UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
DEPARTAMENTO DE GENÉTICA E EVOLUÇÃO
PROGRAMA DE PÓS GRADUAÇÃO EM GENÉTICA EVOLUTIVA E BIOLOGIA MOLECULAR

MONIQUE ROMEIRO BRITO

FILOGENIA, BIOGEOGRAFIA E DIVERSIFICAÇÃO DE CACTOS NA AMÉRICA DO SUL COM ENFOQUE NA TRIBO CEREEAE (CACTACEAE)

SÃO CARLOS -SP
2023

MONIQUE ROMEIRO BRITO

FILOGENIA, BIOGEOGRAFIA E DIVERSIFICAÇÃO DE CACTOS NA AMÉRICA DO SUL COM ENFOQUE NA TRIBO CEREEAE (CACTACEAE)

Tese apresentada ao Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular, Universidade Federal de São Carlos, campus São Carlos, como parte dos requisitos para obtenção do título de doutor em Ciências.

Orientador: Dr. Evandro Marsola de Moraes

Financiamento: Fapesp (2018/06937-8)

SÃO CARLOS - SP
2023

Romeiro-Brito, Monique

Filogenia, biogeografia e diversificação de cactos na América do Sul com enfoque na tribo Cereeae (Cactaceae) / Monique Romeiro-Brito -- 2023.
167f.

Tese de Doutorado - Universidade Federal de São Carlos, campus São Carlos, São Carlos
Orientador (a): Evandro Marsola Moraes
Banca Examinadora: Cássio Van den Berg, Suzana de Fátima Alcantara, Matias Köhler, Fernando Faria Franco
Bibliografia

1. Filogenômica. 2. Biogeografia. 3. Taxa de diversificação. I. Romeiro-Brito, Monique. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática (SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Ronildo Santos Prado - CRB/8 7325

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Biológicas e da Saúde
Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular

Folha de Aprovação

Defesa de Tese de Doutorado da candidata Monique Romeiro Brito, realizada em 26/05/2023.

Comissão Julgadora:

Prof. Dr. Evandro Marsola de Moraes (UFSCar)

Prof. Dr. Matias Kohler (UFSCar)

Prof. Dr. Fernando de Faria Franco (UFSCar)

Profa. Dra. Suzana de Fátima Alcantara (UFSC)

Prof. Dr. Cássio Van Den Berg (UEFS)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular.

Este trabalho foi realizado com auxílio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) através dos processos 2018/06937-8 e 2019/03211-9.

Dedico a minha mãe e minha irmã,
meus exemplos de resistência e
resiliência.

“Dentro de mim tem um sonho
Querendo me despertar
Fazendo nascer o dia
Bem afastado do mar
Cercado dessa aridez
Olhando o mundo secar
Tem um lugar encantado
Que faz meu sonho acordar

Uma fazenda no céu
Uma viola no ar
Tendo esse sol como véu
A terra como lugar
Fico plantado na vida
Como semente de amor
Tendo uma rede estendida
Brota uma planta de dor”

Música ‘Aridez’ da banda ‘Flor de Cactus’

AGRADECIMENTOS

Foram muitas as pessoas, instituições e projetos que colaboraram para o desenvolvimento desse projeto de forma direta ou indireta. É com muito carinho e gratidão que lembro delas nessa seção.

À Universidade Federal de São Carlos (UFSCar) e ao Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular (PPGGEV), incluindo a coordenação, professores e secretaria (em especial, a Ivanildes Menezes) pela disponibilidade em auxiliar a qualquer momento e pela infraestrutura.

À Fundação de Amparo à Pesquisa do estado de São Paulo e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior pela concessão da bolsa de projeto de doutorado.

Ao Prof. Dr. Evandro Marsola de Moraes pela orientação, paciência, confiança e ensinamentos essenciais para o desenvolvimento desse projeto e para meu desenvolvimento profissional nesta etapa de imensa importância para minha formação.

Ao Prof. Dr. Fernando F. Franco, por ser meu primeiro incentivador a entrar nos caminhos da ciência desde a iniciação científica e por sempre trazer comentários e discussões enriquecedoras para o desenvolvimento desse projeto de doutorado.

À Profa. Dra. Daniela C. Zappi e ao Dr. Nigel P. Taylor pelo auxílio como especialistas do grupo e pelas discussões muito enriquecedoras para o presente trabalho. Sou imensamente grata por sempre me receberem com carinho e por aprender cada vez mais sobre os cactos e ambientes secos durante visitas e coletas de campo.

À Dr. Juliana F. Martinez e à Msc. Heidi M. Utsonomyia pelo apoio teórico e prático no LAGEvol e pela amizade.

Ao Gerardus Olsthoorn por gentilmente ceder amostras do seu acervo pessoal para o presente trabalho, e sempre por compartilhar seu conhecimento sobre o grupo de estudo.

Ao Dr. Diego R. Gonzaga por gentilmente ceder amostras da Coleção de Cactos e Suculentas do Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, as quais foram fundamentais para o presente trabalho.

À Profa. Dra. Lidyanne Yuriko Saleme Aona por gentilmente ceder amostras do Herbário do Recôncavo da Bahia que foram de extrema importância para completar a amostragem de espécies brasileiras do grupo de estudo.

A todos os meus queridos colegas e amigos do Laboratório de Diversidade Genética e Evolução (LAGEvol) que conheci durante minha passagem no laboratório. Por todos os

labmeetings, pelos cafés da tarde, e por promover uma convivência harmoniosa, prazerosa e colaborativa.

À Dr. Flávia M. Lanna por várias sugestões, discussões e scripts compartilhados sobre aprendizagem de máquina desde o início desse projeto. É sempre bom nos reunirmos para conversar sobre ciência com tanta leveza.

Ao Prof. Dr. Luciano A. Digiampietri por auxiliar e sugerir várias abordagens e ensinamentos sobre aprendizagem de máquina que foram fundamentais para as análises e desenvolvimento desse projeto.

Ao Dr. Matias Köhler, que tive o prazer de conhecer recentemente no LAGEvol e com quem pude compartilhar muitas discussões sobre evolução das cactáceas e sobre a vida na academia.

Aos meus queridos amigos do laboratório, em especial, Profa. Dra. Isabel A. S. Bonatelli, Prof. Dr. Danilo Trabuco Amaral, e Milena C. Telhe. Não tenho palavras suficientes para agradecer a amizade, crédito, companheirismo e colaboração de vocês durante toda essa fase turbulenta e desafiadora. Sou muito feliz por ter amigos e incentivadores como vocês.

Às minhas amigas e amigos de Sorocaba, amizades herdadas do LAGevol para a vida, como a Msc. Juliana R, Bombonato, Msc. Fernanda M. Leite, Msc. Heidi S. M. Utsonomyia, e Milena C. Telhe, e seus companheiros Wesley Matheus, Jefferson, Luciano e Leonardo.

Aos meus grandes amigos desde a graduação, Msc Bianca M. Vendramini, Dr. Antonio M. C. Neto, Lay G. Basílio, Msc. Eric Y. Kataoka, Msc. Fabia S. Steyer, Dra. Ariana S. Moraes, e Msc. Andreliza Terciotii, Dra. Gabriela de Oliveira, Msc. Tamara Fonseca pelo carinho, companheirismo, reencontros, e por sempre compartilharmos bons momentos e boas risadas.

Ao meu companheiro e melhor amigo Dr. Antonio M. C. Neto que esteve comigo em momentos maravilhosos e difíceis da minha vida pessoal e por todo seu apoio e incentivo nesses momentos desafiadores.

Às minhas gatinhas, Kyoko e Kyara, que supriram com muito aconchego o longo período de pandemia.

Aos meus amigos de infância que permanecem comigo até hoje, em especial, Ana Carolina N. Tírico e Fernanda Casimiro, por carinho e estímulo dessa amizade maravilhosa.

À minha mãe Maria Romeiro e minha irmã Michelle Romeiro Brito, meus exemplos de mulheres fortes e independentes. Muito obrigada pelo apoio incondicional e por estarem presentes em todos os momentos importantes da minha vida.

RESUMO GERAL

A expansão de ambientes áridos e semiáridos no globo é frequentemente associada às alterações climáticas que ocorreram durante o Mioceno. Apesar da relevância desses eventos climáticos para a diversificação de plantas em ambientes xerofíticos, poucos estudos têm sido conduzidos sobre o papel da expansão da aridez como mecanismo de diversificação na América do Sul. Aliados à expansão da aridez, outros fatores bióticos podem estar envolvidos na diversificação de táxons associados a ambientes xéricos, como observado em estudos com representantes da família Cactaceae. Devido à diversificação recente dessa família, filogenias moleculares empregando poucos marcadores moleculares têm tido dificuldade em distinguir as relações de parentesco entre seus principais clados. Uma estratégia para obter maior acurácia e suporte nas inferências filogenéticas é utilizar dados genômicos aliados à abordagens filogenéticas coalescentes. O presente estudo tem como objetivo principal investigar fatores bióticos e abióticos associados à diversificação da tribo Cereeae, uma importante linhagem de cactos da América do Sul, e identificar padrões evolutivos e ecológicos em comum à diversificação de táxons que habitam ambientes xéricos na América do Sul. Para isso, foram estabelecidos os seguintes objetivos específicos: identificar regiões gênicas ortólogas nos genomas disponíveis da família Cactaceae e construir um painel de sondas linhagem-específico para desenvolver estudos evolutivos neste grupo, utilizando a abordagem *target capture sequencing* (Capítulo 1); reconstruir uma filogenia robusta da tribo Cereeae, grupo que reúne a maior diversidade de Cactaceae na América do Sul, utilizando a abordagem *target capture sequencing*. Essa filogenia foi utilizada para definir relações supragenéricas e genéricas bem suportadas, sendo utilizadas para guiar o tratamento taxonômico desse grupo (Capítulo 2); e estimar as datas de divergência e eventos biogeográficos, bem como estimar alterações nas taxas de diversificação e a associação das mesmas com variáveis bióticas e abióticas (Capítulo 3). Foram identificadas mais de quinhentas regiões nucleares ortólogas para Cactaceae, e foi construído um painel de sondas de RNA para resgatar essas regiões, o qual foi nomeado Cactaceae591. Essas regiões foram capazes de estabelecer relações filogenéticas a níveis profundos e rasos dentro da família Cactaceae, como demonstrado no capítulo 1. A partir de dados sequenciados pela abordagem *target capture sequencing* utilizando Cactaceae591 e de dados provenientes de repositórios públicos, foi possível ter uma visão abrangente sobre as relações filogenéticas dentro da tribo Cereeae. Os dados genômicos foram essenciais para obter relações filogenéticas altamente suportadas neste grupo, as quais permitiram delimitar agrupamentos supragenéricos e genéricos na tribo, como demonstrado no capítulo 2. Por fim,

o tempo de divergência estimado para os principais clados da tribo Cereeae coincide com o período de expansão de ambientes áridos durante o Mioceno. O clado composto por cactos globosos da subtribo Cereinae foi o único grupo que apresentou uma alteração na taxa de diversificação. O presente estudo aponta a importância de características morfológicas e ecológicas para a diversificação de grupos que habitam regiões secas da América do Sul, como detalhado pelo capítulo 3.

Palavras-chave: Filogenômica, radiação evolutiva, expansão da aridez, biogeografia, taxa de diversificação, aprendizagem de máquina.

GENERAL ABSTRACT

The global expansion of arid and semiarid habitats is often associated with climate changes that occurred during the Miocene and Pleistocene. Despite the relevance of these climatic events for the diversification of plants in xeric (dry) environments, few studies have been conducted related to the role of the expansion of aridity as a mechanism of diversification in South America. Besides the expansion of aridity, biotic factors may be involved with the diversification of taxon associated with xeric habitats, as observed in previous studies within the Cactaceae family. Due to the recent diversification of this family, molecular phylogenies that used few molecular markers have faced difficulties to distinguish the relationship within its main clades. A reasonable strategy to obtain higher accuracy and resolution in phylogenetic inferences is to employ genomic data combined with coalescence phylogenetic approaches. The present study aims to investigate biotic and abiotic factors associated with the diversification of tribe Cereeae, one of the major lineage of cacti in South America, and identify evolutionary and ecological patterns in common with taxa inhabiting xeric environments in South America. For this purpose, the following specific aims were established: identify ortholog genic regions available for Cactaceae and build a lineage-specific panel for shallow and deep-phylogenetic studies in this group using target capture sequencing approach (Chapter1); reconstruct a robust phylogeny for the tribe Cereeae, a group comprising the major diversity of Cactaceae in South America, using target capture sequencing approach. This phylogeny would be useful to delimitate well-supported suprageneric and generic relationships, and guide new taxonomic treatments (Chapter 2); and estimate divergence time and biogeographic events involved in the diversification of tribe Cereeae, as well as estimate shifts in diversification rates related to biotic and abiotic traits (Chapter 3). In brief, we selected more than five hundred ortholog nuclear regions and built a RNA bait panel, named here as Cactaceae591. As shown in chapter 1, these regions were capable of establishing phylogenetic relationships at deeper and shallower levels in Cactaceae. In chapter 2, we used newly generated target capture sequencing data and public databases to obtain a comprehensive view of phylogenetic relationships within the tribe Cereeae. The genomic dataset showed to be fundamental to obtaining highly supported phylogenetic relationships in this group, being the base to delimit suprageneric and generic groupings in this tribe. Finally, in chapter 3, we recovered that the divergence time estimated in tribe Cereeae is congruent with the period of expansion of aridity in Miocene and Pliocene. The group comprising globose cacti of subtribe Cereinae showed a shift in diversification rates during Pleistocene. The present study endorses

the importance of morphological and ecological characteristics for the diversification of taxa inhabiting dry areas in South America.

Keywords: Phylogenomic, evolutionary radiation, expansion of aridity, diversification rate, machine learning.

FIGURE LIST

General Introduction

Chapter 1

Chapter 2

Chapter 3

TABLE LIST

Chapter 1

Chapter 2

Chapter 3

SUMMARY

GENERAL INTRODUCTION

1. Diversification patterns of xeric habitats in South America

South America is home of the most species-rich and biome-diverse regions of the world (Antonelli and Sanmartín 2011). Multiple hypotheses have been proposed to explain the high diversity of this region (Antonelli et al. 2018a). However, the investigations on the patterns of diversification of these regions have long been focused on humid tropical forests (Turcheto-Zolet et al. 2013, Meseguer et al. 2022), especially exploring the importance of these environments in the formation and connection of the biodiversity in Neotropics (Hoorn et al. 2010, Antonelli et al. 2018b). Conversely, the role of dry forests and open ecosystems in Neotropical diversity is still in its infancy (Särkinen et al. 2011, DRYFLOR et al. 2016, Collevatti et al. 2020), considering the role these environments in the connections between humid forests (Thomé et al. 2016, Ledo and Colli 2017, Masa-Iranzo et al. 2021, Lörch et al. 2021) and for biome transitions (Dexter et al. 2018, Ledo et al. 2020, Zizka et al. 2020, Calió et al. 2022).

The diversification and origin of taxa inhabiting drylands in South America have been subjected to increasing attention for the past decade (reviewed by Werneck 2011, Lima et al. 2018, Abraham et al. 2020; Baranzelli et al. 2020, Luebert 2021). Those regions are composed of continuous and adjacent open formations habitats recently named as eastern and western South American Dry Diagonals (eSADD and wSADD, Luebert 2021). The eSADD formally includes semi-arid and dry domains, such as Caatinga, Cerrado, and Chaco (some studies include Chiquitan, Pantanal and Pampas regions, see Neves et al. 2015, Weber et al. 2019, Thode et al. 2019), while wSADD is composed of arid and hyperarid domains, including Patagonia, Monte, Prepuna, Dry Puna, Atacama Desert and Peruvian Desert (Figure 1). Although each domain of SADDs has unique environmental conditions (Werneck 2011, Lima et al. 2018, Luebert 2021), each SADDs shares fauna and flora similarities (Nogueira et al. 2011, Särkinen et al. 2012, DRYFLOR et al. 2016, Prieto-Torres et al. 2019). To date, few studies have focused on biota similarities and biogeographic events linking or segregating the diversification in both SADDs (Luebert 2021).

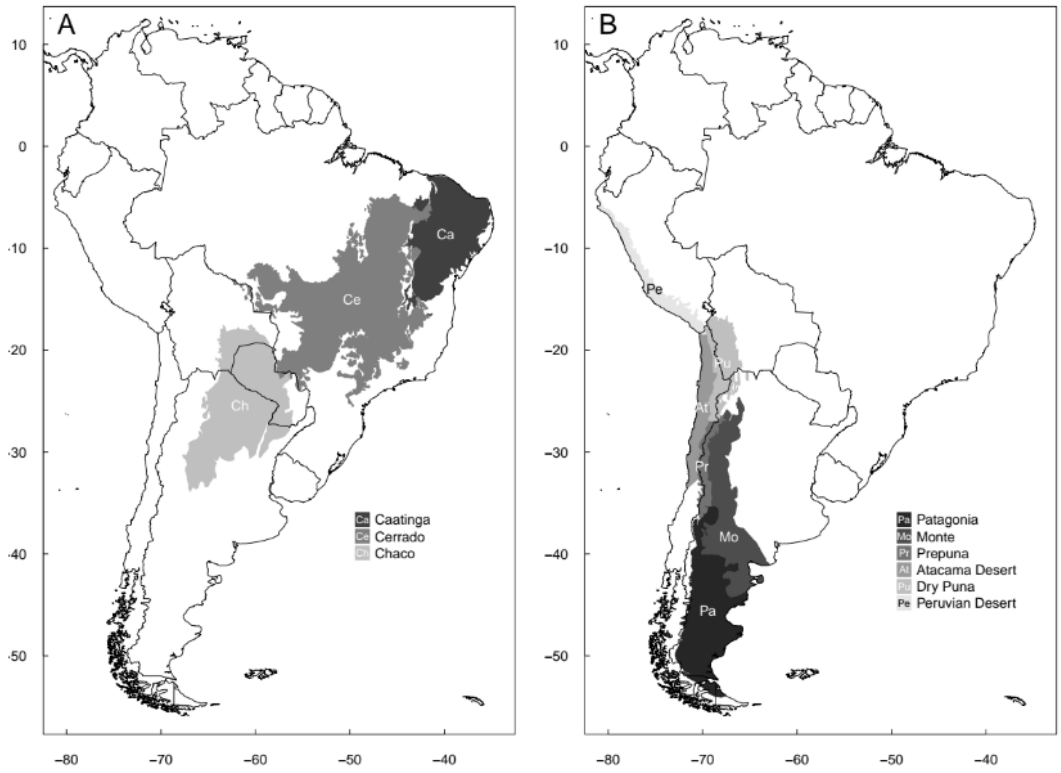


Figure 1. The two South American dry diagonals (SADD). A: Western (wSADD) and B: Eastern (eSADD) (from Luebert, 2021).

According to the Neotropical regionalization proposed by Morrone (2014), the eastern and western SADD are currently assigned to different biogeographic zones, the former corresponding to the Chacoan dominion and the latter to the South American Transition Zone. This distinction may be due to the weak similarities in biota at the species level found in both SADD (Magalhães et al. 2019, Arango et al. 2021, but see Prieto-Torres et al. 2019), and the distinct endemisms of taxa for each zone (Morrone 2014). Despite this, several generic and suprageneric plant lineages are found geographically restricted to eSADD or wSADD, which coincides with the disjunct pattern of seasonally dry tropical forests (SDTFs) in Neotropics (Pennington et al. 2009, Särkinen et al. 2012, DRYFLOR et al. 2016, Gagnon et al. 2019, Luebert et al. 2020, Hurbath et al. 2021). The main hypotheses explaining this long-term disjunction among the SDTF nuclei in Neotropics are related to phylogenetic niche conservatism and dispersal limitations of plant lineages (Pennington et al. 2009, Sarkinen et al. 2012, Hughes et al. 2013, Hurbarth et al. 2021). Moreover, the diversification of major plant lineages inhabiting SDTF nuclei had occurred simultaneously around the globe, where most lineages emerged in the Eocene and Oligocene, but diverged in the mid and late Miocene (Arakaki et al. 2011; Almeida et al. 2018; Gagnon et al. 2019; Hurbath et al. 2021), which

suggest a common trigger promoting the diversification of succulent plant lineages in South America.

The intense Andean uplift during the Miocene has been traditionally considered one of the primary geological events to promote the diversification of Neotropical biota (Hoorn et al. 2010, Luebert and Weigend 2014). Nonetheless, in this epoch, several concomitant events may have promoted the expansion of arid environments and the diversification of grasslands and succulent plants, such as the global cooling and the decline of atmospheric CO_2 (Figure 2, Arakaki et al. 2011, Herbert et al. 2016, Gutiérrez-Ortega et al. 2018). The decrease in sedimentation rates in Miocene and Pliocene had also revealed an expansion of arid environments at the regional scale in Central and Southern Andes (Amidon et al. 2017), and may have corroborated the diversification of succulent plant lineages in South America (Amaral et al. 2021a; Fernandes et al. 2022). Beyond aridity in itself, the exhumation of sediments and erosion of rocks generating topographic and edaphic heterogeneity has a relevant role in promoting the in situ diversification of SDTF and xeric lineages in Miocene (Fernandes et al. 2022, Pérez-Escobar et al. 2022).

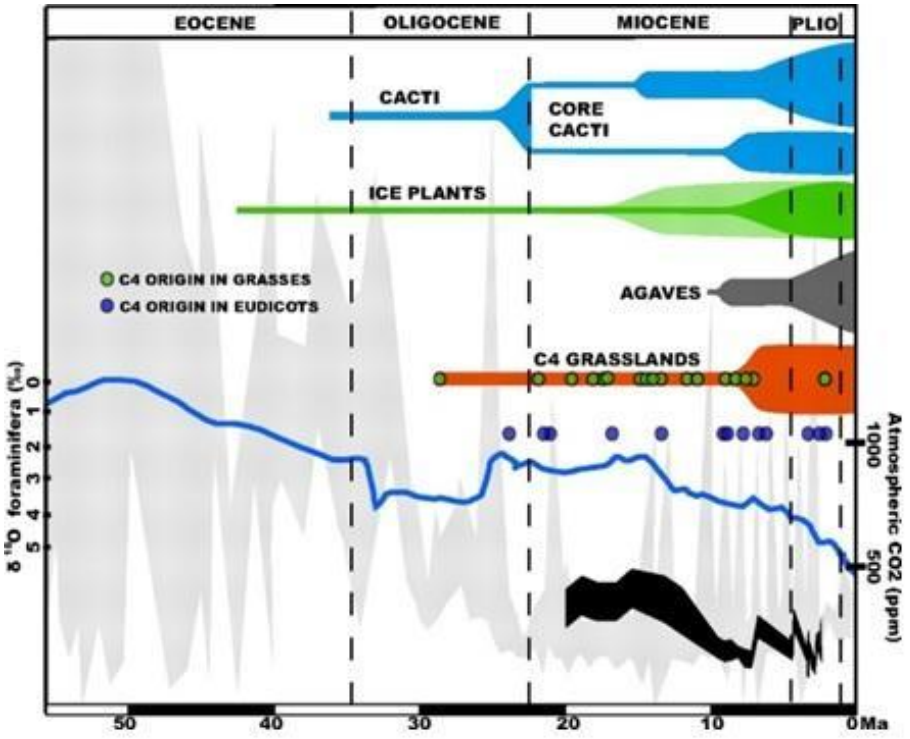


Figure 2. Variation of raw atmospheric CO_2 (in gray) and its period of decline in Miocene (in black); relative global temperature (line in blue); origin and diversification of C4 and succulent plant lineages (from Arakaki et al. 2011).

While abiotic factors have been extensively linked with the diversification of succulent lineages (Arakaki et al. 2011, Gutiérrez-Ortega et al. 2017), little is known about the effect of biotic traits shaping the diversification of dry-adapted lineages in Neotropics. Considering several adaptive morphological strategies of plant lineages for inhabiting extreme drought habitats (Folk et al. 2020, Fagundes et al. 2020, Oliveira et al. 2020), both biotic and abiotic traits may have contributed to the diversification of lineages inhabiting xeric biomes. Meanwhile, growth form and reproductive traits were associated with the increase of diversification in several succulent plant lineages (e.g Hernández-Hernández et al. 2014, Tripp and Tsai 2017, Alcantara et al. 2018, Larocca et al. 2022) and with biome shift among dry and semiarid regions in South America (Amaral et al. 2021). Therefore, the interaction of abiotic and biotic traits could help understand the complex patterns involving the rapid diversification events, and biogeographic disjunction patterns among xeric areas in South America.

2. Tribe Cereeae (Caryophylalles, Cactaceae) as a biological model for investigating diversification patterns in xeric habitats of South America

Cactaceae is one of the most diverse plant groups among the succulent plants of the order Caryophyllales (Hernández-Ledesma et al. 2015, Folk et al. 2021). The family harbors more than 1,800 species (Korotkova et al. 2021) and is one of the most threatened groups in the world (Goettsch et al. 2015). Cacti species are well-known for their conspicuous adaptation strategies, such as the loss and reduction of leaves, shoots modified into areoles, and the development of succulent cortex and parenchyma (Gibson and Nobel 1986, Anderson 2001; Hernández-Hernández et al. 2011). The family distribution is predominantly associated with dry and arid environments of the American continent (Guerrero et al. 2018). South America holds three centers of diversity for Cactaceae: one in the Central Andes and two in Eastern Brazil (in Caatinga and in the Southeastern Atlantic Forest, Barthlott et al. 2015).

This family provides a model for studying adaptive patterns in dry and arid habitats in South America, mainly the tribe Cereeae sensu lato Nyffeler and Eggli (2010). The tribe Cereeae, originally comprising cereoid groups as Cereus and Praecereus (Anderson 2001, 2005). However, the molecular phylogenetic studies were fundamental to propose a new circunscription, also known as the BCT clade, joining taxa from tribes Browningeae, Cereeae, and Trichocereeae (Nyffeler 2002, Hernández-Hernández et al. 2011). A broadened circumscription for this group was applied using molecular phylogenetics, but many subtribe groups were revealed as polyphyletic (Guerrero et al. 2018). The former tribe Cereeae s.l. includes the major diversity of Cactaceae in South America, comprising few genera that reach Caribbean and Mesoamerica regions (*Pilosocereus*, *Melocactus*, *Cereus*, *Harrisia*, Fig. 3). This group presents a disjunct distribution in South America and is mainly found in Central-Southern Andes and in Eastern Brazil (Fig. 3, Barthlott et al. 2015).

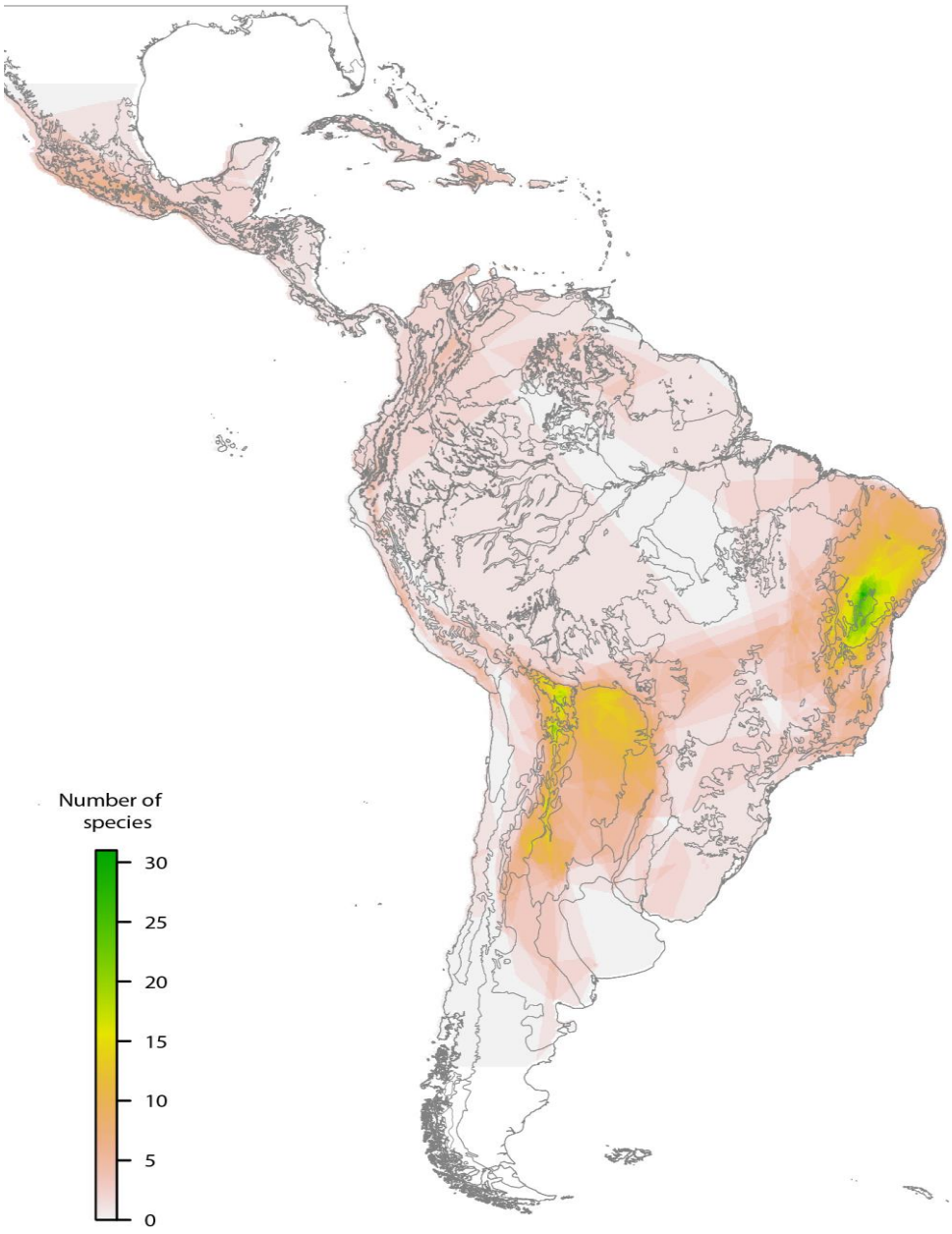


Figure 3. Distribution and richness of species from tribe Cereeae across the Neotropical ecoregions delimited by Dinerstein et al. (2017). The map was produced by overlaying species range maps estimated from geographic coordinates obtained in GBIF and specieslink and with R package SpeciesGeoCoder (Töpel et al. 2016).

The tribe exhibits one of the highest diversification rates within the family as a putative result of a combination of climatic, morphological, and ecological conditions (Hernández-Hernández et al. 2014). The diversification of this tribe coincides with the global expansion of aridity in the Miocene, and its higher diversification rates are associated with derived pollination syndromes and columnar growth forms (Hernández-Hernández et al. 2014;

examples of variation of flower and growth forms in Fig. 4). To date, few studies have explored other morphological and ecological traits that may have underpinned the high diversification rates within this group (e.g. Amaral et al. 2021a). Indeed, this group faces a huge challenge due to the lack of phylogenetic support between suprageneric and generic families (Schlumpberger and Renner 2012, Lendel, 2013, Fantinati et al. 2021) and the non-monophyly of several genera as currently circumscribed. By so, a robust and well-supported phylogenetic hypothesis is fundamental to follow towards the inference of the major phylogenetic relationships in this tribe, allowing to assess in an evolutionary framework the biotic and abiotic traits involved in the diversification and biogeographical patterns.

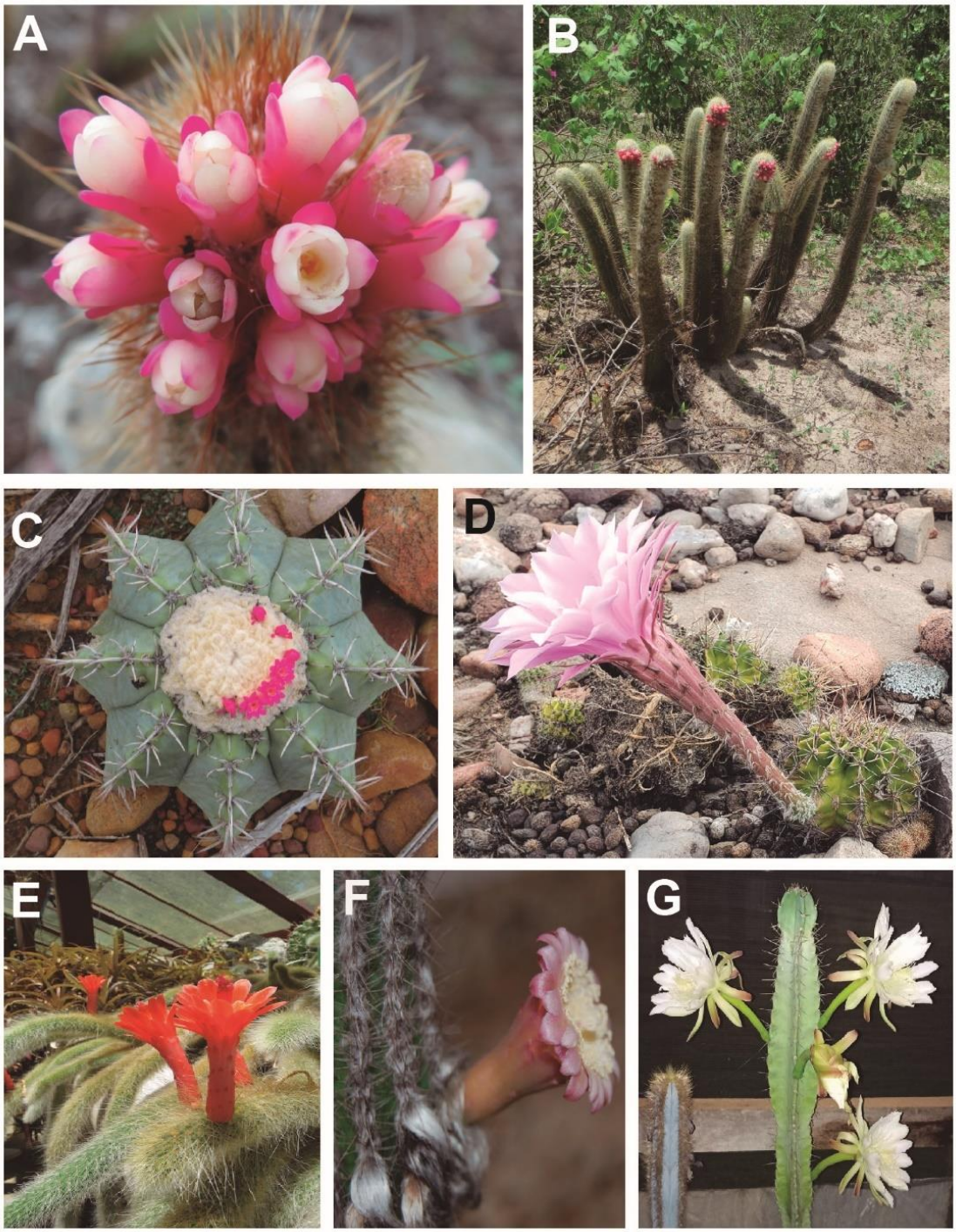


Figure 4. Representatives of growth form, flower form, and pollination syndrome variation in tribe Cereeae s.l.. Growth form can be roughly typified as columnar (photos B and G) and globose (photos C and D). Flowers range from 10 mm, as found in representatives of *Micranthocereus* Backeb. (photos A and B) and *Melocactus* Link & Otto (photo C) to 200 mm, as found in *Echinopsis* Zucc. (photo D) and *Cereus* Mill. (photo G). This group is pollinated by hummingbirds (photos A, C, and E), moths (photos D and G), and bats (photo F). A and B: *Micranthocereus polyanthus* (Werderm.) Backeb.; C) *Melocactus glaucescens* Buining & Brederoo; D) *Echinopsis oxygona* (Link) Zucc. ex Pfeiff.; E) *Cleistocactus winteri* D.R.Hunt; F) *Pilosocereus albisummus* P.J.Braun & Esteves; G) *Cereus hankeanus* F.A.C. Weber. Photo credits: A, B, C and E: M.C. Telhe; D: M. Köhler; F: G. Olsthoorn; G: M. Romeiro-Brito.

3. Resolving phylogenetic relationships in recent radiations of cactus lineages

The recent radiation of cacti lineages and long generation time may lead to a lack of phylogenetic signal (Townsend 2007) and increasing incomplete lineage sorting (Maddison 1997) in this group, which make relationship resolution challenging within major groups in Cactaceae (Copetti et al. 2017). Whereas the lack of phylogenetic relationships results in polytomies (e.g. Calvente et al. 2017), the incomplete lineage sorting may produce misleading phylogenetic relationships due to incongruent histories among gene trees and species trees (e.g. Romeiro-Brito et al. 2023). The use of independent nuclear loci combined with phylogenetic inferences that deal with gene tree heterogeneity may overcome both concerns (Edwards et al. 2016).

As high-throughput sequencing (next-generation sequencing) techniques become affordable for non-model organisms, we can obtain multiple independent regions for evolutionary studies in plants (McKain et al. 2020). Cactaceae also follow this trend, given the increasing use of next-generation approaches in the last decade (Franco et al. 2022). Several genomic resources are now available for this group (e.g. Copetti et al. 2017, Walker et al. 2018, Wang et al. 2019, Amaral et al. 2021b), which enables the screening of phylogenetic informative regions to resolve evolutionary relationships at multiple levels in Cactaceae. These ortholog regions are easily isolated through target enrichment strategies, a technique that has become increasingly popular in phylogenetic and systematic studies in plants (McDonnel et al. 2021, Baker et al. 2021). Hence, obtaining phylogenetic informative regions for Cactaceae through target enrichment sequencing provides a promising tool for a better understanding of complex evolutionary patterns within this family (e.g. Acha and Majure 2022).

4. Aims of this thesis

The main aim of this study is to investigate the role of biotic and abiotic traits in the diversification of taxa inhabiting xeric areas in South America, using the tribe Cereeae as a biological model. We aimed to answer the following questions: Did the main lineages of the tribe Cereeae diversify at the same time as the expansion of aridity in South America, during

the Miocene and Pliocene epochs (Amidon et al. 2017)? Were the diversification rates in this group associated with abiotic (climatic and edaphic metrics) and biotic traits (morphological and functional traits)? Which biogeographic events were involved in the disjunction pattern among Eastern Brazil and Central-Southern Andes? For this, we set the following specific aims:

(1) To develop a lineage-specific probe set isolating hundreds of ortholog nuclear regions by target sequencing approach to assess a well-supported phylogeny across multiple levels in the family Cactaceae (Chapter 1);

(2) To obtain a robust phylogenetic inference and delimit suprageneric and generic relationships within the tribe Cereeae, comparing the performance of concatenated and coalescent methods (Chapter 2);

(3) To estimate divergence times and main biogeographic events related to the diversification of the major lineages of tribe Cereeae, and infer the ancestral range distribution of these groups (Chapter 3);

(4) To estimate time-dependent and trait-dependent diversification rates in tribe Cereeae and identify biotic and abiotic traits associated with diversification rates using machine learning (Chapter 3).

To maintain the flow of reading for readers and curious, I would suggest including here a brief paragraph presenting the following contents and organization of your thesis, e.g.:

In the following, each chapter is presented as individual manuscripts already published or prepared for submission to publication in scientific journals....

5. References

Abraham E. M., Rodríguez M. D., Rubio M. C., Guida-Johnson B., Gomez L. and Rubio, C. (2020). Disentangling the concept of “South American Arid Diagonal”. Journal of Arid Environments, 175, 104089.

Alcantara S., Ree R. H. and Mello-Silva, R. (2018). Accelerated diversification and functional trait evolution in Velloziaceae reveal new insights into the origins of the campos rupestres’ exceptional floristic richness. Annals of Botany, 122, 165-180.

Almeida F. R., Amorim A. M. A. and van den Berg, C. (2018). Timing the origin and past connections between Andean and Atlantic Seasonally Dry Tropical Forests in South America: Insights from the biogeographical history of *Amorimia* (Malpighiaceae). Taxon, 67, 739-751.

Amidon W. H., Fisher G. B., Burbank D. W., Ciccioli P. L., Alonso R. N., Gorin A. L .et al. (2017). Mio-Pliocene aridity in the south-central Andes associated with Southern Hemisphere cold periods. Proceedings of the National Academy of Sciences, 114, 6474-6479.

Amaral D. T., Minhós-Yano I., Oliveira J. V. M., Romeiro-Brito M., Bonatelli I. A. S., Taylor N. P. et al. (2021a). Tracking the xeric biomes of South America: The spatiotemporal diversification of Mandacaru cactus. Journal of Biogeography, 48, 3085-3103.

Amaral D. T., Bombonato J. R., da Silva Andrade S. C., Moraes E. M. and Franco, F. F. (2021b). The genome of a thorny species: comparative genomic analysis among South and North American Cactaceae. Planta, 254, 44.

Anderson E. F. (2001) The cactus family. Timber Press, Portland, Oregon, USA

Antonelli A., Ariza M., Albert J., Andermann T., Azevedo J., Bacon C. et al. (2018a). Conceptual and empirical advances in Neotropical biodiversity research. PeerJ, 6, e5644.

Antonelli A., Zizka A., Carvalho F. A., Scharn R., Bacon C. D., Silvestro D. and Condamine, F. L. (2018b). Amazonia is the primary source of Neotropical biodiversity. Proceedings of the National Academy of Sciences, 115, 6034-6039.

Antonelli A. and Sanmartín I. (2011). Why are there so many plant species in the Neotropics?. Taxon, 60, 403-414.

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. Proceedings of the National Academy of Sciences, 108, 8379-8384.

Arango A., Villalobos F., Prieto-Torres D. A. and Guevara R. (2021). The phylogenetic diversity and structure of the seasonally dry forests in the Neotropics. Journal of Biogeography, 48, 176-186.

Baker W. J., Dodsworth S., Forest F., Graham S. W., Johnson M. G., McDonnell A. et al. (2021). Exploring Angiosperms353: An open, community toolkit for collaborative phylogenomic research on flowering plants. American Journal of Bontany, 108, 1059-1065.

Baranzelli M. C., Cosacov A., Rocamundi N., Issaly E. A., Aguilar D. L., Camps G. A. et al. (2020). Volcanism rather than climatic oscillations explains the shared phylogeographic patterns among ecologically distinct plant species in the southernmost areas of the South

American Arid Diagonal. Perspectives in Plant Ecology, Evolution and Systematics, 45, 125542.

Barthlott W., Burstedde K., Geffert J. L., Ibisch P. L., Korotkova N., Miebach A. et al. (2015). Biogeography and biodiversity of cacti. Universitätsverlag Isensee.

Calió M. F., Thode V. A., Bacon C. D., Silvestro D., Antonelli A. and Lohmann, L. G. (2022). Spatio-temporal evolution of the catuaba clade in the Neotropics: Morphological shifts correlate with habitat transitions. Journal of Biogeography, 49, 1086-1098.

Calvente A., Moraes E. M., Lavor P., Bonatelli I. A., Nacaguma P., Versieux L. M. et al. (2017). Phylogenetic analyses of *Pilosocereus* (Cactaceae) inferred from plastid and nuclear sequences. Botanical Journal of the Linnean Society, 183, 25-38.

Collevatti R. G., Lima N. E. and Vitorino, L. C. (2020). The Diversification of Extant Angiosperms in the South America Dry Diagonal. In Neotropical Diversification: Patterns and Processes (pp. 547-568). Springer, Cham

Copetti D., Búrquez A., Bustamante E., Charboneau J. L., Childs K. L., Eguiarte L. E. et al. (2017). Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. Proceedings of the National Academy of Sciences, 114, 12003-12008.

Dexter K. G., Pennington R. T., Oliveira-Filho A. T., Bueno M. L., Miranda P. L. S. and Neves D. M. (2018). Inserting tropical dry forests into the discussion on biome transitions in the tropics. Frontiers in Ecology and Evolution, 6, 104.

Dinerstein E., Olson D., Joshi A., Vynne C., Burgess N. D. Wikramanayake E. et al. (2017). An ecoregion-based approach to protecting half the terrestrial realm. BioScience, 67, 534-545.

Dryflor, Banda-R K., Delgado-Salinas A., Dexter K. G., Linares-Palomino R., Oliveira-Filho A. et al. (2016). Plant diversity patterns in neotropical dry forests and their conservation implications. Science, 353, 1383-1387.

Edwards S. V., Xi Z., Janke A., Faircloth B. C., McCormack J. E., Glenn T. C. et al. (2016). Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Molecular Phylogenetics and Evolution, 94, 447-462.

Fagundes M. V., Souza A. F. and Ganade, G. (2022). Functional traits above and below ground allow species with distinct ecological strategies to coexist in the largest seasonally dry tropical forest in the Americas. Frontiers in Forests and Global Change, 5:9300

Fantinati M. R. Soffiatti, P. and Calvente, A. (2021). A new phylogenetic hypothesis for Cereinae (Cactaceae) points to a monophyletic subtribe. Systematic Botany, 46, 689-699.

Fernandes M. F., Cardoso D., Pennington R. T. and de Queiroz, L. P. (2022). The origins and historical assembly of the Brazilian Caatinga seasonally dry tropical forests. Frontiers in Ecology and Evolution, 10,723286.

Folk R. A., Siniscalchi C. M. and Soltis, D. E. (2020). Angiosperms at the edge: Extremity, diversity, and phylogeny. Plant, Cell & Environment, 43, 2871-2893.

Franco F. F., Amaral D. T., Bonatelli I. A., Romeiro-Brito M., Telhe M. C. and Moraes, E. M. (2022). Evolutionary genetics of cacti: research biases, advances and prospects. Genes, 13, 452.

Gagnon E., Ringelberg J. J., Bruneau A., Lewis G. P. and Hughes C. E. (2019). Global succulent biome phylogenetic conservatism across the pantropical Caesalpinia group (Leguminosae). New Phytologist, 222, 1994-2008.

Gibson A. C. and P. S. Nobel. (1986). The cactus primer. Harvard University Press, Cambridge, Massachusetts, USA

Goettsch B., Hilton-Taylor C., Cruz-Piñón G., Duffy J. P., Frances A., Hernández H. M. et al.. (2015). High proportion of cactus species threatened with extinction. Nature plants, 1, 1-7.

Guerrero P. C., Majure L. C., Cornejo-Romero A. and Hernández-Hernández T. (2018). Phylogenetic relationships and evolutionary trends in the cactus family. Journal of Heredity, 110, 4-21.

Gutiérrez-Ortega J. S., Yamamoto T., Vovides A. P., Angel Pérez-Farrera M., Martinez J. F., Molina-Freaner F. et al. (2018). Aridification as a driver of biodiversity: a case study for the cycad genus *Dioon* (Zamiaceae). Annals of Botany, 121, 47-60.

Herbert T. D., Lawrence K. T., Tzanova A., Peterson L. C., Caballero-Gill R. and Kelly C. S. (2016). Late Miocene global cooling and the rise of modern ecosystems. Nature Geoscience, 9, 843-847.

Hernández-Hernández T., Hernández H. M., De-Nova J. A., Puente R., Eguiarte L. E. and Magallón S. (2011). Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). American Journal of Botany, 98, 44-61.

Hernández-Hernández 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 Phytologist, 202, 1382-1397.

Hernández-Ledesma P., Berendsohn W. G., Borsch T., Von Mering S., Akhani H., Arias S. et al. (2015). A taxonomic backbone for the global synthesis of species diversity in the angiosperm order Caryophyllales. Willdenowia, 45, 281-383.

Hoorn C., Wesselingh F. P., Ter Steege H., Bermudez M. A., Mora A., Sevink J. et al. (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science, 330, 927-931.

Hughes C. E., Pennington R. T. and Antonelli A. (2013). Neotropical plant evolution: assembling the big picture. Botanical Journal of the Linnean Society, 171, 1-18.

Hurbath F., Stubbs R. L., Cordeiro I. and Cellinese N. (2021). Biogeography of succulent spurges from Brazilian Seasonally Dry Tropical Forest (SDTF). Taxon, 70, 153-169.

Korotkova N., Aquino D., Arias S., Eggli U., Franck A., Gómez-Hinostrosa C. et al. (2021). Cactaceae at Caryophyllales. org–a dynamic online species-level taxonomic backbone for the family. Willdenowia, 51, 251-270.

Larocca P. F., Mancio J. S., Padilha P., Mello-Silva R. and Alcantara, S. (2022). Recent divergence in functional traits affects rates of speciation in the Neotropical Velloziaceae (Pandanales). Botanical Journal of the Linnean Society, 199, 144-172.

Ledo R. M. D. and Colli G. R. (2017). The historical connections between the Amazon and the Atlantic Forest revisited. Journal of biogeography, 44, 2551-2563.

Ledo R. M. D., Domingos F. M., Giugliano L. G., Sites Jr J. W., Werneck F. P. and Colli, G. R. (2020). Pleistocene expansion and connectivity of mesic forests inside the South American Dry Diagonal supported by the phylogeography of a small lizard. Evolution, 74, 1988-2004.

Lendel, A. (2013). South American Cacti in time and space: studies on the diversification of the tribe Cereeae, with particular focus on subtribe Trichocereinae (Cactaceae). PhD Thesis, University of Zurich, Switzerland.

Lima N. E. D., Carvalho A. A., Lima-Ribeiro M. S. and Manfrin, M. H. (2018). Caracterização e história biogeográfica dos ecossistemas secos neotropicais. Rodriguésia, 2209-2222.

Lörch M., Mutke J., Weigend M. and Luebert, F. (2021). Historical biogeography and climatic differentiation of the *Fulcaldea-Archidasyphyllum-Arnaldoa* clade of Barnadesioideae (Asteraceae) suggest a Miocene, aridity-mediated Andean disjunction associated with climatic niche shifts. Global and Planetary Change, 201, 103495.

Luebert F. and Weigend M. (2014). Phylogenetic insights into Andean plant diversification. Frontiers in Ecology and Evolution, 2, 27.

Luebert F., Lörch, M., Acuña R., Mello-Silva R., Weigend M. and Mutke, J. (2020). Clade-specific biogeographic history and climatic niche shifts of the southern Andean-southern Brazilian disjunction in plants. In Neotropical Diversification: Patterns and Processes (pp. 661-682). Springer, Cham

Luebert F. (2021). The two South American dry diagonals. Frontiers of Biogeography, 13.

Maddison W. P. (1997). Gene trees in species trees. Systematic biology, 46, 523-536.

Magalhaes I. L. F., Neves D. M., Santos F. R., Vidigal T. H. D. A., Brescovit A. D. and Santos A. J. (2019). Phylogeny of Neotropical *Sicarius* sand spiders suggests frequent transitions from deserts to dry forests despite antique, broad-scale niche conservatism. Molecular Phylogenetics and Evolution, 140, 106569.

Masa-Iranzo I., Sanmartín I., Caruzo M. B. R. and Riina, R. (2021). Skipping the Dry Diagonal: spatio-temporal evolution of *Croton* section *Cleodora* (Euphorbiaceae) in the Neotropics. Botanical Journal of the Linnean Society, 197, 61-84.

McDonnell A. J., Baker W. J., Dodsworth S., Forest F., Graham S. W., Johnson M. G. and Wickett N. J. (2021). Exploring Angiosperms353: Developing and applying a universal toolkit for flowering plant phylogenomics. Applications in plant sciences, 9, e11443

McKain M. R., Johnson M. G., Uribe-Convers S., Eaton D. and Yang, Y. (2018). Practical considerations for plant phylogenomics. Applications in plant sciences, 6, e1038.

Meseguer A. S., Michel A., Fabre P. H., Escobar O. A. P., Chomicki G., Riina R. et al. (2022). Diversification dynamics in the Neotropics through time, clades, and biogeographic regions. Elife, 11, e74503.

Morrone J. J. (2014). Cladistic biogeography of the Neotropical region: identifying the main events in the diversification of the terrestrial biota. Cladistics, 30, 202-214.

Neves D. M., Dexter K. G., Pennington R. T., Bueno M. L. and Oliveira Filho A. T. (2015). Environmental and historical controls of floristic composition across the South American Dry Diagonal. Journal of Biogeography, 42, 1566-1576.

Nogueira C., Ribeiro S., Costa G. C. and Colli G. R. (2011). Vicariance and endemism in a Neotropical savanna hotspot: distribution patterns of Cerrado squamate reptiles. Journal of Biogeography, 38, 1907-1922.

Nyffeler, R. (2002). Phylogenetic relationships in the cactus family (Cactaceae) based on evidence from trnK/matK and trnL-trnF sequences. American Journal of Botany, 89, 312-326.

Nyffeler, R. and Eggli, U. (2010). A farewell to dated ideas and concepts: molecular phylogenetics and a revised suprageneric classification of the family Cactaceae. Schumannia, 6, 109-149.

Oliveira A. C. P., Nunes A., Rodrigues R. G. and Branquinho, C. (2020). The response of plant functional traits to aridity in a tropical dry forest. Science of The Total Environment, 747, 141177.

Pennington R. T., Lavin M. and Oliveira-Filho A. (2009). Woody plant diversity, evolution, and ecology in the tropics: perspectives from seasonally dry tropical forests. Annual Review of Ecology, Evolution, and Systematics, 40, 437-457.

Pérez-Escobar O. A., Zizka A., Bermúdez M. A., Meseguer A. S., Condamine F. L., Hoorn C. et al. (2022). The Andes through time: evolution and distribution of Andean floras. Trends in Plant Science, 27, 364-378.

Prieto-Torres D. A., Rojas-Soto O. R., Bonaccorso E., Santiago-Alarcon D. and Navarro-Sigüenza, A. G. (2019). Distributional patterns of Neotropical seasonally dry forest birds: a biogeographical regionalization. Cladistics, 35, 446-460.

Romeiro-Brito M., Khan G., Perez M. F., Zappi D. C., Taylor N. P., Olsthoorn G. et al. (2023). Revisiting phylogeny, systematics, and biogeography of a Pleistocene radiation. American Journal of Botany, 110, e16134.

Särkinen T., Iganci J. R., Linares-Palomino R., Simon M. F. and Prado, D. E. (2011). Forgotten forests-issues and prospects in biome mapping using Seasonally Dry Tropical Forests as a case study. BMC Ecology, 11, 1-16.

Särkinen T., Pennington R. T., Lavin M., Simon M. F. and Hughes C. E. (2012). Evolutionary islands in the Andes: persistence and isolation explain high endemism in Andean dry tropical forests. Journal of Biogeography, 39, 884-900.

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. American Journal of Botany, 99, 1335-1349.

Thode V. A., Sanmartín I. and Lohmann, L. G. (2019). Contrasting patterns of diversification between Amazonian and Atlantic forest clades of Neotropical lianas (*Amphilophium*, Bignonieae) inferred from plastid genomic data. Molecular Phylogenetics and Evolution, 133, 92-106.

Thomé M. T. C., Sequeira F., Brusquetti F., Carstens B., Haddad C. F., Rodrigues M. T. and Alexandrino J. (2016). Recurrent connections between Amazon and Atlantic forests shaped diversity in Caatinga four-eyed frogs. Journal of Biogeography, 43, 1045-1056.

Töpel M., Zizka A., Calió M. F., Scharn R., Silvestro D. and Antonelli A. (2017). SpeciesGeoCoder: Fast categorization of species occurrences for analyses of biodiversity, biogeography, ecology, and evolution. Systematic Biology, 66, 145-151.

Townsend J. P. (2007). Profiling phylogenetic informativeness. Systematic Biology, 56, 222-231.

Tripp E. A. and Tsai, Y. H. E. (2017). Disentangling geographical, biotic, and abiotic drivers of plant diversity in neotropical *Ruellia* (Acanthaceae). PLoS One, 12, e0176021.

Turchetto-Zolet A. C., Pinheiro F., Salgueiro F. and Palma-Silva, C. (2013). Phylogeographical patterns shed light on evolutionary process in South America. Molecular Ecology, 22, 1193-1213.

Walker J. F., Yang Y., Feng T., Timoneda A., Mikenas J., Hutchison V. et al. (2018). From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. American Journal of Botany, 105, 446-462.

Wang N., Yang Y., Moore M. J., Brockington S. F., Walker J. F., Brown J. W. et al. (2019). Evolution of Portulacineae marked by gene tree conflict and gene family expansion associated with adaptation to harsh environments. Molecular biology and evolution, 36, 112-126.

Weber M. M., Novaes R. L. M., Delgado-Jaramillo M., Barbier E., Cláudio V. C., Bernard E. and Moratelli, R. (2019). Is *Myotis lavali* (Chiroptera, Vespertilionidae) endemic to the South American dry diagonal?. Journal of Mammalogy, 100, 1879-1888.

Werneck, F. P. (2011). The diversification of eastern South American open vegetation biomes: historical biogeography and perspectives. Quaternary Science Reviews, 30(13-14), 1630-1648.

Zizka A., Carvalho-Sobrinho J. G., Pennington R. T., Queiroz L. P., Alcantara S., Baum D. A. et al. (2020). Transitions between biomes are common and directional in Bombacoideae (Malvaceae). Journal of Biogeography, 47, 1310-1321.

CHAPTER 1: A target capture probe set useful for deep- and shallow-level phylogenetic studies in Cactaceae

Monique Romeiro-Brito[1], Milena C. Telhe[1], Danilo T. Amaral[2], Fernando F. Franco[1], Evandro M. Moraes[1]

[1] Departamento de Biologia. Centro de Ciências Humanas e Biológicas. Universidade Federal de São Carlos (UFSCar). Sorocaba, Brazil.

[2] Programa de Pós-graduação em Biologia Comparada. Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto. Universidade de São Paulo (USP). Ribeirão Preto, Brazil

Published in Genes 2022, 13, 707. https://doi.org/10.3390/ genes130407

genes

MDPI

Article

A target Capture Probe Set Useful for Deep- and Shallow-Level Phylogenetic Studies in Cactaceae

Monique Romeiro-Brito [1], Milena Cardoso Telhe [1], Danilo Trabuco Amaral [1,2], Fernando Faria Franco [1] and Evandro Marsola Moraes [1,*]

[1] Departamento de Biologia, Centro de Ciências Humanas e Biológicas, Universidade Federal de São Carlos (UFSCar), Sorocaba 18052-780, SP, Brazil; romeiro.monique@gmail.com (M.R.-B.); milena.telhe@gmail.com (M.C.T.); danilo.trabuco@gmail.com (D.T.A.); franco@ufscar.br (F.F.F.)

[2] Programa de Pós-graduação em Biologia Comparada, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo (USP), Ribeirão Preto 14040-901, SP, Brazil

* Correspondence: emarsola@ufscar.br

Abstract: The molecular phylogenies of Cactaceae have enabled us to better understand their systematics, biogeography, and diversification ages. However, most of the phylogenetic relationships within Cactaceae major groups remain unclear, largely due to the lack of an appropriate set of molecular markers to resolve its contentious relationships. Here, we explored the genome and transcriptome assemblies available for Cactaceae and identified putative orthologous regions shared among lineages of the subfamily Cactoideae. Then we developed a probe set, named Cactaceae591, targeting both coding and noncoding nuclear regions for representatives from the subfamilies Pereskioideae, Opuntioideae, and Cactoideae. We also sampled inter- and intraspecific variation to evaluate the potential of this panel to be used in phylogeographic studies. We retrieved on average of 547 orthologous regions per sample. Targeting noncoding nuclear regions showed to be crucial to resolving inter- and intraspecific relationships. Cactaceae591 covers 13 orthologous genes shared with the Angiosperms353 kit and two plastid regions largely used in Cactaceae studies, enabling the phylogenies generated by our panel to be integrated with angiosperm and Cactaceae phylogenies, using these sequences. We highlighted the importance of using coalescent-based species tree approaches on the Cactaceae591 dataset to infer accurate phylogenetic trees in the presence of extensive incomplete lineage sorting in this family.

Keywords: phylogenomics; cacti; coalescent inference; rapid radiation; target capture sequencing

Citation: Romeiro-Brito, M.; Telhe, M.C.; Amaral, D.T.; Franco, F.F.; Moraes, E.M. A target Capture Probe Set Useful for Deep- and Shallow-Level Phylogenetic Studies in Cactaceae. *Genes* **2022**, *13*, 707. https://doi.org/10.3390/genes13040707

Academic Editor: Roberta Bergero

Received: 18 March 2022
Accepted: 15 April 2022
Published: 17 April 2022

1. Introduction

Molecular phylogenetics has been a fundamental approach to understand the systematics and diversification of major plant groups, such as in the family Cactaceae [1]. The use of a few traditional molecular markers (e.g., plastid DNA regions) has been the primary source for phylogenetic and dating inference in Cactaceae [2]. However, the evolutionary relationship within tribes and genera remains largely inconclusive, partly due to the lack of phylogenetic signals in these molecular markers (see References [3–7] for examples). The poor phylogenetic resolution within Cactaceae is likely a reflection of the recent and rapid diversification experienced by the family since the late Miocene [8]. This tempo and mode of diversification may generate complex patterns of gene divergence, including large-scale incomplete lineage sorting (deep coalescence; see Reference [9]). Moreover, the reticulate evolution events experienced for some clades may also obscure the phylogenetic signals in recent radiation groups [10]. Accessing multiple independent loci allied to phylogenetic approaches that allow for gene tree heterogeneity [11] may be a better strategy to resolve challenging groups and recover highly supported phylogenies within this family.

For the past few decades, high-throughput sequencing techniques have become more affordable for non-model plants [12] and, more recently, have been applied to evolutionary

studies in Cactaceae (see References [13–15]). Different sequencing strategies have fueled phylogenomic studies in cacti according to the research interest, including whole-genome (WGS) and transcriptome sequencing [14,16], reduced genome representation libraries (e.g., RAD-seq and GBS; see References [17,18]), and genome skimming for whole-plastome assembly [19–22]. In addition to allowing the test of evolutionary hypotheses in ways that were not previously feasible, these data represent an invaluable resource to survey for phylogenetically informative markers able to resolve contentious relationships at multiple levels of Cactaceae diversity. For instance, screening these genomic data for single-copy loci can provide a customized set of markers that is able to generate robust phylogenomic hypotheses when combined with hybridization-based target sequencing. Although all genomic strategies display pros and cons, especially regarding non-model species, the target-sequencing approach has intrinsic advantages. Among them, the most important are (i) the feasibility of capturing homologous single-copy genes across many individuals and species, thus decreasing the frequency of missing data and complicating downstream analyses; (ii) the requirement of a lower quality of DNA extraction, allowing for the use of herbarium specimens; and (iii) the more effective data integration with previous datasets [12].

Target sequencing by hybridization capture in plants is a cost-effective enrichment method for sequencing hundreds of putatively single-copy nuclear regions in multiple samples [23]. By screening the coding sequence variations in transcriptome data from related taxa, researchers can design sequence capture probes targeting orthologous coding sequences conserved across taxa. Furthermore, target sequencing allows us to capture specific single-copy genes of interest, such as those candidates to be under selection, which are accessed only casually, using non-targeted sequencing strategies. This method has become popular in angiosperm systematics and phylogenetic studies [24,25] and is largely fostered by the Angiosperms353 kit [26]. Acha and Majure [27] used Cactaceae genomic resources to develop a customized probe set for capturing coding sequences of 120 single-copy nuclear loci and complemented it with a subset of the Angiosperms353 probe set. Although universal kits such as Angiosperms353 have been successfully applied at both deep and shallow phylogenetic scales of many angiosperm families (see Reference [28]) and allow phylogenetic data integration among different taxa, lineage-specific probe sets are preferable in some cases. As an example, lineage-specific probe sets may be more effective in accessing relationships over multiple phylogenetic scales within rapidly radiating clades [29–31]. Furthermore, given the widespread occurrence of genomic duplications in angiosperms, lineage-specific kits are less prone to sample paralogous genes (but see Reference [32]), and this may affect the accuracy of phylogenetic inference.

Lineage-specific probe sets are generally designed to target conserved exon regions (see Reference [27]), while noncoding regions (i.e., introns and intergenic spacers) are generally sampled as off-target sequences (i.e., DNA sequences flanking the target exonic regions obtained by the Hyb-Seq approach [33]). Obtaining noncoding nuclear regions enables the accession of selectively unconstrained genomic variation that, allied with slow-evolving coding regions, may provide accurate phylogenetic inferences at multiple taxonomic scales [34–36], including rapid radiations [37]. Although noncoding regions have posted a huge impact on resolving shallow-level phylogenies [38,39], few studies have included them on customized probe sets (see References [40,41]).

Here, we present a newly developed probe set targeting both the coding and noncoding sequences of 591 putative orthologous regions for the subfamily Cactoideae, named Cactaceae591. We hypothesized that this probe set would be useful to perform phylogenetic studies across the phylogeographic–phylogenetic continuum of Cactaceae, as it includes regions with different levels of variation. We tested the informativeness of this panel sequencing 83 accessions for two datasets sampling: (1) at the family level on Cactaceae, including subfamilies Pereskioideae, Opuntioideae, and Cactoideae, with broad sampling of the tribe Cereeae; and (2) at the species level on Cactaceae, considering interspecific relationships and intraspecific variation within a group of closely related species of the genus

Cereus Mill. [17]. We compared the performance of concatenated and coalescent-based phylogenetic reconstructions for both datasets. We highlighted the potential of Cactaeceae591 to produce well-resolved phylogenetic inferences at different taxonomic levels of the family Cactaceae. Furthermore, by overlapping 13 loci with Angiosperms353, the data from both probe sets may be partially aligned and integrated for phylogenetic purposes.

2. Materials and Methods

2.1. Target Loci Selection, Probe Design, and Library Preparation

We used Orthofinder v2.3.14 [42] to search for putative orthologous nuclear regions across 14 de novo transcriptomes and four genome assemblies available for Cactaceae in public repositories (References [14,16,43]; Supplementary Table S1). We constrained our search for orthologous nuclear regions shared among at least seven transcriptomes (at least one from the tribe Cereeae) and at least one of the four genome datasets selected in this study. We annotated the identified orthologous nuclear regions by using *Arabidopsis thaliana* (L.) Heynh. transcripts [44] and Caryophyllales representatives available on the PLAZA 5.0 dicots database [45], including an annotated cactus genome (*Selenicereus undatus* (Haw.), D.R.Hunt, designated as *Hylocereus undatus* by the authors [46]). The annotation was performed by using BLASTn with a minimum of 100 bp overlap and an e-value lower than 1×10^{-15} for *A. thaliana* transcripts. The annotation of Caryophyllales representatives considered metrics of a minimum identity of 85% and an e-value lower than 1×10^{-80}, according to the genetic proximity across our samples and the Caryophyllales sequence taxa. We also obtained information about gene ontology from the PLAZA 5.0 dicots database (Supplementary Table S2). We removed cytoplasmic and contaminant (e.g., bacteria) sequences. After selecting the final putative orthologous nuclear regions on our dataset, we surveyed for shared target nuclear regions to those in the expanded version (mega353 target file [47]) of the Angiosperms353 kit [26], using BLASTn. We considered a match between our panel and the Angiosperm353 regions with an e-value lower than 1×10^{-10}, with a minimum of 50 bp overlap, and those that matched only one hit with the Angiosperm353 loci. Additionally, to enable Cactaceae591 datasets to be integrated with an array of previous Cactaceae phylogenetics studies (e.g., Reference [8]), we also targeted two plastid regions widely used in phylogenetic studies of Cactaceae in our probes: the protein-coding gene *rbcL* and the intergenic spacer *trnK-matK*.

The final dataset was composed of 591 orthologous target sequences, including coding and noncoding nuclear regions and two plastid regions. The noncoding regions are composed of the intronic, intergenic spacer, and anonymous (non-annotated) nuclear regions. The target regions identified in our survey ranged from 280 to 3590 bp. A set of 28,570 probes of 120 bp was designed by RAPiD Genomics LLC (Appendix A) with 3× tilling density. Approximately 58.1% of the probes were designed on exons, 32.27% on intronic and intergenic regions, and 9.63% on anonymous regions. A lower probe density was designed for plastid *rbcL* and *trnK-matK* regions to avoid overrepresentation of these regions in the dataset.

2.2. Sampling, DNA Extraction, and Library Preparation

We designed our sampling to evaluate the performance of Cactaceae591 at the family and species levels (for more details, see Supplementary Table S3). The family-level dataset (hereafter, the Cactaceae dataset) was generated by sampling 57 species from the subfamilies Pereskioideae, Opuntioideae, and Cactoideae, with an emphasis on the tribe Cereeae (sensu [48]). This sampling covers divergence events spanning from the Oligocene (~26.9 MyA) