inactivating Protein: characterization, evolution, and utilization in phylogenetic analyses of *Agave*. *Front. Plant Sci.* 11:573. doi: 10.3389/fpls.2020.00573

Losos, J. B. (2010). Adaptative radiation, ecological opportunity and evolutionary determinism. *Am. Nat.* 175, 623–639. doi: 10.1086/652433

Magallón, S., and Castillo, A. (2009). Angiosperm diversification through time. *Am. J. Bot.* 96, 349–365. doi: 10.3732/ajb.0800060

Magallón, S., Gómez−Acevedo, S., Sánchez−Reyes, L. L., and Hernández−Hernández, T. (2015). A metacalibrated time−tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* 207, 437–453. doi: 10.1111/nph.13264

Magallón, S., and Sanderson, M. J. (2001). Absolute diversification rates in angiosperm clades. *Evolution* 55, 1762–1780. doi: 10.1554/0014-3820(2001)055[1762:adriac]2.0.co;2

Mastretta-Yanes, A., Moreno-Letelier, A., Piñero, D., Jorgensen, T. H., and Emerson, B. C. (2015). Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. *J. Biogeogr.* 42, 1586–1600. doi: 10.1111/jbi.12546

McKain, M. R., McNeal, J. R., Kellar, P. R., Eguiarte, L. E., Pires, J. C., and Leebens-Mack, J. (2016). Timing of rapid diversification and convergent origins of active pollination within Agavoideae (Asparagaceae). *Am. J. Bot.* 103, 1717–1729. doi: 10.3732/ajb.1600198

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). "Creating the CIPRES science gateway for inference of large phylogenetic trees," in *Proceedings of the Gateway Computing Environments Workshop*, New Orleans, LA, 1–8.

Molina-Freaner, F., and Eguiarte, L. E. (2003). The pollination biology of two paniculate agaves (Agavaceae) from Northwestern Mexico: contrasting roles of bats as pollinators. *Am. J. Bot.* 90, 1016–1024. doi: 10.3732/ajb.90.7.1016

Morales, E. (2000). El método comparativo en ecología vegetal. *Boletín Sociedad Botánica México* 66, 37–52.

Morán-Zenteno, D., and Wilson, J. L. (1994). *The Geology of the Mexican Republic*. Tulsa: American Association of Petroleum Geologists.

Moreno-Valdez, A., Honeycutt, R. L., and Grant, E. W. (2004). Colony dinamics of *Leptonycteris nivallis* (Mexican long-nosed bat) related to flowering *Agave* in northern Mexico. *J. Mammal.* 8, 453–459. doi: 10.1644/1545-1542(2004)085<0453:cdolnm>2.0.co;2

Müller, J., Müller, K. F., Neinhuis, C., and Quandt, D. (2006). *PhyDE–Phylogenetic Data Editor*. Available online at: http://www.phyde.de (accessed January 20, 2020).

Nassar, J. M., Beck, H., Da, S. L., Stenberg, L., and Fleming, T. H. (2003). Dependence on cacti and agaves in nectar-feeding bats from Venezuelan arid zones. *J. Mammal.* 1, 106–116. doi: 10.1644/1545-1542(2003)084<0106:docaai>2.0.co;2

Nosil, P. (2012). *Ecological Speciation*. Oxford: Oxford University.

Olson, M. E., and Arroyo-Santos, A. (2009). Thinking in continua: beyond the "adaptive radiation" metaphor. *BioEssays* 31, 1337–1346. doi: 10.1002/bies.200900102

Pagel, M., Meade, A., and Barker, D. (2004). Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* 53, 673–684. doi: 10.1080/10635150490522232

Pellmyr, O. (2003). Yuccas, yucca moths, and coevolution: a review. *Ann. Missouri Bot. Garden* 90, 35–55. doi: 10.2307/3298524

Percival, M., and Morgan, P. (1965). Observations on the floral biology of digitalis species. *New Phytol.* 64, 1–23. doi: 10.1111/j.1469-8137.1965

Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., et al. (2014). BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol. Evol.* 5, 701–707. doi: 10.1111/2041-210x.12199

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

Rezende, E., and Garland, T. (2003). *Comparaciones Interespecíficas y Métodos Filogenéticos Estadísticos*. Chile: Universidad Católica de Chile.

Rocha, M., Good-Ávila, S. V., Molina-Freaner, F., Arita, H. T., Castillo, A., García-Mendoza, A., et al. (2006). Pollination biology and adaptive radiation of Agavaceae, with special emphasis on the genus *Agave*. *Aliso J. Syst. Evol. Bot.* 22, 329–344. doi: 10.5642/aliso.20062201.27

Rocha, M., Valera, A., and Eguiarte, L. E. (2005). Reproductive ecology of five sympatric *Agave Littaea* (Agavaceae)species in central Mexico. *Am. J. Bot.* 92, 1330–1341. doi: 10.3732/ajb.92.8.1330

Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed modes. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180

RStudio Team (2020). *RStudio: Integrated Development for R*. Boston, MA: RStudio, PBC.

Schaffer, M. W., and Schaffer, M. V. (1979). The adaptative significance of variations in reproductive habit in the Agavaceae II. Pollinator foraging behavior and selection for increased reproductive expenditure. *Ecology* 60, 1051–1069. doi: 10.2307/1936872

Schaffer, W. M., and Schaffer, M. V. (1977). The reproductive biology of Agavaceae. I. Pollen and nectar production in four Arizona agaves. *Southwestern Nat.* 22, 157–168. doi: 10.2307/3669806

Scheinvar, E., Gamez, N., Catellanos-Morales, G., Aguirre-Planter, E., and Eguiarte, L. E. (2017). Neogene and Pleistocene history of *Agave lechuguilla* in the Chihuahuan desert. *J. Biogeogr.* 44, 322–334. doi: 10.1111/jbi.12851

Scheinvar, G. E. (2018). *Filogeografía de Agave Lechuguilla y Patrones de Distribución de Agave en México*. Ph.D. thesis, Universidad Nacional Autónoma de México, Mexico city.

Schlumpberger, B. O. (2012). "A survey on pollination modes in cacti and a potential key innovation," in *Evolution of Plant-Pollinator Interactions*, ed. S. Patiny (Cambridge: Cambridge University Press), 301–319. doi: 10.1017/cbo9781139014113.011

Schlumpberger, B. O., and Renner, S. S. (2012). Molecular phylogenetics of *Echinopsis* (Cactaceae): polyphyly at all levels and convergent evolution of pollination modes and growth forms. *Am. J. Bot.* 99, 1335–1349. doi: 10.3732/ajb.1100288

Schluter, D. (1996). Ecological causes of adaptative radiation. *Am. Nat.* 148, 40–64.

Schluter, D. (2000). *The Ecology of Adaptative Radiation*. Oxford: Oxford University.

Schluter, D. (2001). Ecology and the origin of the species. *Trends Ecol. Evol.* 16, 372–380. doi: 10.1016/s0169-5347(01)02198-x

Scott, A. H., and Arnold, M. L. (1995). Spurring plant diversification: are floral nectar spurs a key innovation? *Proc. R. Soc. Biol. Sci.* 262, 343–348. doi: 10.1098/rspb.1995.0215

Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., et al. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92, 142–166. doi: 10.3732/ajb.92.1.142

Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., et al. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am. J. Bot.* 94, 275–288. doi: 10.3732/ajb.94.3.275

Silva-Montellano, A., and Eguiarte, L. E. (2003). Geographic patterns in the reproductive ecology of *Agave lechuguilla* (Agavaceae) in the Chihuahuan desert. I. Floral characteristics, visitors and fecundity. *Am. J. Bot.* 90, 377–387. doi: 10.3732/ajb.90.3.377

Slauson, L. A. (2000). Pollination biology of two chiropterophilous agaves in Arizona. *Am. J. Bot.* 87, 825–836. doi: 10.2307/2656890

Slauson, L. A. (2001). Insights on the pollination biology of *Agave* (Agavaceae). *Haseltonia* 8, 10–23.

Smith, C. I., Pellmyr, O., Althoff, D. M., Balcazar-Lara, M., Leebens-Mack, J., and Segraves, K. A. (2008). Pattern and timing of diversification in *Yucca* (Agavaceae): specialized pollination does not escalate rates of diversification. *Proc. R. Soc.* 275, 249–258. doi: 10.1098/rspb.2007.1405

Solano, E., and Feria, T. P. (2007). Ecological niche modeling and geographic distribution of the genus *Polianthes L.* (Agavaceae) in Mexico: using niche modeling to improve assessments of risk status. *Biodivers. Conserv.* 16, 1885–1900. doi: 10.1007/s10531-006-9091-0

Solano, E., Terrazas, T., and González-Becerril, A. (2013). Comparative anatomy of the stem, leaf and inflorescence basal axis of *Polianthes L.* (Asparagaceae, Agavoideae) species. *Feddes Repertorium* 124, 105–115. doi: 10.1002/fedr.201300017

Stamatakis, A. (2014). RAxML versión 8: a tool for phylogenetic análisis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Starr, G. D. (1997). A revision of the genus *Hesperaloe* (Agavaceae). *Madroño* 44, 282–296.

Starr, G. D., Etter, J., and Kristen, M. (2018). *Agave cremnophila* (Agavaceae), a New Species from Southeastern Oaxaca Mexico. *Cactus Succulent J.* 90, 39–45. doi: 10.2985/015.090.0105

Tidwell, W. D., and Parker, L. R. (1990). *Protoyucca shadishii* gen. et sp. nov., an arborescent monocotyledon with secondary growth from the middle Miocene of northwestern Nevada, USA. *Rev. Palaeobot. Palynol.* 62, 79–95. doi: 10.1016/0034-6667(90)90018-e

Trejo-Salazar, R. E., Eguiarte, L. E., Suro-Piñera, D., and Medellin, R. A. (2016). Save our bats, ¿save our tequila: industry and science join forces to help bats and agaves. *Nat. Areas* 4, 523–530.

Trejo-Salazar, R. E., Scheinvar, E., and Eguiarte, L. E. (2015). ¿Quién poliniza realmente a los agaves? Diversidad de visitantes florales en 3 especies de *Agave* (Agavoideae: Asparagaceae). *Rev. Mexicana Biodiversidad* 86, 358–369.

Van der Niet, T., and Johnson, S. D. (2012). Phylogenetic evidence for pollinator-driven diversification of angiosperm. *Trends Ecol. Evol.* 27, 356–361.

Wikstrom, N., Savolainen, V., and Chase, M. W. (2001). Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. Biol. Sci.* 268, 2211–2220. doi: 10.1098/rspb.2001.1782

Zamora-Abrego, J. G., Manríquez-Morán, N. L., Ortíz-Yusty, C. E., and Ortega-León, A. M. (2013). "Uso de técnicas moleculares como herramienta para conservar la diversidad biológica," in *en Biología Molecular Aplicada a la Producción Animal y la Conservación de Especies Silvestres*, ed. H. A. López (Medellín: Universidad Nacional de Colombia).

Zhang, D., and Zheng, Y. (2014). Molecular control of grass inflorescence development. *Ann. Rev. Plant Biol.* 65, 553–578. doi: 10.1146/annurev-arplant-050213-040104

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 08 January 2021
doi: 10.3389/fpls.2020.578338

Leaf Venation and Morphology Help Explain Physiological Variation in *Yucca brevifolia* and *Hesperoyucca whipplei* Across Microhabitats in the Mojave Desert, CA

Amber R. Jolly, Joseph Zailaa, Ugbad Farah, Janty Woojuh, Félicia Makaya Libifani, Darlene Arzate, Christian Alex Caranto, Zayra Correa, Jose Cuba, Josephina Diaz Calderon, Nancy Garcia, Laura Gastelum, Ivette Gutierrez, Matthew Haro, Monserrat Orozco, Jessica Lamban Pinlac, Andoni Miranda, Justin Nava, Christina Nguyen, Edgar Pedroza, Jennyfer Perdomo, Scott Pezzini, Ho Yuen and Christine Scoffoni**

Department of Biological Sciences, California State University, Los Angeles, Los Angeles, CA, United States

OPEN ACCESS

Edited by:
Karolina Heyduk, University of Hawaii, United States

Reviewed by:
Antonella Gori, University of Florence, Italy
Ricardo A. Marenco, National Institute of Amazonian Research (INPA), Brazil

***Correspondence:**
Amber R. Jolly
arjolly@uci.edu
Christine Scoffoni
cscoffo@calstatela.edu

Specialty section:
This article was submitted to Functional Plant Ecology, a section of the journal Frontiers in Plant Science

Received: *30 June 2020*
Accepted: *11 December 2020*
Published: *08 January 2021*

Citation:
Jolly AR, Zailaa J, Farah U, Woojuh J, Libifani FM, Arzate D, Caranto CA, Correa Z, Cuba J, Calderon JD, Garcia N, Gastelum L, Gutierrez I, Haro M, Orozco M, Pinlac JL, Miranda A, Nava J, Nguyen C, Pedroza E, Perdomo J, Pezzini S, Yuen H and Scoffoni C (2021) Leaf Venation and Morphology Help Explain Physiological Variation in Yucca brevifolia and Hesperoyucca whipplei Across Microhabitats in the Mojave Desert, CA. Front. Plant Sci. 11:578338. doi: 10.3389/fpls.2020.578338

Different microclimates can have significant impact on the physiology of succulents that inhabit arid environments such as the Mojave Desert (California). We investigated variation in leaf physiology, morphology and anatomy of two dominant Mojave Desert monocots, *Yucca brevifolia* (Joshua tree) and *Hesperoyucca whipplei,* growing along a soil water availability gradient. Stomatal conductance (g_s) and leaf thickness were recorded in the field at three different sites (north-western slope, south-eastern slope, and alluvial fan) in March of 2019. We sampled leaves from three individuals per site per species and measured in the lab relative water content at the time of g_s measurements, saturated water content, cuticular conductance, leaf morphological traits (leaf area and length, leaf mass per area, % loss of thickness in the field and in dried leaves), and leaf venation. We found species varied in their g_s: while *Y. brevifolia* showed significantly higher g_s in the alluvial fan than in the slopes, *H. whipplei* was highest in the south-eastern slope. The differences in g_s did not relate to differences in leaf water content, but rather to variation in number of veins per mm^2 in *H. whipplei* and leaf width in *Y. brevifolia*. Our results indicate that *H. whipplei* displays a higher water conservation strategy than *Y. brevifolia*. We discuss these differences and trends with water availability in relation to species' plasticity in morphology and anatomy and the ecological consequences of differences in 3-dimensional venation architecture in these two species.

Keywords: 3D venation, chaparral yucca, Joshua tree, LMA, monocots, succulence

INTRODUCTION

Plants have different physiological thresholds to survive drought and understanding these thresholds can help predict community structure as climate change persists (Choat et al., 2018). The structure and function of plant populations are greatly determined by the environment, especially in deserts, a quickly growing biome (Cloudsley-Thompson, 1978; Archer and Predick, 2008) where

temperature and moisture limitations fluctuate drastically (St. Clair and Hoines, 2018). Plants native to arid ecosystems have adapted special anatomical, morphological and physiological traits, enabling them to survive and thrive under conditions of water scarcity and extreme temperatures. The Mojave Desert is dominated by winter precipitation making it an ideal home for the iconic *Yucca brevifolia*, commonly known as the Joshua tree (Archer and Predick, 2008). However, the standing groves of the *Y. brevifolia* and another important Mojave Desert succulent, *Hesperoyucca whipplei,* are threatened because of the ongoing major migration of their ideal habitats due to climate change (Wells and Woodcock, 1985; Cole et al., 2011).

Y. brevifolia is a C3 monocotyledonous evergreen species endemic to the Mojave Desert of southwest California, Nevada, Utah and Arizona. It typically resides at the elevations ranging from 1,000 to 2,000 m (Brittingham and Walker, 2000). *Y. brevifolia* grows on average 3.7 cm per year and has a median life expectancy of about 90 years (Gilliland et al., 2006), but little is known of the physiology of this species. Previous studies investigated rosette and flower panicle orientations and found that both were predominately oriented toward the south to maximize solar radiation (Warren et al., 2016) and these south facing rosettes tend to have higher stomatal conductance (Rasmuson et al., 1994).

However, a lack of knowledge still exists of the hydraulic properties of this species and how these relate to anatomy. *Y. brevifolia* has a discontinuous distribution in the Mojave Desert, with highest population density on well-drained alluvial fans adjacent to desert mountain ranges (Cole et al., 2011). Because of this narrow range of abiotic climate conditions, *Y. brevifolia* populations have been shifting to higher elevations with climate change (Harrower and Gilbert, 2018). Indeed, the elevated temperatures consequent to climate change result in shorter frost seasons, and thus, seasonal water availability, causing the reduction and range shift of *Y. brevifolia* populations (Archer and Predick, 2008). Differences in shifts occurs throughout the entire range of this species. Indeed, the loss of the distribution of *Y. brevifolia* has been suggested in primarily its southern and central ranges, highlighting potential habitat expansion in the north and east though the occupation of these new areas are noted to be unlikely to happen via natural seed dispersal (Cole et al., 2011). Additionally, the northern range expansion could be limited by factors such as invasive grasses which may intensify wildfires (Sweet et al., 2019). Increase in fires in the northern range could further limit *Y. brevifolia*'s expansion there, as a study at the southern range found that increased fires due to prolonged droughts increased mortality of *Y. brevifolia*, especially in the younger individuals (Defalco et al., 2010). Finally, the distribution of the mutualistic yucca moths (*Tegeticula synthetica* and *Tegeticula antithetica*) must match the expansion into refugia regions (Sweet et al., 2019), which is dependent on the elevation gradient (Harrower and Gilbert, 2018). A finer scale study, focused on the southern range of *Y. brevifolia* found that under a 3°C temperature increase scenario, 90% of the present distribution would be reduced (Barrows and Murphy-Mariscal, 2012). Warming temperatures also negatively impacted seed germination (Keeley and Meyers, 1985), which could further exacerbate the population decline of this species with climate change. These alarming findings point to the urge to better characterize the physiology and anatomy of this species in hopes to better understand how it might acclimate to warmer and drier environments with climate change. Here, we explore physiological and anatomical variation in *Y. brevifolia* across microhabitats varying in soil composition and moisture.

Hesperoyucca whipplei, commonly known as the chaparral yucca, is a C3 species more widely distributed than *Y. brevifolia. H. whipplei* extends north from San Diego into the Mojave Desert of Southern California and is discontinuous throughout its range (Haines, 1941; Aker, 1982; Davis et al., 1994). Contrary to *Y. brevifolia*, *H. whipplei* populations usually occur in shallow soil or rocky outcrops and inhabit elevations from sea level up to 2,450 m (Haines, 1941). Little is known about the hydraulic properties of *H. whipplei* and how this species may respond to a changing climate. Previous work has mapped the historic extant of *H. whipplei* in the Death Valley region and found that there is evidence that its distribution has diminished over time there possibly due the increase in summer temperature and aridity (Wells and Woodcock, 1985). However, a study comparing seedlings of *Y. brevifolia* and *H. whipplei* showed that when grown under elevated CO_2, *H. whipplei* was able to maintain photosynthesis despite a decline in g_s at high temperatures, whereas *Y. brevifolia* was more sensitive to high temperatures, experiencing a decline in both photosynthesis and g_s (Huxman et al., 1998). These results suggest that *H. whipplei* might have greater potential to acclimate to climate change. But these differences in physiology were never investigated in the field across microhabitats varying in soil moisture.

We investigated the leaf physiological, morphological, and anatomical properties of these two ecologically important desert succulents, *Y. brevifolia* and *H. whipplei,* growing in three different microhabitats in the West Mojave Desert: a north-western slope, south-eastern slope, and an alluvial fan. Alluvial fans are known to have deep, sandy soils with high moisture and nutrient levels, while water run-off areas such as slopes consist of shallow, rocky soils with poor nutrient levels and water retention (Brooks, 1999). We hypothesize that the highest g_s would be observed at sites most representative of the species preferred habitat: *Y. brevifolia* at the alluvial fan and *H. whipplei* at the south-eastern slope (i.e., the site with least water and nutrient retention as a result of soil topography and the increased potential evapotranspiration due to the sun's diurnal trajectory). We also expect these physiological differences to be constrained by morpho-anatomical traits across sites, with individuals with lower vein density and thinner leaves displaying lower g_s values.

MATERIALS AND METHODS

Study Species and Site Description

Data was collected in March of 2019 (March 10, 15, and 24) () on two desert succulent species (*Yucca brevifolia* and *Hesperoyucca whipplei*) from at the Puma Canyon Ecological Reserve (34°24′22.6″N 117°37′11.3″W), from three sites differing in soil moisture and composition (n = 3–8 individuals per species per

site; **Table 1**). Our sites consisted of a north-western slope (NW) and a south-eastern slope (SE) exhibiting rocky soils, separated by an alluvial fan with sandy soil (AF) (**Figure 1**). Elevation across sites ranged from 1,300 to 1,400 m. To characterize the environmental conditions in the three microhabitats, we measured air temperature and vapor-pressure deficit (VPD) throughout the days of measurement with a Delta-T AP4 Leaf Porometer (Dynamax, TX, United States). We quantified differences in soil moisture at 20–30 cm below the surface with a soil moisture meter at six different locations per site, with three measurements per locations (dry/wet index from 1–10; Dr. meter Hygrometer Moisture Sensor, Taiwan). Notably, no variation in either air temperature or VPD was recorded across sites during sampling days (**Table 1**). Soil moisture significantly varied across sites, with the alluvial fan being significantly moister than the south-eastern slope (**Table 1**).

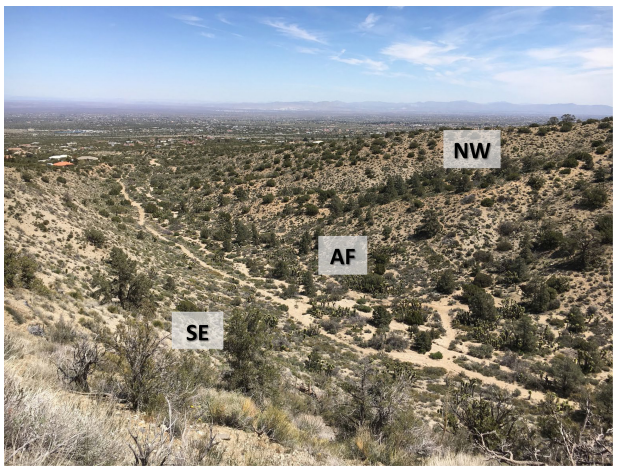


FIGURE 1 | Study sites in the West Mojave Desert. Overview of the three sites (NW, north-western slope; AF, alluvial fan; SE, south-eastern slope) in Puma Canyon in Piñon Hills, California. *n* = 3–8 individuals per species per site.

Physiological Traits

Individuals from *Y. brevifolia* and *H. whipplei* were chosen 2 and 1 m apart, respectively, to avoid sampling clonal sprouts. Stomatal conductance (g_s) was measured in the field between approximately 10:00 and 15:00 h on the abaxial surface of 3–5 leaves from each of 6–8 individuals per species and site using a Delta-T AP4 Leaf Porometer (Dynamax, TX, United States). We tested whether g_s measurements were influenced by the time of measurement and by the VPD at the time of measurement by plotting g_s against each one of these variables at each site for each species (**Supplementary Figures S1**, **S2**). We found no significant correlations, except for *H. whipplei* that displayed a negative correlation between time and g_s in the AF, and a positive correlation between g_s and VPD in the AF and SE slope. Notably, the g_s measurements taken at a later time of day were also those with very low VPD. Such conditions would likely induce stomatal closure. We thus decided to remove those data points from the analysis. We note that our results were not affected by the inclusion or exclusion of these points.

We collected 21 leaves (three leaves from five individuals; including some for which we had measured g_s on) per species per site from the top third of the rosette and rapidly placed them in a Whirlpak bag (Nasco, GA, United States). Bags were exhaled into to increase relative humidity and CO_2 concentration, causing stomata to close and preventing further water loss from the leaf samples. Individual bags were then placed in large black plastic bags filled with moist paper towels to further prevent sample dehydration, placed in a cooler and transported to the laboratory within a few hours. Once in the lab, 15 leaves from three individuals were removed from the bags and we measured both leaf mass using an analytical balance (±10 μg; model XS205; Mettler Toledo) and leaf thickness (using calipers; ±0.01 mm; Fowler). These leaves and remaining samples were then rehydrated overnight in beakers with ultrapure water, covered in plastic bags filled with wet paper towels to ensure high relative humidity around the rehydrating leaves. From these rehydrated leaves, the minimum epidermal conductance (g_{min}) was obtained from six, and the 15 leaves measured for initial mass and thickness were remeasured the next day after rehydration. We placed 10 of these leaves in a drying oven at 75°C, and kept 5 leaves from three individuals for anatomical measurements (see below). After a minimum of 3 days in the oven, leaves were measured for dry mass. We calculated the relative water content for each leaf using Eq. (1):

$$RWC = \frac{(leaf\ mass\ in\ field - leaf\ dry\ mass)}{rehydrated\ leaf\ mass - leaf\ dry\ mass} * 100 \quad (1)$$

Saturated water content was calculated using Eq. (2):

$$SWC = \frac{dry\ leaf\ mass}{rehydrated\ leaf\ mass - leaf\ mass\ in\ the\ field} \quad (2)$$

The percent loss thickness (PLT) of leaves collected in the field was calculated using Eq. (3):

$$\mathrm{PLT} = 100 - \left(average\ initial\ thickness * \frac{100}{average\ rehydrated\ thickness}\right) \quad (3)$$

The minimum epidermal conductance, also referred to as cuticular conductance (g_{min}) was measured using the

TABLE 1 | Environmental conditions (air temperature, soil moisture, and vapor-pressure deficit) at the three sites in Puma Canyon Ecological Reserve.

	Air temperature (°C)	Soil moisture (1–10 index)	Vapor-pressure deficit (kPa)
South-eastern slope	6.64 ± 0.21 (a)	2.31 ± 0.13 (b)	1.11 ± 0.008 (a)
Alluvial fan	7.19 ± 0.27 (a)	3.22 ± 0.36 (a)	1.12 ± 0.009 (a)
North-western slope	6.76 ± 0.17 (a)	2.31 ± 0.22 (ab)	1.12 ± 0.009 (a)
One-way ANOVA (site)	$p = 0.751$	$p = 0.019$	$p = 0.999$

Letters in parenthesis indicate significant differences across means at p = 0.05 (n = 9–49).

"clothesline" method (protocol from Sack and Scoffoni, 2011), based on the decline in mass through vapor diffusion of the leaf epidermis. Two leaves from three individuals per site and species were collected in the field as described above. In the lab, leaves were rehydrated overnight. The following day, g_{min} was measured as described in published protocol (Sack and Scoffoni, 2011).

Leaf Morphological Traits

From the 15 leaves collected described above, maximum thickness was obtained from the middle of each leaf using a caliper (±0.01 mm; Fowler) the day after leaves were rehydrated, along with leaf area and leaf length. Individual leaves were photographed and analyzed for area, width, and length in ImageJ (version 1.52a[1]).

To account for the 3D nature of the leaves, leaf surface area (n = 9–10 leaves per species per site) was measured using leaf geometry following previous literature (Huxman et al., 1998), considering 2/3 of the leaf length as a rectangle and 1/3 of the leaf length as a triangle (**Supplementary Figure S3**). To account for leaf tapering, the rectangle calculation only included leaf thickness and width from the top and middle of the leaves, while the triangle calculation only included the leaf thickness and width from the top. All surfaces of the leaf were considered. Area of the top and bottom of the rectangle (representing the adaxial and abaxial surfaces of the bottom two thirds of the leaf) were calculated as follows:

$$Area\ of\ top\ and\ bottom\ rectangles = 2\left(\frac{2}{3} * total\ leaf\ length * average\ leaf\ thickness\right) \quad (4)$$

To measure the area of the left and right sides of the rectangle (representing the sides of the bottom two thirds of the leaf), we used Eq. (5):

$$Area\ of\ left\ and\ right\ rectangles = 2\left(\frac{2}{3} * total\ leaf\ length * average\ leaf\ width\right) \quad (5)$$

To consider all four surfaces of the triangle, the top and bottom (representing the adaxial and abaxial surfaces of the top third of the leaf) were calculated as follows:

$$Area\ of\ top\ and\ bottom\ rectangles = 2\left(\frac{1}{2} * \frac{total\ leaf\ length}{3} * average\ leaf\ width\right) \quad (6)$$

The remaining two faces of the triangle (the sides of the top third of the leaf) were calculated using Eq. (8):

$$Area\ of\ triangles\ on\ left\ and\ right\ side = 2\left(\frac{1}{2} * \frac{total\ leaf\ length}{3} * average\ leaf\ thickness\right) \quad (7)$$

We summed all equations (Eqs. 4–8) to calculate total leaf surface area (cm^2) as follows:

$$Leaf\ surface\ area = Eq.\ 4 + Eq.\ 5 + Eq.\ 6 + Eq.\ 7 \quad (8)$$

[1]https://imagej.nih.gov/ij/

Leaf mass per area (LMA; g m^{-2}) was determined by dividing the dry mass by the leaf surface area.

Leaf Anatomical Traits

Hand cross sections were made using double-edge razor blades from 5 leaves from 3 to 5 individuals, from both species and all three sites, and mounted to slides and scanned (flatbed scanner; Canon Scan Lide 90, 1,200 pixels per inch). The total number of vascular bundles and leaf cross-sectional area were measured using ImageJ. Each trait was measured twice by a different person to ensure accuracy and repeatability in measurements. No significant differences were found across the datasets obtained from the two different individuals (paired t-test, $p > 0.05$). Because of the 3-dimensional (3D) vein architecture that these species display (**Figure 2**), inter-vein density (IVD; μm) was measured using ImageJ (n = 5 leaves per species per site) by selecting a central vein and then measuring the distance to the nearest neighboring veins (number of neighboring veins ranging from 4 to 8 across samples and species). Inter-vein density was calculated as the average distance between the central vein and the neighboring veins. Full leaves were also scanned to measure leaf length using Image J. Because of the 3D arrangement of veins throughout the leaf length, it was not possible to get an accurate measurement of vein length per area (VLA) by clearing leaves. Instead, we cut and scanned (flatbed scanner; Canon Scan Lide 90, 1,200 pixels per inch) six leaf cross sections throughout the leaf length and counted the number of veins in each section. We also measured the cross-sectional area to determine the number of veins per area (mm^2) as follows:

$$Number\ of\ veins\ per\ mm^2 = \frac{number\ of\ veins\ per\ section}{cross\ sectional\ area} \quad (9)$$

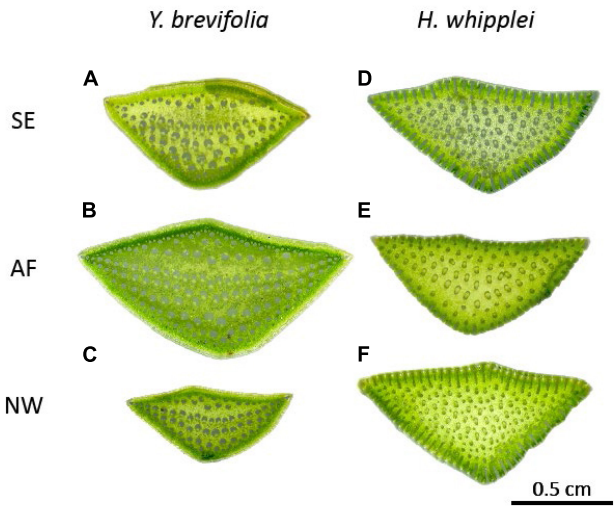


FIGURE 2 | Leaf cross-sections of *Y. brevifolia* and *H. whipplei* from different microhabitats in the West Mojave Desert. *Y. bevifolia* **(A–C)** and *H. whipplei* **(D–F)** cross sections showing vascular bundles. SE, south-eastern slope; AF, alluvial fan; NW, north-western slope. n = 5 leaves per species per site.

The number of veins per mm^2 was calculated for each of the six cross-sectional sections per leaf (n = 5). We then averaged values across sections and leaves and site for each species.

Statistical Analyses

To test for physiological, morphological and anatomical variation across sites and species, we performed a *two-way* ANOVA followed by Tukey-Kramer *post hoc* test in R (R 3.5.3[2]). Prior to all statistical analyses, we log-transformed data to improve normality and heteroscedasticity (Sokal and Rohlf, 1995). We tested for a correlation between leaf width and stomatal conductance using organized least squares (in SMATR; Sokal and Rohlf, 1995; Warton et al., 2006), and for correlations between g_s and time of g_s measurement, and g_s and VPD using Pearson's coefficient.

RESULTS

Physiological Differences Across Sites and Species

We found significant variation in stomatal conductance (g_s) across species and sites. The g_s varied from 96.75 ± 2.9 mmol m^{-2} s^{-1} in the AF to 57.0 ± 2.7 mmol m^{-2} s^{-1} in the SE slope in *Y. brevifolia*, and from 78.3 ± 4.9 mmol m^{-2} s^{-1} to 53.2. ± 3.32 mmol m^{-2} s^{-1} in the SE and NW slope, respectively, in *H. whipplei* (**Figure 3**). We found a significant species-site interaction (*two-way* ANOVA; p < 6.29e-11), with *Y. brevifolia* growing at the AF having the highest g_s (**Figure 3**). *H. whipplei* individuals growing in the SE slope exhibited higher g_s than individuals in the AF and NW slope (**Figure 3**).

Relative water content did not differ across sites (*two-way* ANOVA; p = 0.262), and no species-site interaction was found (*two-way* ANOVA; p = 0.342). However, we found a significant difference in RWC between species (*two-way* ANOVA; p = 5.65e-8), with RWC in *Y. brevifolia* being on average 8% lower than in *H. whipplei* (**Figure 4**). No significant differences across sites or species were found in SWC (*two-way* ANOVA; p < 0.05).

The minimum epidermal conductance did not differ between species (*two-way* ANOVA; p = 0.2745), and no significant species-site interaction was found (*two-way* ANOVA; p = 0.0857). However, a significant site effect was observed with g_{min} being the highest in individuals growing at the SE slope (*two-way* ANOVA; p = 0.0433) (**Supplementary Table S1**).

Morphological and Anatomical Differences Across Sites and Species

Leaf area was significantly different across species at each site (*two-way* ANOVA; p < 0.04), with *H. whipplei* leaves on average 2.2-fold larger than *Y. brevifolia* (average leaf area across sites = 98.9 ± 7.8 cm^2 in *H. whipplei* and 45.3 ± 4.5 cm^2 in *Y. breviolifa*; *two-way* ANOVA; p < 0.001; **Supplementary Table S1**). Leaf mass per area (LMA) displayed significant differences across species and species-site interaction (*two-way ANOVA*; p = 5.32e^{-4} — 1.93e^{-6}, respectively), though these differences were driven by differences observed between the two species at the NW site, with *H. whipplei* displaying 1.7-fold higher LMA than *Y. brevifolia* (62.3 ± 5.2 vs. 36.1 ± 2.2 g m^{-2}; **Supplementary Table S1**).

Leaf width varied significantly across species, sites and in their interaction (*two-way* ANOVA; p = 0.003). Leaf width was larger on average in *Y. brevifolia*, and highest at the AF (**Figure 5**). However, no significant differences were found in leaf width across sites in *H. whipplei* (*two-way* ANOVA; p > 0.05, **Figure 5**). A significant positive correlation was found between leaf width and *Y. brevifolia* across sites (**Figure 6**), but not across individuals of *H. whipplei* (p > 0.05).

Leaf thickness was on average higher in *H. whipplei* than *Y. brevifolia* across sites (*two-way ANOVA*; p = 0.008; **Figure 7**), and a significant site and species-site interaction was found (*two-way ANOVA*; p < 0.001), with the highest value achieved by *H. whipplei* (5.25 ± 0.22 mm) in the NW slope, and smallest by *Y. brevifolia* (4.19 ± 0.15 mm) growing in the SE slope (**Supplementary Table S1**). No differences were found in leaf PLT across species, nor in site-species interaction (*two-way* ANOVA; p > 0.05). However, a significant site effect was found (*two-way* ANOVA; p = 0.003), with the NW slope showing 7–8% greater thickness shrinkage at midday than the AF and SE slope (**Supplementary Table S1**).

The number of veins per mm^2 significantly varied across species and in their interaction with sites (*two-way* ANOVA; p < 0.0342; **Figure 8**). While *Y. brevifolia* showed no significant variation in the number of veins per mm^2 across sites (3.02 ± 0.22 mm^2 on average across sites), *H. whipplei* was highest in the SE slope (5.35 ± 0.2 mm^2) and lowest in the AF (2.2 ± 0.4 mm^2; **Figure 8**). No significant differences across sites or species were found in IVD (*two-way* ANOVA; p > 0.05) and averaged 307 ± 7.5 μm across sites and species.

DISCUSSION

Our results suggest that *H. whipplei* exhibits a higher water conservation strategy than *Y. brevifolia*, which performed better physiologically in the moister site. Indeed, we found variation in stomatal conductance (g_s) across sites and species with highest values exhibited at the species' preferred habitat (Haines, 1941; Cole et al., 2011). *Y. brevifolia* displayed higher g_s in the alluvial fan while *H. whipplei* displayed higher g_s in the shallow and exposed soils of the south-eastern slope. Notably, a recent meta-analysis of 54 case studies on the variation in gas exchange, carbon isotopes and water use efficiency showed that individuals were physiologically more efficient at their ecological optimum than when growing at their ecological range edge (Abeli et al., 2020). While our study did not directly measure the variation in stomatal conductance across the entire range of the *Y. brevifolia* and *H. whipplei*, our findings that species' physiological maxima were found at the sites that best

[2]https://www.r-project.org/

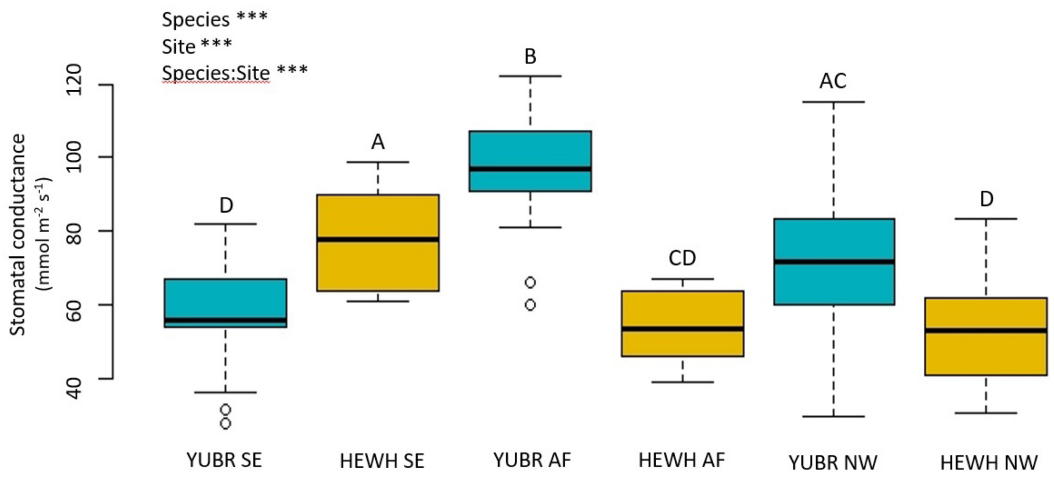


FIGURE 3 | Stomatal conductance (g_s) varied significantly across species and sites. Different letters denote statistically significant differences across species and sites ($p < 0.05$) by *two*-way ANOVA followed by a Tukey-Kramer *post hoc* test. YUBR = *Y. brevifolia*; HEWH = *H. whipplei*; SE, south-eastern slope; AF, alluvial fan; NW, north-western slope. n = 23–25 leaves per species per site. $^{***}p < 0.001$; $^{NS}p > 0.05$.

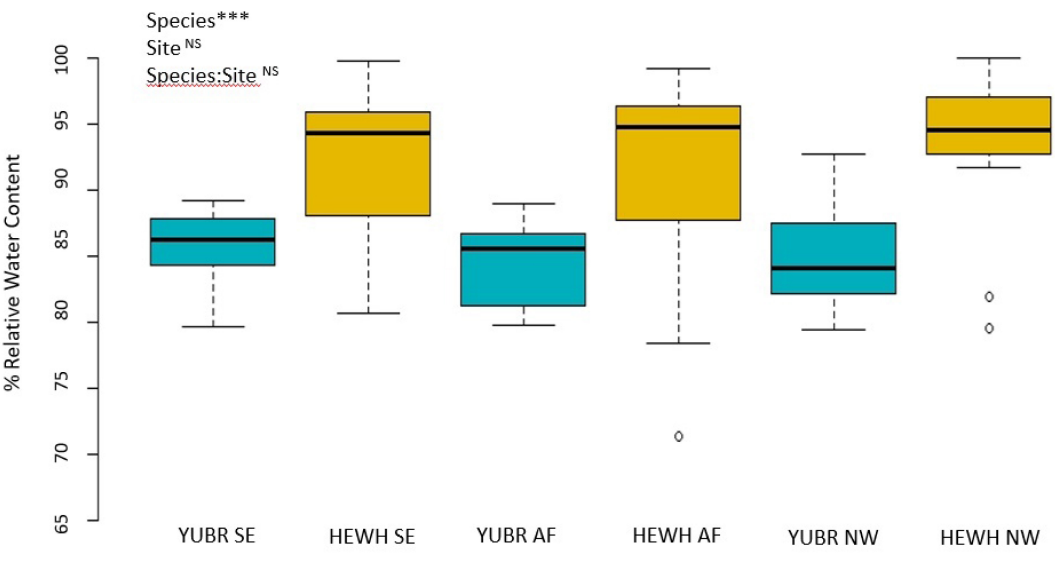


FIGURE 4 | Relative water content (RWC) did not differ across sites. Results from *two-way* ANOVA are shown in the top left corner. YUBR = *Y. brevifolia*; HEWH = *H. whipplei*; SE, south-eastern slope; AF, alluvial fan; NW, north-western slope. n = 13–20 leaves per species per site. $^{***}p < 0.001$; $^{NS}p > 0.05$.

reflected their preferred habitat could suggest stomatal limitation as a potential driver of these succulent monocot population demographics in the Mojave deserts. Future studies are needed to test and confirm this hypothesis. Such studies would help improve our understanding of how population demographics are affected by climate change.

Causes of Stomatal Conductance Variation Across Sites and Species

What is causing the observed variation in g_s across sites within species? Our results indicate that the g_s variation is not driven by differences in either leaf water status as relative water content was similar across sites within species or VPD. A previous study on a different *Yucca* species growing in Colorado (*Y. glauca*) found that the high VPD and temperature values in the summer at midday led to a strong decrease in stomatal conductance and photosynthetic rates (Roessler and Monson, 1985). In contrast, we found for our two species that at the end of winter/start of spring, VPD values were low on average even at midday (**Table 1**), and no negative correlations between g_s and time of measurement or VPD were found (**Supplementary Figures S1**, **S2**). Future studies should focus on investigating physiological variation throughout the year to identify possible seasonal patterns across species and sites. Additionally, long term physiological surveys are critical

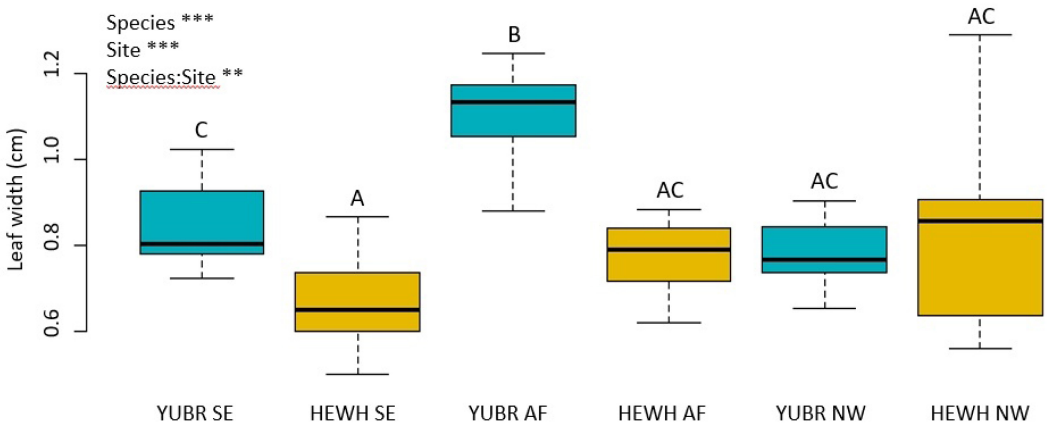


FIGURE 5 | Leaf width varied across species, sites, and had significant species-site interactions. Different letters denote statistically significant differences ($p < 0.05$) by single-factor ANOVA followed by a Tukey-Kramer *post hoc* test. YUBR = *Y. brevifolia*; HEWH = *H. whipplei*; SE, south-eastern slope; AF, alluvial fan; NW, north-western slope. n = 9–11 leaves per species per site. ***$p < 0.001$; **$p < 0.01$.

to understand the future trajectory of these responses especially under a changing climate.

The differences observed in g_s in our study are thus likely driven by underlying anatomical traits. While we were unable to quantify the stomatal density on neither leaves of *H. whipplei* nor *Y. brevifolia* due to the excessive epicuticular wax, we investigated variation in vein density which has been shown to correlate with stomatal density (Li et al., 2013). Water supply and demand are generally well correlated across diverse and closely related species (Scoffoni et al., 2016; Scoffoni and Sack, 2017). We found a coordination between the number of veins/mm^2 of cross-sectional area and g_s in *H. whipplei,* where leaves of individuals growing on the south-eastern slopes had greater number of veins/mm^2, associated to their higher g_s. This higher vein density could help supply water to more numerous stomata, and thus increase stomatal conductance at given water potential. The higher vein density at this site might also enable greater sugar transport if the higher stomatal conductance enabled greater photosynthetic rates, as has been observed across diverse and closely related species (Brodribb et al., 2007; Scoffoni et al., 2016). Indeed, an increase in gas exchange driven by anatomical differentiation in shallow, rocky soils with poor water retention could help explain how individuals of this species are so successful at those sites across their ecological range.

Number of veins/mm^2 could not explain the variation observed in g_s in *Y. brevifolia*, as our results showed that individuals did not differ significantly in that trait across sites. However, we found a strong correlation between leaf width and stomatal conductance across sites in *Y. brevifolia*. Indeed, leaf width was significantly higher in individuals growing at the alluvial fan in *Y. brevifolia* than in those growing on the slopes (**Figure 5**), matching the pattern observed for g_s in that species. Additionally, leaves of *Y. brevifolia* were significantly thicker in individuals growing at the AF compared to the SE slope. The thicker and wider leaves of *Y. brevifolia* at the AF could provide greater water storage, which could help maintain water supply to the transpiration stream and keep more stomata open at a given water potential.

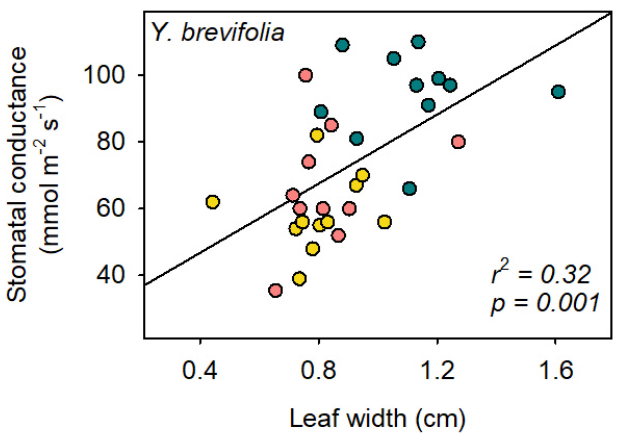


FIGURE 6 | Leaf width correlates with stomatal conductance in *Y. brevifolia*. The r^2- and p-value of the organized least square regression is provided in the panel. *Stomatal conductance* = 52 × *leaf width* + 26. Yellow, south-eastern slope; Blue, alluvial fan; Pink, north-western slope. n = 9–11 leaves per species per site.

Differences in other leaf morphological traits did not explain the variation observed in g_s at the three microhabitats for each species. Smaller leaves have thinner boundary layers which can lead to higher rates of transpiration and thus higher g_s, all else being equal. *H. whipplei* leaves were larger than *Y. brevifolia* but no site effect was found, suggesting that the variation in g_s may not be explained by leaf surface area. Leaf percent loss thickness coordinates with the shrinkage inside of the leaf cells and tissues (Scoffoni et al., 2014). While in most species higher rates of leaf shrinkage can play a role in triggering stomatal closure (Scoffoni et al., 2014), in succulent species an opposite mechanism was proposed (Scoffoni et al., 2014). Indeed, because of the greater proportion of water storage cells in these species,

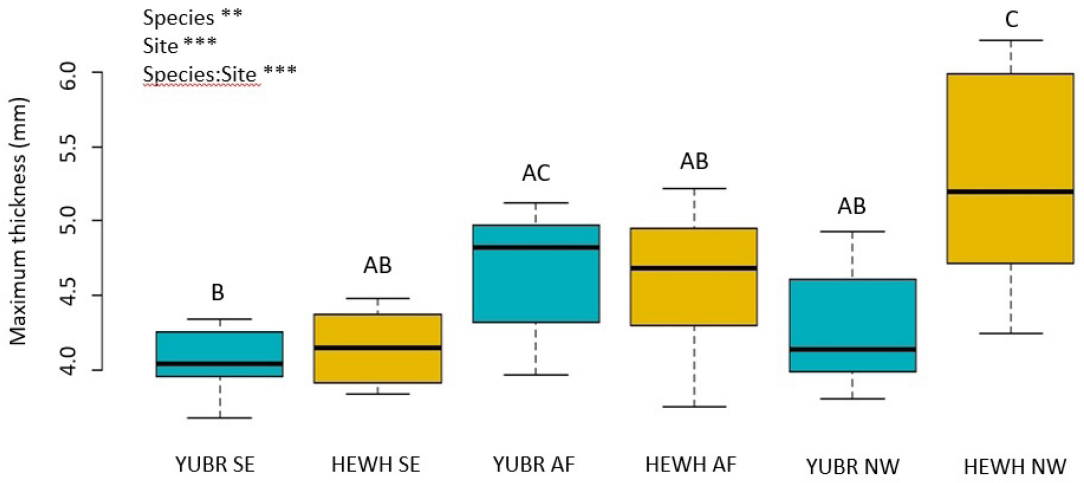


FIGURE 7 | Maximum leaf thickness varied significantly across sites and species. Different letters denote statistically significant differences ($p < 0.05$) by single-factor ANOVA followed by a Tukey-Kramer *post hoc* test. YUBR = *Y. brevifolia*; HEWH = *H. whipplei*; SE, south-eastern slope; AF, alluvial fan; NW, north-western slope. n = 7–10 leaves per species per site. ***$p < 0.001$; **$p < 0.01$.

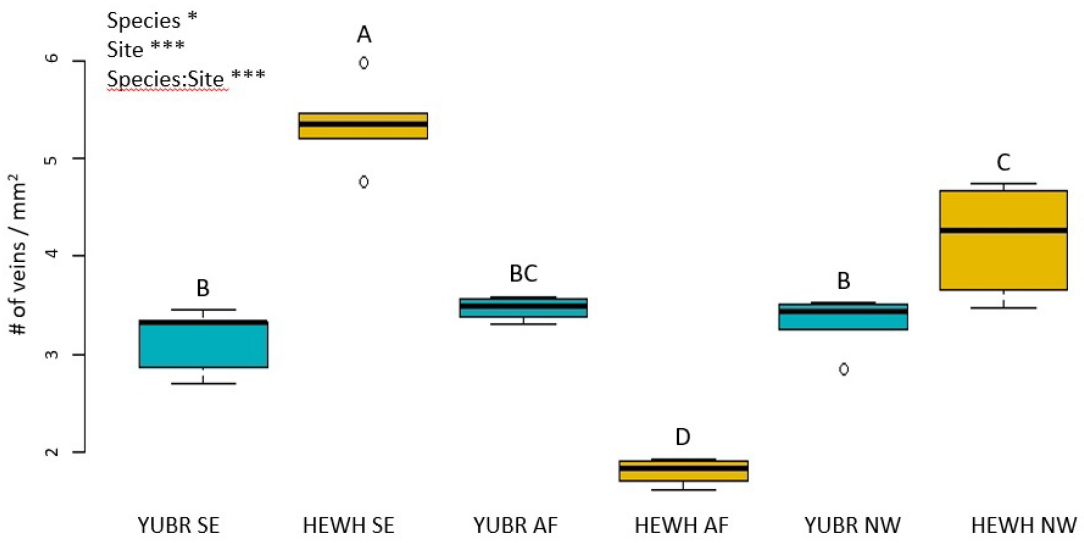


FIGURE 8 | Number of veins per mm^2 area varied across sites and had significant species-site interactions. Different letters denote statistically significant differences ($p < 0.05$) by single-factor ANOVA followed by a Tukey-Kramer *post hoc* test. YUBR = *Y. brevifolia*; HEWH = *H. whipplei*; SE, south-eastern slope; AF, alluvial fan; NW, north-western slope. n = 5 leaves per species per site. ***$p < 0.001$; *$p < 0.05$.

shrinkage might occur without causing physiological damage to the plant (Ogburn and Edwards, 2013) and contribute to the plant's ability to survive during drought. Here we found no significant differences in the percent loss of thickness in the field for neither of our species or in their interaction with sites. Our two species shrunk on average between 9.0 and 19.4% across sites, which is within the range of PLT values at turgor loss point observed across 14 diverse non succulent angiosperm species (average = 17.6 ± 2.43 %, ranging from 4.6% in *Raphiolepis indica* to 38% in *Lantana camara*; Scoffoni et al., 2014). These results suggest that our species were not suffering from drought stress at the time of measurement in the field. Notably, PLT was highest in individuals growing at the NW sites, which is also the site where individuals displayed greater g_{min}. A greater loss of water through the epidermis could potentially help explain greater shrinkage at that site.

Notably, other organs, such as roots, may have a role in explaining the differences observed in gas exchange across species and sites. Both *Y. brevifolia* and *H. whipplei* have contractile roots which are specialized to contract longitudinally (North et al., 2008). *Y. brevifolia* roots appear to play a major role in the rapid uptake and delivery of water as the reduced endodermal suberization of proximal roots and shallow root growth was shown to facilitate water uptake and allow access to

soil columns receiving maximum amounts of moisture (North and Baker, 2007). Future studies should focus on understanding plasticity in root hydraulics and morphology in these species and their impact on gas exchange. Finally, we collected data at the end of winter/start of spring to compare species across sites when water was least limiting and where they could thus exhibit highest rates of gas exchange. Future studies should focus on understanding the plasticity in gas exchange across sites throughout the year, and how much these vary in cumulative gas exchange throughout seasons using these different strategies.

Ecological Implications Behind the Morphoanatomical Drivers of the Physiological Variation Across Sites

The ability of *H. whipplei* to "adjust" its vein density (and potentially its stomatal density) to enable higher gas exchange might allow *H. whipplei* to thrive in the most arid micro-habitats compared to *Y. brevifolia*. Previous studies have found that vein density is increased in species of arid habitats. Indeed, a meta-analysis of 796 species showed a strong positive correlation between vein density and the aridity index (Sack and Scoffoni, 2013). Additionally, previous work on *Eucalyptus* and *Corymbia* species has shown that species of arid habitats tend to overinvest in leaf venation to compensate for leaf morphological constraints on photosynthesis (de Boer et al., 2016). This increase in vein density in arid habitats is argued to help achieve high photosynthetic rates after a rare rain event (Maximov, 1931; Grubb, 1998; Scoffoni et al., 2011; Sack and Scoffoni, 2013). Across its relatively wide ecological range throughout southern California, *H. whipplei* thrives in shallow soil or rocky outcrops, which are characterized by poor water retention, and thus greater aridity. Additionally, we observed bundle sheath extensions in leaves of *H. whipplei* across all three sites, though these appear more pronounced in individuals growing in the slopes than in the alluvial fan (**Figure 2**). Bundle sheath extensions have been shown to help guide light to the mesophyll (Karabourniotis et al., 2000; Nikolopoulos et al., 2002; Kenzo et al., 2007). In the thick succulent leaves of *H. whipplei*, the more pronounced bundle sheath extensions in individuals growing at the slopes could help increase light availability to the deeper mesophyll of the leaves, and thus engage more numerous cells in photosynthesis. Indeed, we observed that many minor veins are found throughout the cross-section in this species (**Figure 2**), and not simply at the ring layer below the epidermis like other species displaying 3D vein architecture typically show (see section "Three-Dimensional Venation Architecture in Our Desert Succulents" in the discussion below). Future studies should quantify the plasticity in bundle sheath extensions, and major and minor vein architecture in this species, and how these relate to gas exchange and resistance to drought.

Y. brevifolia on the other hand are typically found on well-drained alluvial fans (Cole et al., 2011), where water is more readily accessible than on slopes after a rain event. Our results indicate that unlike *H. whipplei*, the greater stomatal conductance observed in *Y. brevifolia* at the alluvial fan site was not driven by differences in vein densities. There could be a disconnect between stomatal and vein densities across sites in this species, and the extra supply needed to meet the evaporative demand could be provided by either differences in xylem anatomies (greater xylem conduit sizes would provide more efficient flow), or mesophyll pathways which could improve outside-xylem hydraulic conductance (Buckley, 2015). Notably, leaves of *Y. brevifolia* were significantly thicker and wider at the AF sites, with a significant correlation between leaf width and stomatal conductance across sites in that species. This coordination suggests that greater succulence and water storage could help buffer changes in vapor pressure deficit throughout the day and maintain open stomata without causing a decrease in leaf water potential. Future work should investigate plasticity in xylem and outside-xylem hydraulic anatomy of *Y. breviolia* across micro-habitats.

Three-Dimensional Venation Architecture in Our Desert Succulents

Our desert succulents display three-dimensional (3D) vein anatomy which, though extremely rare across species, has been hypothesized to have evolved in highly succulent leaves and other thick and/or rounded leaf species to help reduce the hydraulic path between veins and leaf cells that they supply, and thus improve photosynthesis (Ogburn and Edwards, 2013). Notably, the 3D venation that our two desert succulents display differs from the 3D arrangements that have been observed in other species. Indeed, to our knowledge, species from eight different plant families (Asparagaceae, Amaranthaceae, Asphodelaceae, Crassulaceae, Ellisochloa, Portulacineae, Proteaceae, and Molluginaceae) have previously been shown to display 3D venation with a midrib at the center of the leaf section, circled by a ring of higher order veins (Ogburn and Edwards, 2013; Heyduk et al., 2016), or one central layer of larger veins circled by a ring of higher order veins as observed in *Halothamnus auriculus* and in three other *Yucca* species (Kadereit et al., 2003; Heyduk et al., 2016). Here, our two Asparagaceae species display an extreme type of 3D venation (**Figure 2**): there are no central midrib or single layer of major veins circled by higher order veins, but rather major and minor veins are dispersed throughout the leaf cross-sections in 5–12 planes of veins in *H. whipplei* and 5–8 in *Y. brevifolia* (**Figure 2**). We note however that across sites, minor veins in *Y. brevifolia* appeared to be mainly located in the outer layers (near the epidermis) where the bulk photosynthesizing cells are found, whereas major veins were mainly distributed at the center of the leaf section and are surrounded by mostly water storage cells. However, in *H. whipplei* minor veins appear to be dispersed throughout the cross-section. *Y. brevifolia* has tighter and more forward-facing rosettes than *H. whipplei*, which could explain why veins appear more densely spaced on the sun exposed abaxial leaf surface (**Figure 2**). We hypothesize that the bundle sheath extensions in this species might allow light penetration deeper into the mesophyll and thus photosynthesis in cells throughout the cross-section. Future work quantifying

the detailed leaf cross-sectional anatomy is needed to test these hypotheses.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

AJ and CS wrote the manuscript with contributions from all authors. All authors contributed to the design of the study, data collection, and analysis.

FUNDING

This work was supported by the Department of Biological Sciences at California State University, Los Angeles, the Louis-Stokes Alliance for Minority Participation Bridge to the Doctorate program at California State University, Los Angeles funded by the National Science Foundation under grant #HRD-1807387, and the NSF-CAREER grant #231715.

ACKNOWLEDGMENTS

We thank Justine Curcio, John Harris, Wendie Marriott, Wendy Walker, Jeremy Yoder, and the Transition Habitat Conservancy staff for guidance in Piñon Hills.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.578338/full#supplementary-material

Supplementary Figure 1 | Testing for an effect of time of measurement on stomatal conductance in *Y. brevifolia* **(A–C)** and *H. whipplei* **(D–F)** growing in different micro-habitats in the West Mojave Desert. SE, south-eastern slope, AF, alluvial fan, and NW, north-western slope. n = 23–25 leaves per species per site. The Pearson correlation coefficient and p-value are provided in each panel.

Supplementary Figure 2 | Testing for an effect of vapor pressure deficit on stomatal conductance in *Y. brevifolia* **(A–C)** and *H. whipplei* **(D–F)** growing in different micro-habitats in the West Mojave Desert. SE, south-eastern slope, AF, alluvial fan, and NW, north-western slope. n = 23–25 leaves per species per site. The Pearson correlation coefficient and p-value are provided in each panel.

Supplementary Figure 3 | Modeled leaf surface area.

Supplementary Table 1 | Averages and raw data for the physiological and morpho-anatomical traits of the study.

REFERENCES

Abeli, T., Ghitti, M., and Sacchi, R. (2020). Does ecological marginality reflect physiological marginality in plants? *Int. J. Deal. Asp. Plant Biol.* 154, 149–157. doi: 10.1080/11263504.2019.1578278

Aker, C. L. (1982). Regulation of flower, fruit and seed production by a monocarpic perennial. *J. Ecol.* 70:357. doi: 10.2307/2259884

Archer, S. R., and Predick, K. I. (2008). Climate Change and Ecosystems of the Southwestern United States. *Rangelands* 30, 23–28. doi: 10.2111/1551-501X200830

Barrows, C. W., and Murphy-Mariscal, M. L. (2012). Modeling impacts of climate change on Joshua trees at their southern boundary: How scale impacts predictions. *Biol. Conserv.* 152, 29–36. doi: 10.1016/j.biocon.2012.03.028

de Boer, H. J., Drake, P. L., Wendt, E., Price, C. A., Schulze, E.-D., et al. (2016). Apparent overinvestment in leaf venation relaxes leaf morphological constraints on photosynthesis in arid habitats. *Plant Physiol.* 172, 2286–2299. doi: 10.1104/pp.16.01313

Brittingham, S., and Walker, L. R. (2000). Facilitation of *Yucca Brevifolia* recruitment by mojave desert shrubs. *West. North Am. Nat.* 60, 374–383.

Brodribb, T. J., Feild, T. S., and Jordan, G. J. (2007). Leaf maximum photosynthetic rate and venation are linked by hydraulics. *Plant Physiol.* 144, 1890–1898. doi: 10.1104/pp.107.101352

Brooks, M. L. (1999). Habitat invasibility and dominance by alien annual plants in the western mojave desert. *Biol. Invasions* 1, 325–337. doi: 10.1023/A:1010057726056

Buckley, T. N. (2015). The contributions of apoplastic, symplastic and gas phase pathways for water transport outside the bundle sheath in leaves. *Plant Cell Environ.* 38, 7–22. doi: 10.1111/pce.12372

Choat, B., Brodribb, T. J., Brodersen, C. R., Duursma, R. A., López, R., and Medlyn, B. E. (2018). Triggers of tree mortality under drought. *Nature* 558, 531–539. doi: 10.1038/s41586-018-0240-x

Clair, S. B. S., and Hoines, J. (2018). Reproductive ecology and stand structure of Joshua tree forests across climate gradients of the Mojave Desert. *PLoS One* 13:e0193248. doi: 10.1371/journal.pone.0193248

Cloudsley-Thompson, J. L. (1978). Human Activities and Desert Expansion. *Geogr. J.* 144, 416–423. doi: 10.2307/634818

Cole, K. L., Ironside, K., Eischeid, J., Garfin, G., Duffy, P. B., and Toney, C. (2011). Past and ongoing shifts in Joshua tree distribution support future modeled range contraction. *Ecol. Appl.* 21, 137–149. doi: 10.1890/09-1800.1

Davis, F. W., Stine, P. A., and Stoms, D. M. (1994). Distribution and conservation status of coastal sage scrub in southwestern California. *J. Veg. Sci.* 5, 743–756. doi: 10.2307/3235887

Defalco, L. A., Esque, T. C., Scoles-Sciulla, S. J., and Rodgers, J. (2010). Desert wildfire and severe drought diminish survivorship of the long-lived Joshua tree (*Yucca brevifolia*; Agavaceae). *Am. J. Bot.* 97, 243–250. doi: 10.3732/ajb.0900032

Gilliland, K. D., Huntly, N. J., and Anderson, J. E. (2006). Age and population structure of Joshua trees (*Yucca brevifolia*) in the northwestern Mojave Desert. *West. North Am. Nat.* 66, 202–208. doi: 10.3398/1527-0904200666

Grubb, P. J. (1998). A reassessment of the strategies of plants which cope with shortages of resources. *Perspect. Plant Ecol. Evol. Syst.* 1, 3–31. doi: 10.1078/1433-8319-00049

Haines, L. (1941). Variation in *Yucca whipplei*. *Madroño* 6, 33–45.

Harrower, J., and Gilbert, G. S. (2018). Context-dependent mutualisms in the Joshua tree–yucca moth system shift along a climate gradient. *Ecosphere* 9:e02439. doi: 10.1002/ecs2.2439

Heyduk, K., McKain, M. R., Lalani, F., and Leebens-Mack, J. (2016). Evolution of a CAM anatomy predates the origins of Crassulacean acid metabolism in the Agavoideae (Asparagaceae). *Mol. Phylogenet. Evol.* 105, 102–113. doi: 10.1016/j.ympev.2016.08.018

Huxman, T. E., Hamerlynck, E. P., Loik, M. E., and Smith, S. D. (1998). Gas exchange and chlorophyll fluorescence responses of three south-western Yucca species to elevated CO2 and high temperature. *Plant Cell Environ.* 21, 1275–1283. doi: 10.1046/j.1365-3040.1998.00396.x

Kadereit, G., Borsch, T., Weising, K., and Freitag, H. (2003). Phylogeny of Amaranthaceae and Chenopodiaceae and the Evolution of C 4 Photosynthesis. *Int. J. Plant Sci.* 164, 959–986. doi: 10.1086/378649

Karabourniotis, G., Bornman, J. F., and Nikolopoulos, D. (2000). A possible optical role of the bundle sheath extensions of the heterobaric leaves of Vitis vinifera

and Quercus coccifera. *Plant Cell Environ.* 23, 423–430. doi: 10.1046/j.1365-3040.2000.00558.x

Keeley, J. E., and Meyers, A. (1985). Effect of Heat on Seed Germination of Southwestern Yucca Species. *Southwest. Nat.* 30:303. doi: 10.2307/3670744

Kenzo, T., Ichie, T., Watanabe, Y., and Hiromi, T. (2007). Ecological distribution of homobaric and heterobaric leaves in tree species of Malaysian lowland tropical rainforest. *Am. J. Bot.* 94, 764–775. doi: 10.3732/ajb.94.5.764

Li, S., Zhang, Y.-J., Sack, L., Scoffoni, C., Ishida, A., Chen, Y.-J., et al. (2013). The Heterogeneity and Spatial Patterning of Structure and Physiology across the Leaf Surface in Giant Leaves of *Alocasia macrorrhiza*. *PLoS One* 8:e66016. doi: 10.1371/journal.pone.0066016

Maximov. (1931). The physiological significance of the xeromorphic structure of plants. *J. Ecol.* 1931, 272–282.

Nikolopoulos, D., Liakopoulos, G., Drossopoulos, I., and Karabourniotis, G. (2002). The relationship between anatomy and photosynthetic performance of heterobaric leaves. *Plant Physiol.* 129, 235–243. doi: 10.1104/pp.010943

North, G. B., and Baker, E. A. (2007). Water uptake by older roots: evidence from desert succulents. *HortScience* 42, 1103–1106. doi: 10.21273/HORTSCI.42.5.1103

North, G. B., Brinton, E. K., and Garrett, T. Y. (2008). Contractile roots in succulent monocots: convergence, divergence and adaptation to limited rainfall. *Plant Cell Environ.* 31, 1179–1189. doi: 10.1111/j.1365-3040.2008.01832.x

Ogburn, R. M., and Edwards, E. J. (2013). Repeated origin of three-dimensional leaf venation releases constraints on the evolution of succulence in plants. *Curr. Biol. CB* 23, 722–726. doi: 10.1016/j.cub.2013.03.029

Rasmuson, K., Anderson, J., and Huntly, N. (1994). Coordination of branch orientation and photosynthetic physiology in the Joshua tree (*Yucca brevifolia*). *Gt. Basin Nat.* 54:1994.

Roessler, P. G., and Monson, R. K. (1985). Midday depression in net photosynthesis and stomatal conductance in Yucca glauca: Relative contributions of leaf temperature and leaf-to-air water vapor concentration difference. *Oecologia* 67, 380–387. doi: 10.1007/BF00384944

Sack, L., and Scoffoni, C. (2013). Leaf venation: structure, function, development, evolution, ecology and applications in the past, present and future. *N. Phytol.* 198, 983–1000. doi: 10.1111/nph.12253

Sack, L.,and Scoffoni, C. (2011). Minimum epidermal conductance (gmin, a.k.a. cuticular conductance). *PrometheusWiki*. 35, 257–262.

Scoffoni, C., and Sack, L. (2017). The causes and consequences of leaf hydraulic decline with dehydration. *J. Exp. Bot.* 68, 4479–4496. doi: 10.1093/jxb/erx252

Scoffoni, C., Chatelet, D. S., Pasquet-Kok, J., Rawls, M., Donoghue, M. J., Edwards, E. J., et al. (2016). Hydraulic basis for the evolution of photosynthetic productivity. *Nat. Plants* 2:16072. doi: 10.1038/nplants.2016.72

Scoffoni, C., Rawls, M., McKown, A., Cochard, H., and Sack, L. (2011). Decline of leaf hydraulic conductance with dehydration: relationship to leaf size and venation architecture. *Plant Physiol.* 156, 832–843. doi: 10.1104/pp.111.173856

Scoffoni, C., Vuong, C., Diep, S., Cochard, H., and Sack, L. (2014). Leaf shrinkage with dehydration: coordination with hydraulic vulnerability and drought tolerance. *Plant Physiol.* 164, 1772–1788. doi: 10.1104/pp.113.221424

Sokal, R. R., and Rohlf, F. J. (1995). *Biometry: The Principles and Practice of Statistics in Biological Research*, 3 Edn. New York, NH: W.H. Freeman.

Sweet, L. C., Green, T., Heintz, J. G. C., Frakes, N., Graver, N., Rangitsch, J. S., et al. (2019). Congruence between future distribution models and empirical data for an iconic species at Joshua Tree National Park. *Ecosphere* 10:e02763. doi: 10.1002/ecs2.2763

Warren, S. D., Baggett, L. S., and Warren, H. (2016). Directional floral orientation in joshua trees (*Yucca brevifolia*). *West. North Am. Nat.* 76, 374–378. doi: 10.3398/064.076.0313

Warton, D. I., Wright, I. J., Falster, D. S., and Westoby, M. (2006). Bivariate line-fitting methods for allometry. *Biol. Rev. Camb. Philos. Soc.* 81, 259–291. doi: 10.1017/S1464793106007007

Wells, P. V., and Woodcock, D. (1985). *Full-glacial vegetation of Death Valley, California: juniper woodland opening to yucca semidesert. Madrono USA*. Available online at: http://agris.fao.org/agris-search/search.do?recordID=US8612110 (accessed on March 8, 2020)

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 15 January 2021
doi: 10.3389/fpls.2020.573767

Hybridization History and Repetitive Element Content in the Genome of a Homoploid Hybrid, *Yucca gloriosa* (Asparagaceae)

Karolina Heyduk[1,2], Edward V. McAssey[1,3,4], Jane Grimwood[5], Shengqiang Shu[6], Jeremy Schmutz[5,6], Michael R. McKain[7] and Jim Leebens-Mack[8]*

[1] School of Life Sciences, University of Hawai'i at Mānoa, Honolulu, HI, United States, [2] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, United States, [3] Department of Biological Sciences, Quinnipiac University, Hamden, CT, United States, [4] Department of Biology and Environmental Science, University of New Haven, West Haven, CT, United States, [5] HudsonAlpha Institute for Biotechnology, Huntsville, AL, United States, [6] Lawrence Berkeley National Laboratory, US Department of Energy Joint Genome Institute, Berkeley, CA, United States, [7] Department of Biological Sciences, University of Alabama, Tuscaloosa, AL, United States, [8] Department of Plant Biology, University of Georgia, Athens, GA, United States

OPEN ACCESS

Edited by:
Hanna Weiss-Schneeweiss, University of Vienna, Austria

Reviewed by:
Tony Heitkam, Technische Universität Dresden, Germany
Steven Dodsworth, University of Bedfordshire, United Kingdom

***Correspondence:**
Karolina Heyduk
heyduk@hawaii.edu

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *17 June 2020*
Accepted: *17 December 2020*
Published: *15 January 2021*

Citation:
Heyduk K, McAssey EV, Grimwood J, Shu S, Schmutz J, McKain MR and Leebens-Mack J (2021) Hybridization History and Repetitive Element Content in the Genome of a Homoploid Hybrid, Yucca gloriosa (Asparagaceae). Front. Plant Sci. 11:573767. doi: 10.3389/fpls.2020.573767

Hybridization in plants results in phenotypic and genotypic perturbations that can have dramatic effects on hybrid physiology, ecology, and overall fitness. Hybridization can also perturb epigenetic control of transposable elements, resulting in their proliferation. Understanding the mechanisms that maintain genomic integrity after hybridization is often confounded by changes in ploidy that occur in hybrid plant species. Homoploid hybrid species, which have no change in chromosome number relative to their parents, offer an opportunity to study the genomic consequences of hybridization in the absence of change in ploidy. *Yucca gloriosa* (Asparagaceae) is a young homoploid hybrid species, resulting from a cross between *Yucca aloifolia* and *Yucca filamentosa.* Previous analyses of ∼11 kb of the chloroplast genome and nuclear-encoded microsatellites implicated a single *Y. aloifolia* genotype as the maternal parent of *Y. gloriosa.* Using whole genome resequencing, we assembled chloroplast genomes from 41 accessions of all three species to re-assess the hybrid origins of *Y. gloriosa*. We further used re-sequencing data to annotate transposon abundance in the three species and mRNA-seq to analyze transcription of transposons. The chloroplast phylogeny and haplotype analysis suggest multiple hybridization events contributing to the origin of *Y. gloriosa*, with both parental species acting as the maternal donor. Transposon abundance at the superfamily level was significantly different between the three species; the hybrid was frequently intermediate to the parental species in TE superfamily abundance or appeared more similar to one or the other parent. In only one case—*Copia* LTR transposons—did *Y. gloriosa* have a significantly higher abundance relative to either parent. Expression patterns across the three species showed little increased transcriptional activity of transposons, suggesting that either no transposon release occurred in *Y. gloriosa*

upon hybridization, or that any transposons that were activated via hybridization were rapidly silenced. The identification and quantification of transposon families paired with expression evidence paves the way for additional work seeking to link epigenetics with the important trait variation seen in this homoploid hybrid system.

Keywords: homoploid, hybrid, *Yucca*, chloroplast, transposable element, genomic shock

INTRODUCTION

Hybridization between related species has the potential to generate novel genotypic and phenotypic combinations, sometimes resulting in the origin of new species. Understanding the factors that promote the process of hybridization, as well as the maintenance of newly created hybrids, has been of considerable interest to both the fields of ecology and evolution (Gross and Rieseberg, 2005). As the generation of biodiversity is of primary importance to evolutionary biology, many studies have sought to determine whether or not newly created hybrids are reproductively isolated from parental species and are capable of persisting in a hybrid state for many generations. The tools aimed at studying plant hybridization include observational studies of plants and their pollinators in the wild (Leebens-Mack and Milligan, 1998; Hersch and Roy, 2007), reciprocal transplant studies across multiple environments (Wang et al., 1997), manual pollinations between related species (Sun et al., 2018), cytogenetics (Thórsson et al., 2001), and population genomics (Bredeson et al., 2016). Hybridization can result in allopolyploid individuals, in which hybridization occurs at the same time as chromosome doubling, as well as homoploids, in which there is no change in chromosome number (for a review, see Rieseberg, 1997; Soltis and Soltis, 2009). Transposable element content and abundance has been hypothesized to contribute to genome dominance in allopolyploid species (Edger et al., 2017; Bird et al., 2018), but change in ploidy makes it difficult to assess its importance relative to hybridization in the genesis of a new species. Homoploid hybrid species provide an opportunity to focus on the effects of hybridization while controlling for ploidy level (Ungerer et al., 2009; Staton et al., 2012).

Investigation of hybridization almost always begins with a detailed understanding of the genetics and life history of the putative parental and hybrid species. In the case of wild sunflowers, numerous studies have focused on how *Helianthus annuus* and *H. petiolaris* have hybridized multiple independent times to form three homoploid hybrid species: *H. anomalus*, *H. deserticola*, and *H. paradoxus* (Rieseberg, 1991; Rieseberg et al., 2003). These hybrid species are morphologically distinct from their parents and each other (Rieseberg et al., 2003), display varying levels of salt tolerance (Welch and Rieseberg, 2002; Karrenberg et al., 2006), show gene expression differences (Lai et al., 2006), and exhibit population genetic patterns consistent with selective sweeps (Sapir et al., 2007). The repeated formation of homoploid hybrids in *Helianthus* has increased our understanding of hybrid speciation from both ecological and genomic perspectives, yet it is only one example of homoploid hybridization in flowering plants. Another well-studied example of homoploid hybridization is in *Iris nelsonii,* a hybrid suspected to have genetic contributions from more than two species based on patterns of both nuclear and plastid genetic variation (Arnold, 1993). The fitness of the hybrid species relatives to the parental species varies depending on the moisture of the environments, implying that genotype-by-environment interactions differentially affect parental and hybrid genotypes, a phenomenon that can lead to hybrid speciation (Johnston et al., 2001).

While hybridization's effect on the generation of biodiversity and the movement of adaptive traits between species has been well established, the effect on the genome is only recently being fully understood. McClintock (1984), described hybridization as a "challenge" or "shock" for the genome; the merger of two separate genomes in a single nucleus results in a completely novel genomic environment. Post hybridization, alleles once restricted to separate species now interact in a new cellular setting, allowing for the formation of novel phenotypes, epistatic interactions, and potentially significant and rapid evolutionary change. Possible outcomes of hybridization and subsequent genome shock include: alteration of gene expression (Hegarty et al., 2009; Xu et al., 2009); chromosomal rearrangements (Rieseberg et al., 1995; Lai et al., 2005; Danilova et al., 2017); genome dominance, in which one progenitor genome expresses and/or retains more genes (Rapp et al., 2009; Bardil et al., 2011; Schnable et al., 2011; Yoo et al., 2013; Edger et al., 2017; Bird et al., 2018); epigenetic perturbation (Salmon et al., 2005), which in turn can lead to a release of silencing of repetitive elements and allows for subsequent repeat proliferation (Ungerer et al., 2006; Parisod et al., 2009).

Repetitive elements in particular have been implicated in the divergence of hybrid species from their progenitors. For example, RNA-seq suggests that established homoploid hybrid sunflowers, as opposed to newly synthesized hybrids, have elevated transposon expression levels (Renaut et al., 2014). In two of these hybrid sunflower species fluorescent *in situ* hybridization studies identified expansions of *Gypsy* retrotransposons relative to the progenitor species (Staton and Ungerer, 2009). *Gypsy* and *Copia* elements are typically the most abundant superfamilies in plant genomes, and are both Class I retrotransposons that replicate via a "copy and paste" mechanism (Wessler et al., 1995), in contrast to the variety of Class II DNA transposons that replicate via a "cut and paste" mechanism (Feschotte and Pritham, 2007). Transposons can affect traits by disrupting genes, duplicating or re-organizing genes (Xiao et al., 2008), or they can land upstream and create new patterns of gene expression (Studer et al., 2011). The accumulation of transposons contributes to a large proportion of genome size variation seen in plants (Tenaillon et al., 2011), and ectopic recombination between transposable elements can result in

genomic deletions and are a major force in genome evolution (Devos et al., 2002).

While homoploid hybrid systems are relatively rare, recent efforts to sequence the genomes of *Yucca* (Asparagaceae) species allows us to investigate the effects of hybridization on a homoploid genome. *Yucca aloifolia* L. and *Yucca filamentosa* L. are emergent models in understanding the evolution of CAM photosynthesis, as the species use CAM and C_3, respectively (Heyduk et al., 2016). The two species also hybridize to form *Y. gloriosa* L. (Rentsch and Leebens-Mack, 2012), which is photosynthetically intermediate and a relatively recently derived homoploid hybrid species (Trelease, 1902). All three species have genome sizes of ∼2.8 Gb (Heyduk unpublished) and are sympatric in the Southeastern United States, with *Y. filamentosa* found across a broader range of the eastern seaboard, including into New England and the Midwest; *Y. aloifolia* is restricted largely to the Southeastern United States and reaches only as far north as North Carolina. *Yucca gloriosa* is even more restricted than either parent in its range, found only in the coastal dune systems of the Atlantic seaboard and, based on herbarium records, along the coast of the Gulf of Mexico. It is thought that *Y. aloifolia* was introduced into the Southeastern United States from Mexico or the Caribbean by Spanish colonists (Trelease, 1902; Groman and Pellmyr, 2000). Perhaps as a result of the human-involved introduction, *Y. aloifolia* has escaped the dependence on the obligate *Yucca*-yucca moth pollination mutualism and can be pollinated by the yucca moth *Tegeticula yuccasella* (Leebens-Mack and Pellmyr, 2004) or introduced generalist honeybees (*Apis mellifera*) (Rentsch and Leebens-Mack, 2014). *Yucca filamentosa* still retains its obligate pollination mutualism with the yucca moths (*Tegeticula yuccasella and T. cassandra*) (Pellmyr, 1999), and overlaps in flowering time with *Y. aloifolia* briefly and only in some years, suggesting that hybridization between the two species may be rare.

Previous work suggested no variation in chloroplast or microsatellite repeats in a small sampling of *Y. aloifolia* genotypes, and further indicated that *Y. aloifolia* is the maternal parent in any hybridization events that led to *Y. gloriosa* (Rentsch and Leebens-Mack, 2012). Through a whole genome sequencing project that aims to assemble the genomes of *Y. aloifolia* and *Y. filamentosa*, resequencing was performed on individuals of all three *Yucca* species. Using the resequencing data, we sought to re-test hypotheses on the number and direction of hybridization events in *Y. gloriosa*. Specifically, we assembled maternally inherited chloroplast genomes, which can inform not only the evolutionary history, but also the direction of hybridization. We assessed whether all hybrid Y. gloriosa individuals were nested within Y. aloifolia on a phylogenetic tree and haplotype network, consistent with the hypothesis of a single hybridization event with *Y. aloifolia* as the maternal parent. We further examined the repeat landscape of all three species to determine if repeat content in the hybrid is intermediate between the two parents, or if transgressive repeat abundance exists, suggesting a degree of post-hybridization genomic shock. Finally, using existing RNA-sequencing datasets in the three species of *Yucca*, we examined the activity of repeats using mRNA reads as a proxy. Through the use of high throughput genomic data, we find that *Y. gloriosa* is the result of repeated and bi-directional hybridization events that evidently led to minimal repeat proliferation. Our findings further suggest that there is little evidence of repetitive element release in *Y. gloriosa* as a result of hybridization.

MATERIALS AND METHODS

DNA Sampling, Library Preparation, and Sequencing

Clones of 41 individuals (5 from *Y. aloifolia*, 24 from *Y. gloriosa*, and 12 form *Y. filamentosa*) were collected throughout the Southeastern United States from 2013 to 2015 and planted in the University of Georgia greenhouse (**Figure 1** and **Supplementary Table 1**). In 2018, approximately 100 mg of fresh tissue was harvested from fully expanded leaves and kept on ice until DNA extraction, using a CTAB protocol with sorbitol addition that removes secondary compounds before DNA purification (Doyle, 1987; Štorchová et al., 2000). DNA was visualized on a 1.5% agarose gel to measure integrity and quantified via Qubit. Samples were shipped to the HudsonAlpha Institute for Biotechnology, where Illumina 350 basepair PCRfree fragment libraries were constructed using standard protocols. Each library was uniquely barcoded and sequenced on a NovaSeq 6000 with paired end 150bp reads. Data is available on the NCBI Sequence Read Archive (for a full list of SRA accessions, see **Supplementary Table 1**).

Chloroplast Genome Assembly and Analysis

Raw reads were first quality trimmed using Trimmomatic v 0.36 (Bolger et al., 2014). Due to the sheer size of the sequence data per individual—roughly 400–800 million reads—a subset of four million paired-end reads was randomly sub-sampled from each library's trimmed dataset in order to speed up computational analyses. The sub-sampled data were used as input into the program Fast-Plast[1], which assembles plastid genomes by first mapping reads to a reference plastid genome (here we used a previously assembled *Y. filamentosa* chloroplast genome; McKain et al., 2016).

Chloroplast genomes of *Agave americana* (NCBI accession: KX519714.1, Abraham et al., 2016) and *Nolina atopocarpa* (NCBI accession: NC_032708.1) were used as outgroups for phylogenetic analyses. All *Yucca* chloroplast assemblies as well as *Agave* and *Nolina* were aligned using MAFFT (Katoh and Standley, 2013). The alignment was manually inspected for misaligned regions, and as a result three *Yucca* genotypes (*Y. aloifolia* YA7, and *Y. gloriosa* YG13 and YG61) containing considerable misalignments indicative of a sub-optimal genome assembly were not included in further analyses. The second inverted repeat (IR) region was removed before tree estimation: an aligned *Y. aloifolia* chloroplast genome sequence (YA23) was annotated for the IR by conducting a

[1]https://github.com/mrmckain/Fast-Plast

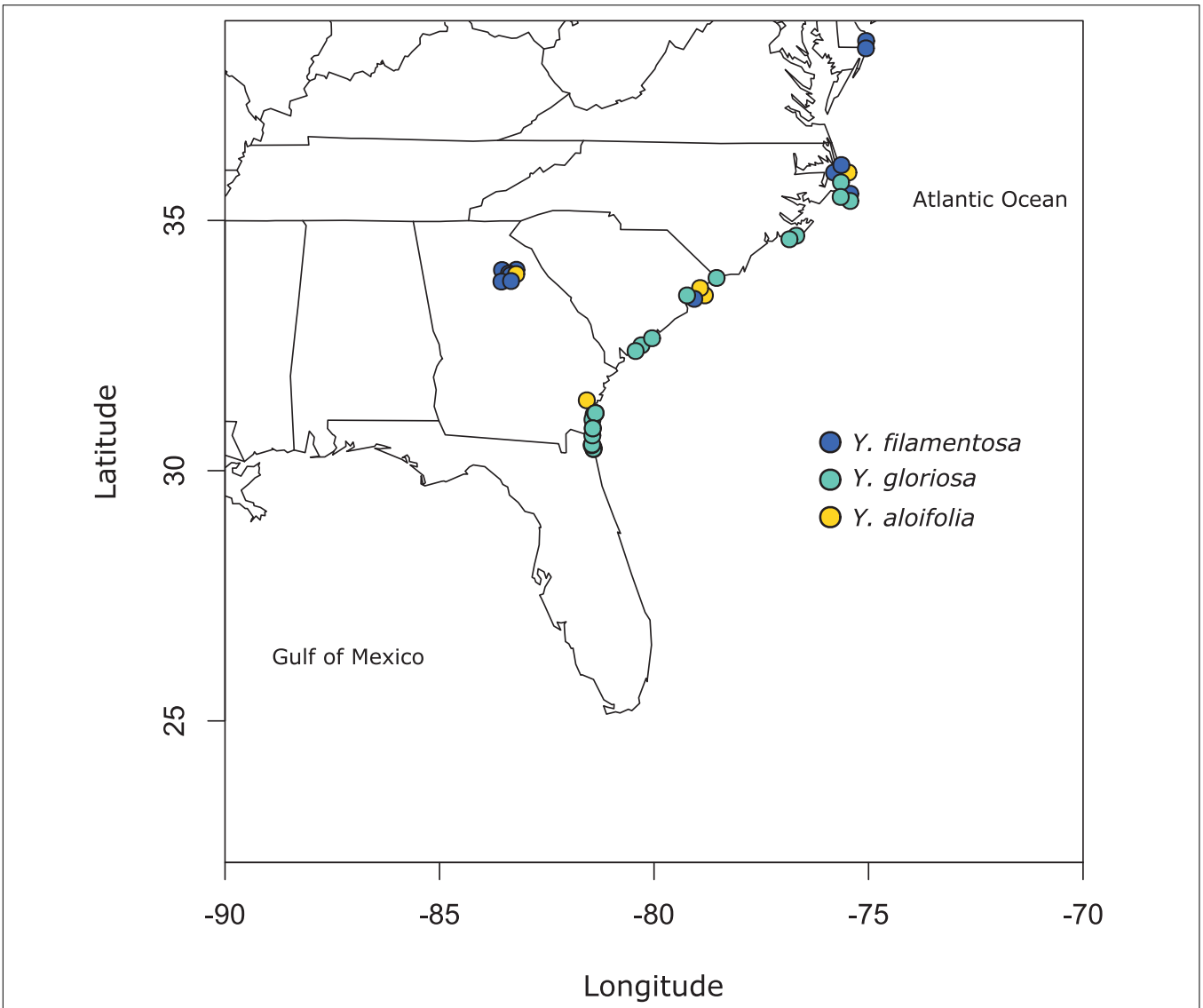


FIGURE 1 | Map of populations sampled for this study. See **Supplementary Table 1** for full geographic locality information.

BLASTn (Altschul et al., 1990) against itself. The position of an inverted self-hit in YA23 was used to remove the second IR from the multi-species alignment. The optimal model of molecular evolution (GTR + Gamma) was determined using JModelTest v2 and BIC penalized-likelihood (Darriba et al., 2012) on the CIPRES gateway (Miller et al., 2010). The multiple sequence alignment was then used to estimate a chloroplast phylogeny using RAxML v8.2.11, with 500 bootstrap replicates (Stamatakis, 2006). The entire chloroplast alignment (with both IR) of the *Yucca* species without outgroups was also used to construct a median joining chloroplast haplotype network using PopArt (epsilon = 0) (Leigh and Bryant, 2015). Chloroplast genome assemblies were annotated in Geneious Prime 2019.2.3, using the built-in annotation tool with the previously published *Y. filamentosa* annotation as a reference (NCBI accession: KX931467, McKain et al., 2016). Chloroplast genome assemblies are available on NCBI's GenBank (**Supplementary Table 1**), and the plastid alignment and newick files can be found on github[2].

Repetitive Content Classification and Analysis

In a similar fashion to the chloroplast sequence processing, one million trimmed paired-end reads were randomly sub-sampled for an analysis of transposon content. In order to ensure that only nuclear repetitive sequences were being analyzed, reads were first mapped to *Yucca* chloroplast and mitochondrial genome sequences (reference files are available at JGI Genome Portal[3]) using Bowtie v2 with default settings (Langmead and Salzberg, 2012) to be flagged for removal. The nuclear data were retained

[2]https://github.com/kheyduk/Yucca_plastome

[3]https://genome.jgi.doe.gov

and further processed in preparation for downstream steps, including: converting bam mapping files to fastq files using SAMTools v1.9 (Li et al., 2009) and BEDTools v2.26 (Quinlan and Hall, 2010), interleaving fastq files so that pairs are found sequentially in a single file (script[4], from Boisvert et al., 2010), and converting fastq files to fasta files with the FASTX-Toolkit v 0.14[5].

Transposome (Staton and Burke, 2015) was used to cluster and identify repetitive DNA sequences in all 41 *Yucca* genotypes using a *Yucca*-specific reference. Briefly, RepeatModeler (Smit and Hubley, 2008) was used to predict repeat families *de novo* on the assembled *Yucca* genomes; RepeatModeler uses both RECON (Bao and Eddy, 2002) and RepeatScout (Price et al., 2005) to identify repeat family consensus sequences. To remove false positives (e.g., repetitive domains within genes), the predicted RepeatModeler consensus sequences were searched for functional PFAM and Panther domains. If no domains—or only known transposable element domains—were found in a given putative repeat family, it was retained as a true repeat; if only false positive domains were identified, the family was removed from further analysis. Putative repeat families that had a combination of transposable element and false positive domains, or had otherwise unknown domain classes, underwent manual curation.

For annotating *Y. aloifolia* and *Y. filamentosa* repeats via Transposome, we used the species-specific RepeatModeler families (repetitive element reference files are available at JGI Genome Portal; see footnote). For *Y. gloriosa* hybrid individuals, we concatenated the two parental repeat databases. Finally, we used the following parameters in our usage of Transposome: percent identity = 90%, a required fraction of overlap between pairwise matches of 0.55, a minimum cluster size of 100, a merge threshold of 1,000, and a BLAST e-value of 1. Cross-species comparisons of transposons included the average amount of total repetitive DNA as well as the relative amounts (genomic proportion) of transposon superfamilies. In R v. 3.6.1 and v. 4.0.2 (R Core Team, 2019), we used ANOVA to determine whether there were significant differences between species in the relative amount of repetitive DNA in each of the 10 transposon superfamilies and abundance of individual family lineages within a superfamily. *Post hoc* tests were conducted with the emmeans package in R[6]. Additionally, a data matrix containing each individual's relative amount of repetitive DNA for each of the 10 superfamilies served as the input for a principal components analysis, using the prcomp() function in R. Throughout, we use the transposable element classification system described in Wicker et al. (2007).

Repetitive Element Activity via mRNAseq

Many repetitive elements contain sequences that are involved in their replication and therefore are transcribed into mRNA; transcripts produced from these repeats can be detected by mRNA sequencing (Hollister et al., 2011; Dion-Côté et al., 2014). While read counts from mRNA sequencing are a proxy for transcription of a repeat, no assumptions can be made as to the successful integration of a repeat copy into the genome post transcription; a variety of genomic mechanisms exist to silence and degrade repetitive element-derived transcripts (Lisch, 2009; Fultz et al., 2015). Nevertheless, as a first approximation of repeat activity, we used previously published mRNA-seq data on the three species of *Yucca* analyzed here (Heyduk et al., 2019). Briefly, RNA from leaf tissue was collected from all three species of *Yucca* growing in growth chambers set to 30°C/18°C day/night temperatures, with ∼400 μmol m^{-2} s^{-1} of light at leaf level, and 40% humidity in a 12 h day/night light regime. While the previous study further assessed gene expression under drought, here only libraries from well-watered plants taken during the daytime were analyzed. The original study used 2–3 genotypes per species, each of which had 2–3 replicates that were taken from different time points during the day. Because replication within a genotype is confounded with time, we limited our analyses to considering only species-specific differences rather than examining genotypic differences within species. Final species-level replication varied from 6 in *Y. aloifolia* to 9 in *Y. gloriosa* and *Y. filamentosa*.

RNA reads were mapped to the same repeat databases used in Transposome; *Y. aloifolia* and *Y. filamentosa* reads were mapped to each species' specific repeat reference, while *Y. gloriosa* reads were mapped to a merged parental reference. RNA reads were mapped via Kallisto v 0.43 using default parameters (Bray et al., 2016). For *Y. gloriosa*, counts were summed in cases where both parental species had a consensus sequence for a given repeat family. Libraries were first normalized by the Trimmed Mean of M-values (TMM) (Robinson and Oshlack, 2010) as implemented in EdgeR (Robinson et al., 2010), then scaled by overall abundance of that repeat family as estimated by Transposome. To scale, a matrix consisting of all repeat abundances across all genotypes from the three *Yucca* species was scaled by the maximum abundance of all families identified by Transposome. These scaled abundance values were then used as a multiplier of the TMM normalized read counts. By normalizing by genomic abundance, expression of repeats could then be compared across genotypes and species that have varying genomic fraction of the repeat families. Once normalized and scaled, we tested for significant expression within species using a glm intercept model in the glm.nb() function in the R package MASS (Venables and Ripley, 2013), which employs a negative binomial model appropriate for count data that exhibits a degree of over dispersion. Differentially expressed repeats between species were also tested with a negative binomial model, and *post hoc* tests were done using the emmeans() function from the R package emmeans.

RESULTS

Plastid Phylogenetic and Haplotype Analyses

Despite the relatedness between the three *Yucca* species studied here, there was enough divergence between the species' chloroplast genomes to identify highly supported clades of chloroplast haplotypes (**Figure 2**). *Y. gloriosa* genotypes were

[4]https://github.com/sebhtml/ray/blob/master/scripts/interleave-fastq.py

[5]http://hannonlab.cshl.edu/fastx_toolkit/

[6]https://github.com/rvlenth/emmeans

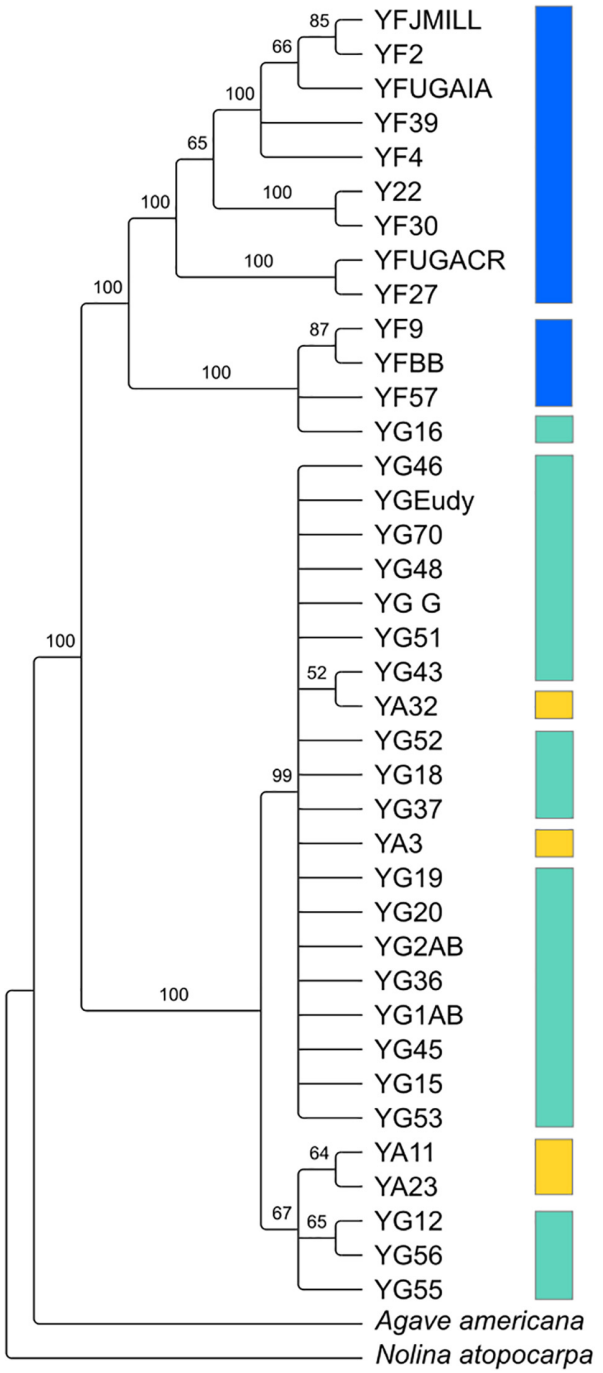


FIGURE 2 | RAxML estimated phylogeny of the plastome using only one inverted repeat copy in the alignment. Bootstrap support indicated on the branches, with nodes that had less than 50 bootstrap support collapsed. Colors indicate the three species: *Y. aloifolia* (yellow), *Y. filamentosa* (blue), and *Y. gloriosa* (teal).

found nested within three separate clades (**Figure 2**). A single *Y. gloriosa* genotype, YG16, was within a clade that otherwise contained all of the *Y. filamentosa* individuals that were analyzed. Three *Y. gloriosa* genotypes (YG12, YG55, and YG56) were placed in a clade with two *Y. aloifolia* genotypes (YA23 and YA11). The remaining 18 *Y. gloriosa* genotypes were grouped with the remaining two *Y. aloifolia* individuals (YA3 and YA32).

PopArt haplotype analysis (Leigh and Bryant, 2015) identified the same patterns found in the maximum likelihood-based phylogeny. Over 350 substitutions differentiated the two major groupings of genotypes (*Y. aloifolia* and *Y. filamentosa*-like chloroplast genomes; **Figure 3**). *Yucca filamentosa* had considerably more chloroplast haplotypes compared to *Y. aloifolia* (7 vs. 2, respectively; **Figure 3**). In contrast to previous analysis of nuclear simple repeats (Rentsch and Leebens-Mack, 2012), genetic diversity was seen not only in the *Y. aloifolia* chloroplast genomes but also for *Y. gloriosa*, which had four substitutions separating the different *Y. aloifolia*-like haplotypes, and over 400 substitutions separating the single *Y. filamentosa*-like haplotype from other individuals of *Y. gloriosa*.

Repetitive Fraction of *Yucca* Genomes

The fraction of the genome containing repetitive DNA significantly differed between the three species [$p < 0.001$, $F_{(2,\ 38)} = 17.853$]. While *Y. aloifolia* (mean repetitive genome fraction = 0.658; *SD* = 0.0138) and *Y. gloriosa* (mean = 0.662; *SD* = 0.0215) had statistically indistinguishable amount of repetitive DNA, *Y. filamentosa* was significantly lower than both species (mean = 0.621; *SD* = 0.0167; $p < 0.01$ for both *post hoc* comparisons). Moreover, the fraction of the genome composed of the various repeat families varied across the three species. The most abundant type of repeat in all three genomes were members of the *Gypsy* superfamily (**Figure 4A**), comprising ∼39% of the total genome, although species did not significantly differ in overall *Gypsy* abundance. The second most abundant superfamily in the *Yucca* genomes, at about ∼16.5%, was *Copia* (**Figure 4A**). *Yucca gloriosa* had significantly more *Copia* elements than either parent (*post hoc* comparison of *Y. gloriosa* to either parent $p < 0.001$). The third most abundant repeat superfamily was DNA *Helitrons*, at ∼3.5%, which had significantly different abundances between all three species (*post hoc* comparison $p < 0.01$). In general, the variation in repeat superfamily abundance between the three species was large enough to distinguish each species (**Supplementary Figure 1**), though intraspecific variation in repeat abundance was apparent as well. Certain transposable element superfamilies not typically seen in plant genomes—including non-LTR *Zisupton*, *Novosib*, and *Line-2* (*L2*) elements—were found at non-zero abundance levels in the *Yucca* species (**Supplementary Table 2**), but are not considered further here, as they may be the result of contamination in DNA isolation (e.g., from fungi).

At the family level (repeat lineages within a superfamily), while variation across the three species existed (**Supplementary Table 3**), it was rare that a single family contributed to the differences seen at the superfamily level (**Figure 4A**). In particular, high abundance families in *Copia* showed a pattern of parental-specific ancestry, in that they were found in only one of the two parental species. Furthermore, these high abundance, parental-specific repeat families also had high

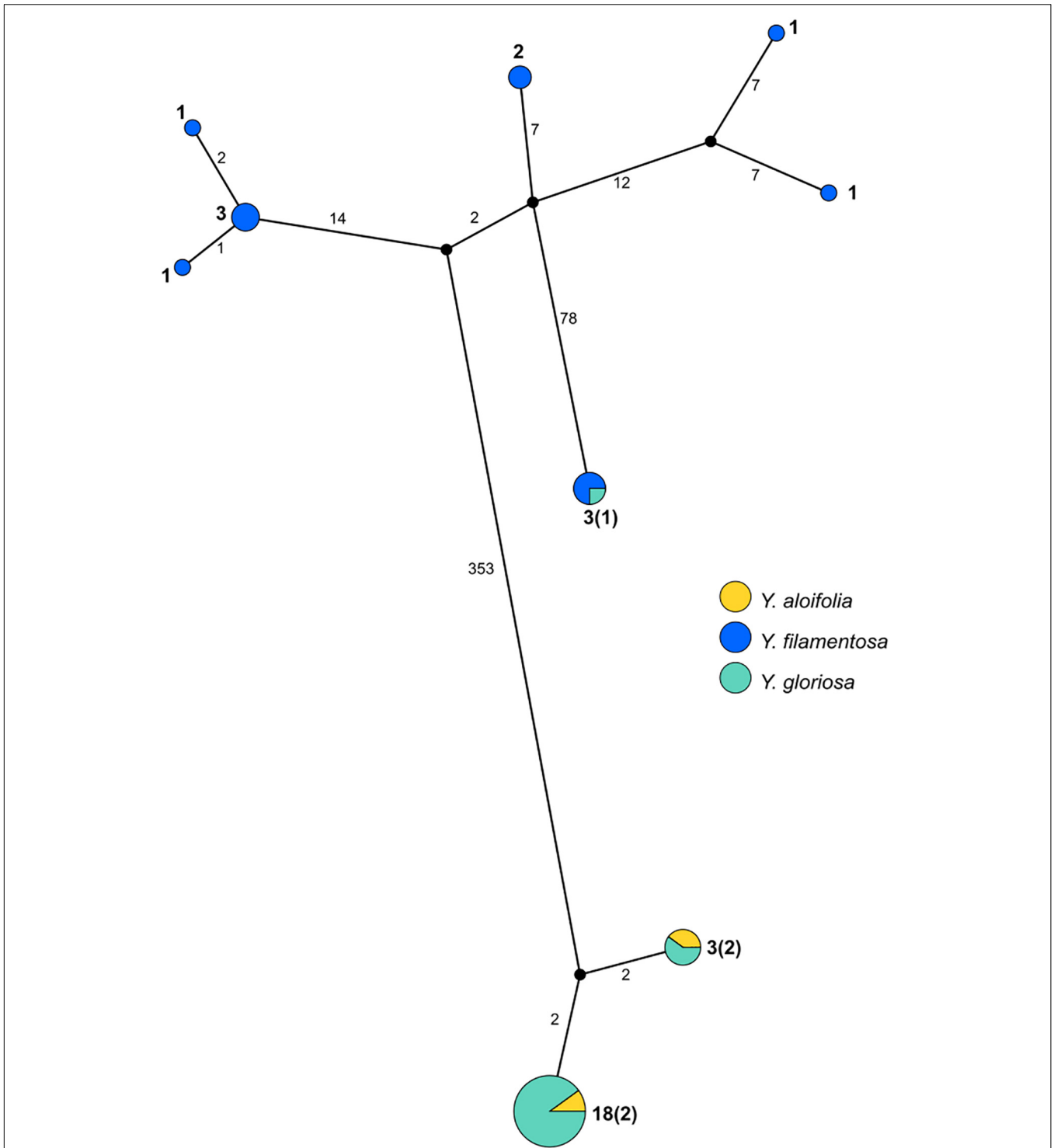


FIGURE 3 | Haplotype network estimated from the entire plastome alignment across all three species, excluding outgroup accessions. Haplotype estimated via PopArt, with number of substitutions separating haplotypes on branches and size (number of individuals) indicated in boldface numbers next to haplotypes, with majority number and, in parentheses, minority number.

abundance inherited in the hybrid, resulting in an overall higher abundance of *Copia* elements in *Y. gloriosa* relative to either parent (**Figures 4B,C**). In contrast, high abundance *Gypsy* elements were not parental-specific, and the hybrid therefore had equivalent abundance values as the parental abundance (**Figures 4B,C**). These high abundance, parental-specific families

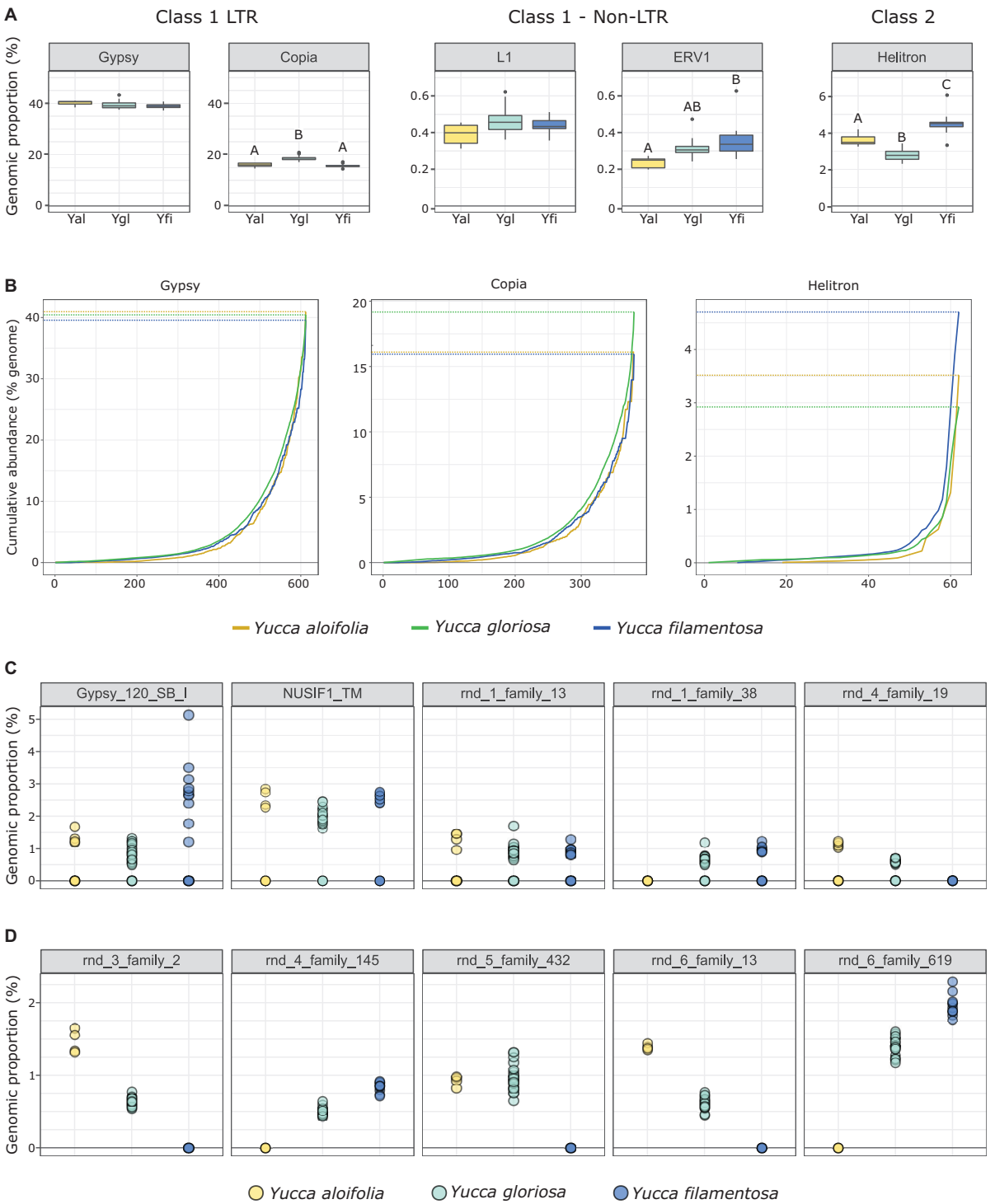


FIGURE 4 | (A) The genomic proportion (as a%) of the genome of a subset of transposable element families in *Y. aloifolia* (yellow), *Y. gloriosa* (teal), and *Y. filamentosa* (blue). Letters indicate significant differences based on Tukey *post hoc* tests from an ANOVA (abundance ~ species) per repeat superfamily; shared letters indicate no significant difference at $p < 0.01$. *y*-axes are scaled by category (LTR, non-LTR, Class II). **(B)** Cumulative abundance of individual families per superfamily, ordered from lowest average abundance to highest average abundance. **(C,D)** Individual genotype abundance values for the five most abundant families in Gypsy **(C)** and Copia **(D)**.

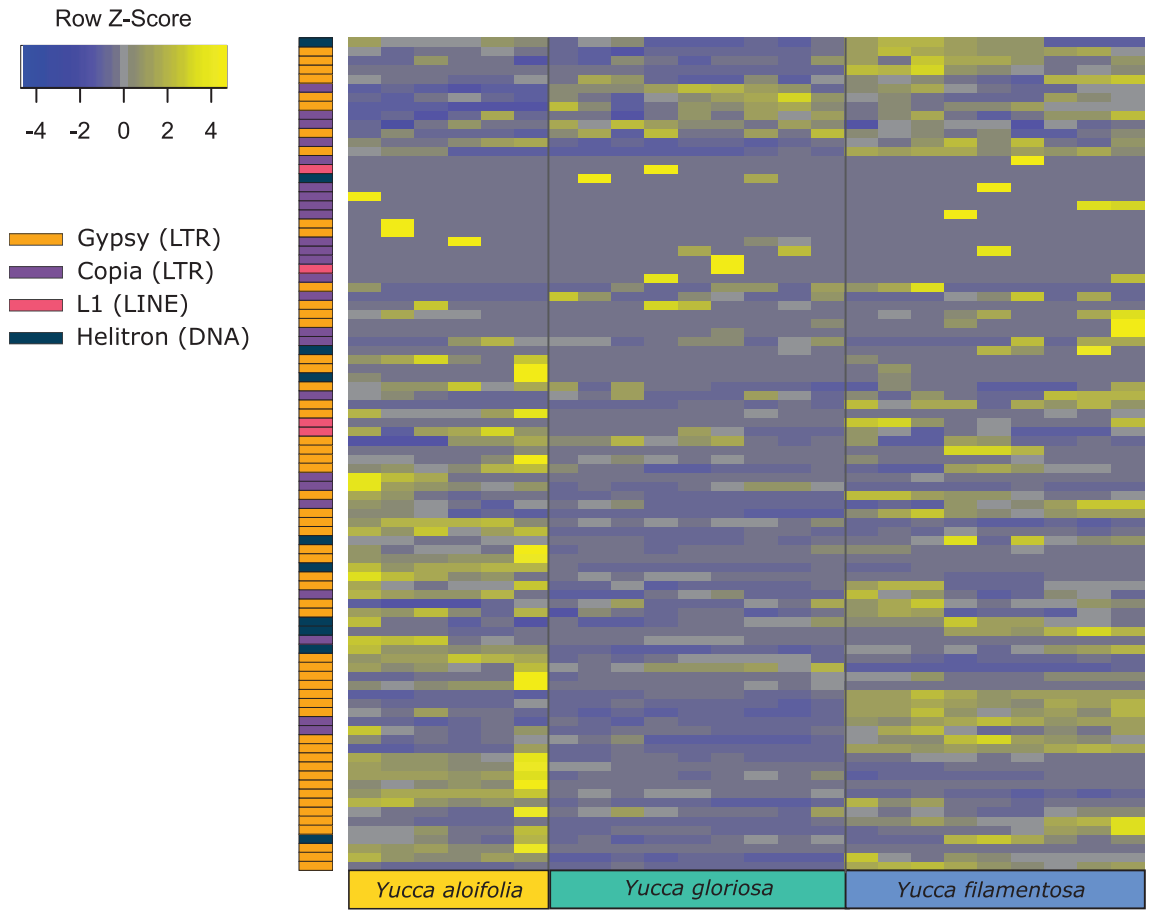


FIGURE 5 | Heatmap of normalized and scaled expression of 92 repeat families that were both present in all three species and had any detectable expression in any library. Individual rows are repeat clusters, and columns represent individual RNA-seq library replicates for each genotype.

in the *Copia* superfamily drive the overall species difference at the superfamily level.

Repeat mRNA Expression

Transposome abundance analysis of *Y. aloifolia* and *Y. filamentosa* identified 504 and 726 repeat families in at least one genotype of either species, respectively; only 231 repeat families were present in both parental species. Of the 231 families present in both species, only 118 and 119 repeat families had significantly non-zero expression in *Y. aloifolia* and *Y. filamentosa*, respectively (Benjamini-Hochberg adjusted $p < 0.01$) (**Supplementary Table 4**), and only 92 families were both present in all three species and had detectable expression in any one library (**Figure 5**). Only 27 families were significantly expressed in both parental species (**Table 1**). Repeat families with significant expression were typically from *Gypsy* (64 and 61% of total families expressed in *Y. aloifolia* and *Y. filamentosa*, respectively) and *Copia* (25, 27%) superfamilies. *Yucca gloriosa* had largely overlapping expression with its parental species; the hybrid shared significant expression of 74 families with *Y. aloifolia* and 70 families with *Y. filamentosa*. *Yucca gloriosa* had only two families that were not also significantly expressed in either parent: one a member of the *Gypsy* superfamily, the other belonging to the *Copia* superfamily, and both had genomic abundance at less than 1%. Overall, normalized and scaled expression values were positively correlated with the genomic abundance of a family across the three species [$r = 0.23$, $t_{(5,214)} = 4.9$, $p < 0.001$ for full data, $r = 0.41$, $t_{(5,204)} = 31.83$, $p < 0.001$ for families with TPM < 2,000].

In comparing the repeat families that are significantly expressed in *any* of the three species, *Y. gloriosa* showed little transgressive expression patterns; in most of the 178 repeat families that had significant *post hoc* comparisons, *Y. gloriosa* was not statistically different than one of its parental species. There were only three repeat families where expression differed significantly in all three species (*post hoc* $p < 0.01$) (**Supplementary Table 5**), and in 5 families, *Y. gloriosa* exhibited an expression level that was significantly different than the pattern shared in the two parental species (*post hoc* $p < 0.01$) (**Figure 6**). In all five cases, *Y. gloriosa* expression was significantly lower than the parental species' expression, though notably not zero. In general, however, the expression levels of repetitive elements in *Y. gloriosa* were shared with one or both parental species. Nine transposons families showed shared expression

TABLE 1 | Mean expression and abundance of repeat families in *Y. aloifolia* and *Y. filamentosa* significantly expressed above zero ($p < 0.01$).

		Yucca aloifolia		*Yucca filamentosa*		*Yucca gloriosa*	
Family	**Superfamily**	**Mean Exp.[a]**	**Mean Abun.[b]**	**Mean Exp.**	**Mean Abun.**	**Mean Exp.**	**Mean Abun.**
Copia_18_BD_I$^{\S\Phi}$	LTR/Copia	15.10	0.01%	16.19	0.01%	5.37	0.004%
Copia12_ZM_I	LTR/Copia	4.49	0.15%	2.91	0.10%	1.86	0.08%
Gypsy_120_SB_I$^{\Phi}$	LTR/Gypsy	12.67	1.21%	24.94	2.62%	8.58	0.91%
Gypsy_3_OS_I$^{*\Phi}$	LTR/Gypsy	16.52	0.12%	97.12	0.29%	9.82	0.09%
Gypsy_4_BD_LTR§	LTR/Gypsy	4.91	0.10%	4.16	0.10%	1.62	0.07%
Gypsy_5B_OS_LTR$^{\S\Phi}$	LTR/Gypsy	66.37	0.53%	38.85	0.27%	14.00	0.19%
Gypsy_8_OS_I$^{*\Phi}$	LTR/Gypsy	70.39	0.12%	372.53	0.34%	45.28	0.13%
Helitron_N117_OS$^{\Phi}$	Helitron	1796.11	0.76%	1953.53	1.19%	989.95	0.67%
Helitron_N29B_OS	Helitron	3.77	0.10%	6.11	0.11%	3.13	0.09%
Helitron_N84_OS$^{\S\Phi}$	Helitron	16.65	1.20%	12.14	0.71%	4.67	0.46%
Helitron7_OS	Helitron	92.69	0.86%	104.54	1.03%	52.54	0.60%
NUSIF1_TM$^{\S\Phi}$	LTR/Gypsy	507.41	2.80%	405.08	2.64%	128.68	1.87%
rnd_1_family_13$^{\S\Phi}$	LTR/Gypsy	146.24	1.22%	147.59	0.96%	65.60	1.03%
rnd_1_family_14*§	LTR/Gypsy	58.75	0.48%	17.27	0.17%	22.38	0.27%
rnd_1_family_15$^{*\S\Phi}$	LTR/Gypsy	98.84	0.49%	21.90	0.35%	41.25	0.43%
rnd_1_family_20$^{*\S\Phi}$	LTR/Gypsy	65.94	0.43%	20.69	0.15%	9.31	0.14%
rnd_1_family_23$^{*\Phi}$	LTR/Gypsy	15.88	0.11%	45.68	0.35%	12.38	0.17%
rnd_1_family_30$^{*\Phi}$	LTR/Gypsy	8.83	0.20%	14.89	0.28%	5.18	0.24%
rnd_1_family_32	LTR/Gypsy	7.75	0.09%	5.83	0.24%	4.97	0.17%
rnd_1_family_37	LTR/Gypsy	9.69	0.13%	8.24	0.10%	4.28	0.12%
rnd_1_family_47*§	LTR/Gypsy	22.26	0.50%	3.63	0.33%	5.91	0.51%
rnd_1_family_505$^{*\Phi}$	LTR/Copia	2.66	0.13%	20.96	0.36%	3.93	0.18%
rnd_1_family_56	LTR/Gypsy	66.87	0.21%	54.49	0.20%	58.95	0.25%
rnd_1_family_71$^{*\Phi}$	LTR/Gypsy	13.84	0.19%	3.05	0.35%	12.97	0.27%
rnd_1_family_76*§	LTR/Gypsy	56.40	0.28%	10.09	0.20%	13.02	0.23%
rnd_1_family_9*§	LTR/Gypsy	275.20	0.54%	15.66	0.11%	20.41	0.16%
SZ_22_int$^{*\Phi}$	LTR/Gypsy	2.43	0.14%	9.69	0.25%	1.84	0.16%

[a] *Mean expression is TMM normalized and scaled by abundance, then averaged across libraries.*
[b] *Mean abundance is the genomic fraction predicted by Transposome per genotype and averaged across genotypes within each species.*
Indicates repeat family is significantly differentially expressed between the parental species (*), between Y. gloriosa and Y. aloifolia (§), or between Y. gloriosa and Y. filamentosa (Φ), all at $p < 0.01$.
See ***Supplementary Table 5*** *for full test results and ANOVA statistics.*

in *Y. gloriosa* and *Y. filamentosa* that differed significantly from *Y. aloifolia,* and seven transposons had shared expression between *Y. gloriosa* and *Y. aloifolia* that differed significantly from *Y. filamentosa*. The majority of transposons had shared expression between the two parents, but significantly different expression between *Y. gloriosa* and either *Y. aloifolia* ($n = 76$) or *Y. filamentosa* ($n = 77$). There was a single transposon family where the parental species had significantly different expression from each other and *Y. gloriosa*'s expression was not significantly different than either parent.

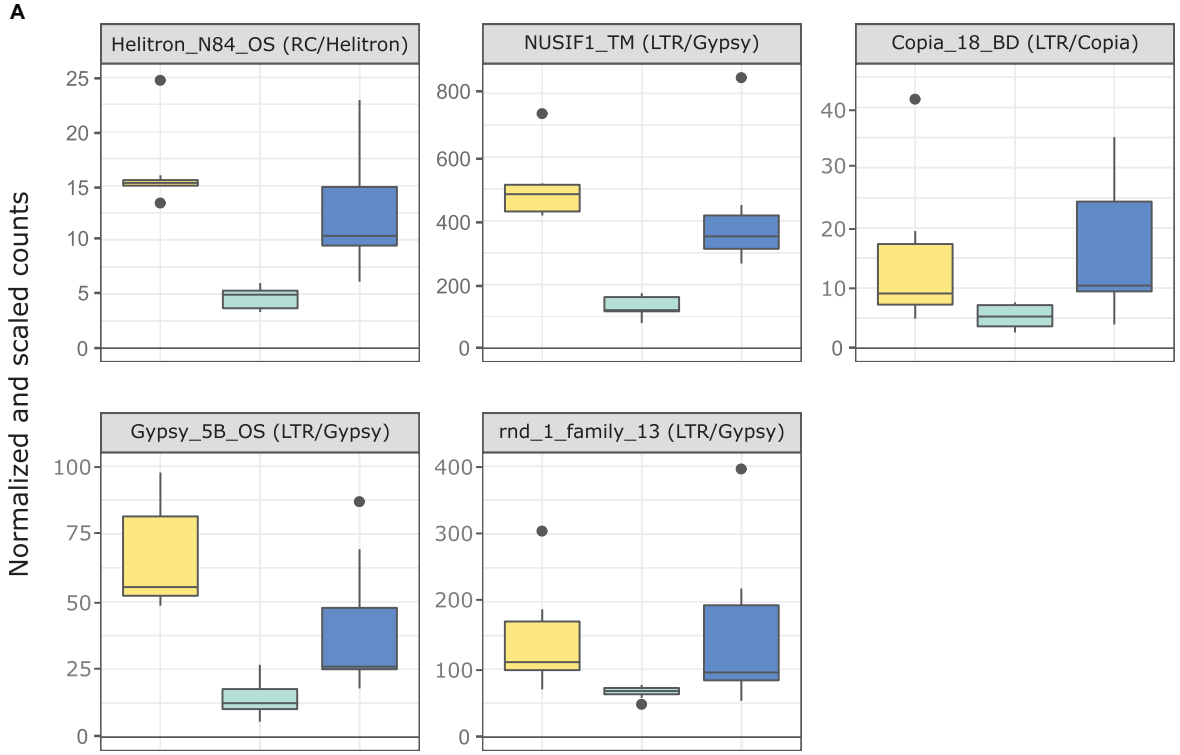
DISCUSSION

By increasing both the number of *Yucca* genotypes and assessing the whole chloroplast genome we have greatly improved resolution of the history of homoploid hybridization in *Yucca* relative to previous analyses of simple sequence repeats and short fragments of the chloroplast (Rentsch and Leebens-Mack, 2012). Whereas the previous work inferred a single, shared plastid haplotype in *Y. aloifolia* and *Y. gloriosa*, our findings implicate multiple origins of *Y. gloriosa* with both *Y. aloifolia* and *Y. filamentosa* acting as maternal parents. Moreover, analyses of nuclear TE abundances document overall quite similar TE landscapes across the three species, but certain families showed species-specific shifts in abundance. Using mRNA to assess current transposon activity, we find little evidence for ongoing release of transposons in the hybrid genome.

Reciprocal Parentage and Multiple Origins

Using 15–40x whole genome resequencing data, chloroplast assemblies for 38 individuals of *Yucca* across three species provided robust re-assessment of the history of this hybrid system. The presence of three separate clades containing *Y. gloriosa* (**Figure 2**) strongly suggests that not only can *Y. aloifolia* act as the maternal parent in the cross, as previously suggested, but that a reciprocal cross with *Y. filamentosa* as the maternal parent was viable enough to produce at least one extant

A



B

Abundance (percent of genome)

Family	*Y. aloifolia*	*Y. gloriosa*	*Y. filamentosa*
Helitron_N84_OS (RC/Helitron)	1.2% (A)	0.46% (C)	0.70% (C)
NUSIF1_TM (LTR/Gypsy)	2.8% (A)	1.87% (B)	2.63% (A)
Copia_18_BD (LTR/Copia)	0.01% (A)	0.004% (B)	0.008% (AB)
Gypsy_5B_OS (LTR/Gypsy)	0.5% (A)	0.1% (B)	0.2% (B)
rnd_1_family_13 (LTR/Gypsy)	1.2% (A)	1.02% (A)	0.96% (A)

FIGURE 6 | (A) Expression plot of the 5 TE families that were significantly differentially expressed between *Y. gloriosa* (teal) and both of its parental species (*Y. aloifolia* = yellow, *Y. filamentosa* = blue). TMM-normalized count data that is further scaled by abundance is plotted. **(B)** Mean percent abundance per species, as estimated by Transposome, and the result of *post hoc* test using emmeans() in R on the results of a negative binomial generalized linear model. Shared letters indicate no significant difference at a $p < 0.01$.

lineage in *Y. gloriosa*. While *Y. filamentosa* acting as the maternal parent in at least one cross is a parsimonious explanation for the data, the presence of a *Y. filamentosa* chloroplast in *Y. gloriosa* could also be due to a backcrossing event in which a *Y. gloriosa* pollen grain sired a seed on a *Y. filamentosa* individual. Such a backcross is unlikely to have happened recently. All of the genotypes of *Y. gloriosa* in this study (n = 24), as well as 2 and 6 of the genotypes for *Y. aloifolia* and *Y. filamentosa* from this study, respectively, have been phenotyped extensively for photosynthesis related traits (Heyduk et al., 2020), and a recent backcrossed hybrid would be expected to have photosynthetic physiology more similar to *Y. filamentosa* than *Y. aloifolia*, as the parents are strongly divergent in whether they use C_3 photosynthesis or CAM, respectively. However, the genotype of *Y. gloriosa* with the *Y. filamentosa* chloroplast haplotype (YG16) has strong signatures of CAM, including nocturnal CO_2 uptake as well as acid accumulation, traits which are diagnostic of the CAM phenotype displayed by *Y. aloifolia* (Heyduk et al., 2020). Additionally, the three species are very easy to distinguish in the field by leaf morphology: *Y. filamentosa* has filamentous leaf margins, *Y. aloifolia* has serrated leaf margins, and *Y. gloriosa* has smooth leaf margins. However these observations cannot rule out a more ancient backcrossing event, in which an original *Y. filamentosa* x *Y. gloriosa* cross's progeny thereafter crossed only

within *Y. gloriosa*, which over time would largely dampen the addition of the *Y. filamenotsa* nuclear genome but the chloroplast haplotype would remain.

The two clades of *Y. gloriosa* individuals that group with *Y. aloifolia* further support the inference that *Y. gloriosa* is derived from multiple hybridization events. However, as with the one instance of a *Y. filamentosa* chloroplast in *Y. gloriosa*, it is difficult to rule out recent backcrossing as the source of this observation (though leaf margins of all *Y. gloriosa* individuals sampled here had smooth margins that are diagnostic of this species in the wild). Earlier literature suggests that *Y. aloifolia* was introduced into the Southeastern United States ∼500 years ago by European colonists, who likely brought the plant from the Caribbean or Central America (Trelease, 1893; Groman and Pellmyr, 2000; Pellmyr, 2003). Within that time frame, however, the age of the hybridization events remains unknown. Additional analysis of re-resequencing data will assist in determining the number and timing of putative hybridization events. For example, the length of parental haplotype segments in a hybrid genome is related to the degree of recombination across the hybrid genome; short haplotype blocks would indicate a greater degree of recombination and, therefore, an older hybridization event. On the other hand, longer intact parental haplotype blocks in the hybrid may point to more recent hybridization. Moreover, the length of these haplotype blocks will vary between individuals, and may point to a mixture of both older and younger hybridization events within *Y. gloriosa*.

Previous work on the three *Yucca* species suggested that all *Y. aloifolia* and *Y. gloriosa* individuals shared a single chloroplast haplotype (Rentsch and Leebens-Mack, 2012). Comparisons across the entire chloroplast genome show that four nucleotide differences separated the two clades of *Y. aloifolia* and *Y. gloriosa* individuals. Over 400 genetic changes separate the *Y. filamentosa* and YG16 haplotypes from all *Y. aloifolia* and the remaining *Y. gloriosa* haplotypes. In agreement with the previous work, this study documents low plastid genetic diversity within *Y. aloifolia* and most *Y. gloriosa* samples. *Yucca aloifolia* is introduced into the Southeastern United States and likely suffered a bottleneck, resulting in lower overall diversity. The current sample of *Y. gloriosa* individuals identified one individual with a *Y. filamentosa*-derived haplotype. Additionally, this analysis identified seven discrete haplotypes within *Y. filamentosa*, which parallels the greater number of alleles per locus in *Y. filamentosa* suggested by previous work (Rentsch and Leebens-Mack, 2012).

Any attempt at describing the frequency of hybrid formation will be largely affected by the number of individuals in the germplasm collection. The original collection area spanned a large portion of the Southeastern United States in order to capture a significant amount of genetic diversity within the genus. Collections of *Y. gloriosa* likely represent many of the extant populations, but the ranges of both *Y. aloifolia* and *Y. filamentosa* are much larger than sampled here. As a result, any interpretation of geographic patterns to the chloroplast phylogeny or haplotype network are hampered by relatively low sampling of the parental genetic diversity. For example, the single *Y. gloriosa* individual found with a *Y. filamentosa* chloroplast (YG16) was collected in South Carolina, while *Y. filamentosa* individuals with the most similar haplotypes were collected in Delaware, North Carolina, and South Carolina. This haplotype grouping is clearly not geographically localized to one portion of the Atlantic coast and could be the result of missing genetic diversity in our analysis. Additionally, the Southeastern United States coastline experiences hurricanes and/or tropical storms on nearly an annual basis. Such storms have the potential to both disperse genets as well as eradicate entire populations and could make geographic interpretation of extant diversity difficult.

Transposable Element Abundance and Amplification

Genome resequencing provides a relatively unbiased sampling of the genome, allowing us to estimate the genomic fraction composed of transposable elements. Among sequenced plant genomes, transposable element contribution to genome size ranges from 14% in *Eragrostis tef* to 85% in *Zea mays* (Wendel et al., 2016). While all three *Yucca* species described in this work fall within the described range, the three species varied in the total amount of repetitive DNA with *Y. filamentosa* having significantly less repetitive DNA that *Y. aloifolia* and *Y. gloriosa* (62% vs. 65/66%). However, variation in abundance of particular repeat superfamilies does suggest superfamily-specific changes between the three species. *Copia* elements, the second most abundant superfamily of repeat in all three species, were more abundant in *Y. gloriosa* relative to both parents, suggesting an amplification of this superfamily post-hybridization. While Class 2 elements represent a relatively small proportion of *Yucca* genomes, *Helitrons* were found more often in *Y. filamentosa* compared to either *Y. aloifolia* or *Y. gloriosa*. *Helitrons* are capable of generating a tremendous amount of structural novelty, including the ability to capture and re-distribute pieces of genes (Yang and Bennetzen, 2009). As genomes become available for these species, it will be possible to analyze the extent to which all types of transposable elements have facilitated structural rearrangements and have affected expression of neighboring genes.

Previous work in various hybrid systems has shown incredible changes to the genomes post-hybridization. In a wallaby x kangaroo cross, reduced methylation of the genome resulted in the proliferation of a novel transposable element that caused significant structural changes to the chromosomes (O'Neill et al., 1998). Interspecific hybrids in *Drosophila* had an increase in transposable element mobilization relative to parental species (Vela et al., 2014). Three independent homoploid hybrids in *Helianthus* all show increased genome size due to expansion of repetitive elements, particularly in *Ty3/gypsy-like* LTR elements (Ungerer et al., 2006, 2009). In *Yucca*, however, there seems to be little indication that transposable elements were released from silencing mechanisms and proliferated in the hybrid *Y. gloriosa*. Instead, *Y. gloriosa* shows similar abundance of transposable elements relative to its progenitor species, though with a notable increase in *Copia* elements in the hybrid (**Figure 4**). Extant genotypes of *Y. gloriosa* have little in the way of increased repeat expression (**Figure 6**); whether this means no genomic shock initially happened upon hybridization, or that the

genome has had sufficient time to stabilize repetitive elements, remains unclear.

Finally, the three *Yucca* species provide an excellent system within which to describe the role of repetitive content on novel phenotypic evolution and adaptation. *Yucca gloriosa* has been studied extensively for its intermediate photosynthetic phenotype (Heyduk et al., 2016, 2019). When well-watered, the majority of carbon fixation happens during the day through the C_3 cycle, although low levels of CAM activity are present. When drought stress, *Y. gloriosa* can switch to predominantly CAM photosynthesis, but the degree to which individual genotypes do so varies. The hybrid's photosynthetic phenotype is novel, in that neither parent displays CAM induction upon drought stress, nor the ability to switch from primarily C_3 carbon fixation to primarily CAM. On first glance, negligible differences in repeat content and activity in *Y. gloriosa* relative to its parents suggest that repetitive content is unlikely to underlie the novel photosynthetic phenotype in the hybrid. However, here we only assessed overall abundance and activity in extant individuals; location of repeats in the hybrid relative to the parental species, as well as older repetitive content bursts, still have the potential to create transgressive and novel phenotypes in the hybrid. Repetitive elements can alter gene expression and gene networks by inserting into regulatory regions (Kunarso et al., 2010; Wang et al., 2013), can interfere with alternative splicing (Leprince et al., 2001; Li et al., 2014), and can be a general source of genomic variation and rapid evolution (González et al., 2010; Schrader et al., 2014). Moreover, transposable element activity can increase in response to environmental stressors (Makarevitch et al., 2015) and can play a role in forming stress-induced regulatory networks (Naito et al., 2009). Whether transposable elements are responsible for *Y. gloriosa*'s ability to upregulate CAM photosynthesis under drought stress remains to be tested.

CONCLUSION

Since the chloroplast phylogeny and haplotype network imply multiple hybridization events contributing to the origin of *Y. gloriosa*, new hypotheses regarding the repeatability of transposon accumulation can now be tested. For example, since YG16 appears to most likely be derived from a distinct hybridization event relative to other *Y. gloriosa* genotypes, we can assess whether the genomic organization of its transposable elements is vastly different from the major clade of *Y. gloriosa* genotypes grouping with *Y. aloifolia* (**Figure 2**). Integrating transposable element abundance and expression with other types of genomic data, including RNA-seq and bisulfite sequencing, may help us understand the potential for insertions to differentially regulate genes. The *Yucca* system is particularly powerful, in that the parental species are strongly divergent in photosynthetic pathway and the hybrid segregates for many of the same traits; this provides a framework in which to understand the role of repeats in regulating these genes in *Y. gloriosa*.

Given the massively expanding availability of whole genome sequence data, hypothesis-driven comparative analyses of genome content and structure are becoming more tractable. In this work, reads that normally would have been filtered out were instead analyzed to address whether a hybrid species had multiple and/or reciprocal origins. Furthermore, these reads helped provide a first glance into the repetitive landscape of 40 genotypes across three related species. While whole genomes will ultimately have the greatest ability to answer many of the questions brought up in this work, the approaches used here are quicker, less expensive, and generate many hypotheses for testing at the genome level in the future.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in **Supplementary Table 1**.

AUTHOR CONTRIBUTIONS

JG prepared libraries and sequenced samples. SS annotated repetitive content in the parental genomes. EM and KH conducted all analyses and wrote the manuscript. MM optimized plastome assembly and assisted with the manuscript. JL-M and JS were integral to overall project planning and management and assisted with the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by a DOE Joint Genome Institute Community Science Project award to KH. The work conducted by the US DOE Joint Genome Institute is supported by the Office of Science of the US Department of Energy under Contract no. DE-AC02-05CH11231.

ACKNOWLEDGMENTS

We gratefully acknowledge Amanda L. Cummings for assistance with DNA preparation, the Georgia Advanced Computing Resource Center, and the staff at the University of Georgia greenhouses, in particular Michael Boyd and Gregory Cousins. We thank the DOE Joint Genome Institute and collaborators for pre-publication access to the WGS data from *Yucca* accessions herein and to repeat databases from the genome sequences of *Yucca aloifolia* and *Yucca filamentosa*. This manuscript has been released as a pre-print at bioRxiv, https://doi.org/10.1101/2020.06.14.150078. This is publication #95 from the School of Life Sciences, University of Hawai'i at Mānoa.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.573767/full#supplementary-material

REFERENCES

Abraham, P. E., Yin, H., Borland, A. M., Weighill, D., Lim, S. D., De Paoli, H. C., et al. (2016). Transcript, protein and metabolite temporal dynamics in the CAM plant Agave. *Nat. Plants* 2:16178.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Arnold, M. L. (1993). Iris nelsonii (Iridaceae): origins and genetic composition of a homoploid hybrid species. *Am. J. Bot.* 80, 577–583. doi: 10.1002/j.1537-2197.1993.tb13843.x

Bao, Z., and Eddy, S. R. (2002). Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* 12, 1269–1276. doi: 10.1101/gr.88502

Bardil, A., de Almeida, J. D., Combes, M. C., Lashermes, P., and Bertrand, B. (2011). Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature. *New Phytol.* 192, 760–774. doi: 10.1111/j.1469-8137.2011.03833.x

Bird, K. A., VanBuren, R., Puzey, J. R., and Edger, P. P. (2018). The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol.* 220, 87–93. doi: 10.1111/nph.15256

Boisvert, S., Laviolette, F., and Corbeil, J. (2010). Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J. Comput. Biol.* 17, 1519–1533. doi: 10.1089/cmb.2009.0238

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519

Bredeson, J. V., Lyons, J. B., Prochnik, S. E., Wu, G. A., Ha, C. M., Edsinger-Gonzales, E., et al. (2016). Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* 34, 562–570. doi: 10.1038/nbt.3535

Danilova, T. V., Akhunova, A. R., Akhunov, E. D., Friebe, B., and Gill, B. S. (2017). Major structural genomic alterations can be associated with hybrid speciation in *Aegilops markgrafii* (Triticeae). *Plant J.* 92, 317–330. doi: 10.1111/tpj.13657

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109

Devos, K. M., Brown, J. K. M., and Bennetzen, J. L. (2002). Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* 12, 1075–1079. doi: 10.1101/gr.132102

Dion-Côté, A.-M., Renaut, S., Normandeau, E., and Bernatchez, L. (2014). RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake whitefish species. *Mol. Biol. Evol.* 31, 1188–1199. doi: 10.1093/molbev/msu069

Doyle, J. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.

Edger, P. P., Smith, R., McKain, M. R., Cooley, A. M., Vallejo-Marin, M., Yuan, Y., et al. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* 29, 2150–2167. doi: 10.1105/tpc.17.00010

Feschotte, C., and Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* 41, 331–368. doi: 10.1146/annurev.genet.40.110405.090448

Fultz, D., Choudury, S. G., and Slotkin, R. K. (2015). Silencing of active transposable elements in plants. *Curr. Opin. Plant Biol.* 27, 67–76. doi: 10.1016/j.pbi.2015.05.027

González, J., Karasov, T. L., Messer, P. W., and Petrov, D. A. (2010). Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. *PLoS Genet.* 6:e1000905. doi: 10.1371/journal.pgen.1000905

Groman, J. D., and Pellmyr, O. (2000). Rapid evolution and specialization following host colonization in a yucca moth. *J. Evol. Biol.* 13, 223–236. doi: 10.1046/j.1420-9101.2000.00159.x

Gross, B. L., and Rieseberg, L. H. (2005). The ecological genetics of homoploid hybrid speciation. *J. Hered.* 96, 241–252. doi: 10.1093/jhered/esi026

Hegarty, M. J., Barker, G. L., Brennan, A. C., Edwards, K. J., Abbott, R. J., and Hiscock, S. J. (2009). Extreme changes to gene expression associated with homoploid hybrid speciation. *Mol. Ecol.* 18, 877–889. doi: 10.1111/j.1365-294x.2008.04054.x

Hersch, E. I., and Roy, B. A. (2007). Context-dependent pollinator behavior: an explanation for patterns of hybridization among three species of Indian paintbrush. *Evolution* 61, 111–124. doi: 10.1111/j.1558-5646.2007.00009.x

Heyduk, K., Burrell, N., Lalani, F., and Leebens-Mack, J. (2016). Gas exchange and leaf anatomy of a C3-CAM hybrid, *Yucca gloriosa* (Asparagaceae). *J. Exp. Bot.* 67, 1369–1379. doi: 10.1093/jxb/erv536

Heyduk, K., Ray, J. N., Ayyampalayam, S., Moledina, N., Borland, A., Harding, S. A., et al. (2019). Shared expression of Crassulacean acid metabolism (CAM) genes predates the origin of CAM in the genus *Yucca*. *J. Exp. Bot.* 70, 6597–6609. doi: 10.1093/jxb/erz105

Heyduk, K., Ray, J. N., and Leebens-Mack, J. (2020). Leaf anatomy is not correlated to CAM function in a C3+CAM hybrid species, *Yucca gloriosa*. *Ann. Bot.* 2020:mcaa036. doi: 10.1093/aob/mcaa036

Hollister, J. D., Smith, L. M., Guo, Y.-L., Ott, F., Weigel, D., and Gaut, B. S. (2011). Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2322–2327. doi: 10.1073/pnas.1018222108

Johnston, J. A., Grise, D. J., Donovan, L. A., and Arnold, M. L. (2001). Environment-dependent performance and fitness of *Iris brevicaulis*, I. fulva (Iridaceae), and hybrids. *Am. J. Bot.* 88, 933–938. doi: 10.2307/2657046

Karrenberg, S., Edelist, C., Lexer, C., and Rieseberg, L. (2006). Response to salinity in the homoploid hybrid species *Helianthus paradoxus* and its progenitors *H. annuus* and *H. petiolaris*. *New Phytol.* 170, 615–629. doi: 10.1111/j.1469-8137.2006.01687.x

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kunarso, G., Chia, N.-Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.-S., et al. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* 42, 631–634. doi: 10.1038/ng.600

Lai, Z., Gross, B. L., Zou, Y. I., Andrews, J., and Rieseberg, L. H. (2006). Microarray analysis reveals differential gene expression in hybrid sunflower species. *Mol. Ecol.* 15, 1213–1227. doi: 10.1111/j.1365-294x.2006.02775.x

Lai, Z., Nakazato, T., Salmaso, M., Burke, J. M., Tang, S., Knapp, S. J., et al. (2005). Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics* 171, 291–303. doi: 10.1534/genetics.105.042242

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Leebens-Mack, J., and Milligan, B. (1998). Pollination biology in hybridizing *Baptisia* (Fabaceae) populations. *Am. J. Bot.* 85:500. doi: 10.2307/2446433

Leebens-Mack, J., and Pellmyr, O. (2004). Patterns of genetic structure among populations of an oligophagous pollinating yucca moth (*Tegeticula yuccasella*). *J. Hered.* 95, 127–135. doi: 10.1093/jhered/esh025

Leigh, J. W., and Bryant, D. (2015). popart : full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6, 1110–1116. doi: 10.1111/2041-210x.12410

Leprince, A. S., Grandbastien, M. A., and Meyer, C. (2001). Retrotransposons of the Tnt1B family are mobile in Nicotiana plumbaginifolia and can induce alternative splicing of the host gene upon insertion. *Plant Mol. Biol.* 47, 533–541.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, Q., Xiao, G., and Zhu, Y.-X. (2014). Single-nucleotide resolution mapping of the *Gossypium raimondii* transcriptome reveals a new mechanism for alternative splicing of introns. *Mol. Plant* 7, 829–840. doi: 10.1093/mp/sst175

Lisch, D. (2009). Epigenetic regulation of transposable elements in plants. *Annu. Rev. Plant Biol.* 60, 43–66. doi: 10.1146/annurev.arplant.59.032607.092744

Makarevitch, I., Waters, A. J., West, P. T., Stitzer, M., Hirsch, C. N., Ross-Ibarra, J., et al. (2015). Transposable elements contribute to activation of maize genes in

response to abiotic stress. *PLoS Genet.* 11:e1004915. doi: 10.1371/journal.pgen.1004915

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* 226, 792–801. doi: 10.1126/science.15739260

McKain, M. R., McNeal, J. R., Kellar, P. R., Eguiarte, L. E., Pires, J. C., and Leebens-Mack, J. (2016). Timing of rapid diversification and convergent origins of active pollination within *Agavoideae* (Asparagaceae). *Am. J. Bot.* 103, 1717–1729. doi: 10.3732/ajb.1600198

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). "Creating the CIPRES Science Gateway for inference of large phylogenetic trees," in *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, LA, 1–8.

Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C. N., Richardson, A. O., et al. (2009). Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* 461, 1130–1134. doi: 10.1038/nature08479

O'Neill, R. J., O'Neill, M. J., and Graves, J. A. (1998). Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 393, 68–72. doi: 10.1038/29985

Parisod, C., Salmon, A., Zerjal, T., Tenaillon, M., Grandbastien, M.-A., and Ainouche, M. (2009). Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in Spartina. *New Phytol.* 184, 1003–1015. doi: 10.1111/j.1469-8137.2009.03029.x

Pellmyr, O. (1999). Systematic revision of the yucca moths in the *Tegeticula yuccasella* complex (Lepidoptera: Prodoxidae) north of Mexico. *Syst. Entomol.* 24, 243–271. doi: 10.1046/j.1365-3113.1999.00079.x

Pellmyr, O. (2003). *Yuccas, Yucca moths*, and coevolution: a review. *Ann. Mo. Bot. Gard.* 90, 35–55. doi: 10.2307/3298524

Price, A. L., Jones, N. C., and Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics* 21(Suppl. 1), i351–i358.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R foundation for statistical computing.

Rapp, R. A., Udall, J. A., and Wendel, J. F. (2009). Genomic expression dominance in allopolyploids. *BMC Biol.* 7:18. doi: 10.1186/1741-7007-7-18

Renaut, S., Rowe, H. C., Ungerer, M. C., and Rieseberg, L. H. (2014). Genomics of homoploid hybrid speciation: diversity and transcriptional activity of long terminal repeat retrotransposons in hybrid sunflowers. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130345. doi: 10.1098/rstb.2013.0345

Rentsch, J. D., and Leebens-Mack, J. (2012). Homoploid hybrid origin of Yucca gloriosa: intersectional hybrid speciation in *Yucca* (Agavoideae, Asparagaceae). *Ecol. Evol.* 2, 2213–2222. doi: 10.1002/ece3.328

Rentsch, J. D., and Leebens-Mack, J. (2014). Yucca aloifolia (Asparagaceae) opts out of an obligate pollination mutualism. *Am. J. Bot.* 101, 2062–2067. doi: 10.3732/ajb.1400351

Rieseberg, L. H. (1991). Homoploid reticulate evolution in helianthus (asteraceae): evidence from ribosomal genes. *Am. J. Bot.* 78, 1218–1237. doi: 10.1002/j.1537-2197.1991.tb11415.x

Rieseberg, L. H. (1997). Hybrid origins of plant species. *Annu. Rev. Ecol. Syst.* 28, 359–389. doi: 10.1146/annurev.ecolsys.28.1.359

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., et al. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301, 1211–1216. doi: 10.1126/science.1086949

Rieseberg, L. H., Van Fossen, C., and Desrochers, A. M. (1995). Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* 375, 313–316. doi: 10.1038/375313a0

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Robinson, M. D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11:R25.

Salmon, A., Ainouche, M. L., and Wendel, J. F. (2005). Genetic and epigenetic consequences of recent hybridization and polyploidy in Spartina (Poaceae). *Mol. Ecol.* 14, 1163–1175. doi: 10.1111/j.1365-294x.2005.02488.x

Sapir, Y., Moody, M. L., Brouillette, L. C., Donovan, L. A., and Rieseberg, L. H. (2007). Patterns of genetic diversity and candidate genes for ecological divergence in a homoploid hybrid sunflower, *Helianthus anomalus*. *Mol. Ecol.* 16, 5017–5029. doi: 10.1111/j.1365-294x.2007.03557.x

Schnable, J. C., Springer, N. M., and Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4069–4074. doi: 10.1073/pnas.1101368108

Schrader, L., Kim, J. W., Ence, D., Zimin, A., Klein, A., Wyschetzki, K., et al. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nat. Commun.* 5:5495.

Smit, A. F. A., and Hubley, R. (2008). *RepeatModeler Open-1.0*.

Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/bioinformatics/btl446

Staton, S. E., Bakken, B. H., Blackman, B. K., Chapman, M. A., Kane, N. C., Tang, S., et al. (2012). The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements. *Plant J.* 72, 142–153. doi: 10.1111/j.1365-313x.2012.05072.x

Staton, S. E., and Burke, J. M. (2015). Transposome: a toolkit for annotation of transposable element families from unassembled sequence reads. *Bioinformatics* 31, 1827–1829. doi: 10.1093/bioinformatics/btv059

Staton, S. E., and Ungerer, M. C. (2009). The genomic organization of Ty3/gypsy-like retrotransposons in *Helianthus* (Asteraceae) homoploid hybrid species. *Am. J. Bot.* 96, 1646–1655. doi: 10.3732/ajb.0800337

Štorchová, H., Hrdlièková, R., Chrtek, J., Tetera, M., Fitze, D., and Fehrer, J. (2000). An improved method of DNA isolation from plants collected in the field and conserved in saturated NaCl/CTAB solution. *Taxon* 49, 79–84. doi: 10.2307/1223934

Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat. Genet.* 43, 1160–1163. doi: 10.1038/ng.942

Sun, C., Ma, Z., Zhang, Z., Sun, G., and Dai, Z. (2018). Factors Influencing Cross Barriers in Interspecific Hybridizations of Water Lily. *J. Am. Soc. Hortic. Sci.* 143, 130–135. doi: 10.21273/jashs04302-17

Tenaillon, M. I., Hufford, M. B., Gaut, B. S., and Ross-Ibarra, J. (2011). Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*. *Genome Biol. Evol.* 3, 219–229. doi: 10.1093/gbe/evr008

Thórsson, A. T., Salmela, E., and Anamthawat-Jónsson, K. (2001). Morphological, cytogenetic, and molecular evidence for introgressive hybridization in birch. *J. Hered.* 92, 404–408. doi: 10.1093/jhered/92.5.404

Trelease, W. (1893). Further studies of yuccas and their pollination. *Missouri Bot. Garden Annu. Rep.* 1893, 181–226. doi: 10.2307/2992178

Trelease, W. (1902). The yucceae. *Missouri Bot. Garden Annu. Rep.* 1902, 27–133. doi: 10.2307/2400121

Ungerer, M. C., Strakosh, S. C., and Stimpson, K. M. (2009). Proliferation of Ty3/gypsy-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. *BMC Biol.* 7:40. doi: 10.1186/1741-7007-7-40

Ungerer, M. C., Strakosh, S. C., and Zhen, Y. (2006). Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Curr. Biol.* 16, R872–R873.

Vela, D., Fontdevila, A., Vieira, C., and García Guerreiro, M. P. (2014). A genome-wide survey of genetic instability by transposition in *Drosophila* hybrids. *PLoS One* 9:e88992. doi: 10.1371/journal.pone.0088992

Venables, W. N., and Ripley, B. D. (2013). *Modern Applied Statistics with S-PLUS*. Cham: Springer.

Wang, H., McArthur, E. D., Sanderson, S. C., Graham, J. H., and Freeman, D. C. (1997). Narrow hybrid zone between two species of Big Sagebrush (Artemisia tridentata: Asteraceae). IV. Reciprocal transplant experiments. *Evolution* 51, 95–102. doi: 10.2307/2410963

Wang, X., Weigel, D., and Smith, L. M. (2013). Transposon variants and their effects on gene expression in *Arabidopsis*. *PLoS Genet.* 9:e1003255. doi: 10.1371/journal.pgen.1003255

Welch, M. E., and Rieseberg, L. H. (2002). Habitat divergence between a homoploid hybrid sunflower species, *Helianthus paradoxus* (Asteraceae),

and its progenitors. *Am. J. Bot.* 89, 472–478. doi: 10.3732/ajb.89.3.472

Wendel, J. F., Jackson, S. A., Meyers, B. C., and Wing, R. A. (2016). Evolution of plant genome architecture. *Genome Biol.* 17:37.

Wessler, S. R., Bureau, T. E., and White, S. E. (1995). LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr. Opin. Genet. Dev.* 5, 814–821. doi: 10.1016/0959-437x(95)80016-x

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982. doi: 10.1038/nrg2165

Xiao, H., Jiang, N., Schaffner, E., Stockinger, E. J., and van der Knaap, E. (2008). A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 319, 1527–1530. doi: 10.1126/science.1153040

Xu, Y., Zhong, L., Wu, X., Fang, X., and Wang, J. (2009). Rapid alterations of gene expression and cytosine methylation in newly synthesized *Brassica napus* allopolyploids. *Planta* 229, 471–483. doi: 10.1007/s00425-008-0844-8

Yang, L., and Bennetzen, J. L. (2009). Distribution, diversity, evolution, and survival of Helitrons in the maize genome. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19922–19927. doi: 10.1073/pnas.0908008106

Yoo, M.-J., Szadkowski, E., and Wendel, J. F. (2013). Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110, 171–180. doi: 10.1038/hdy.2012.94

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership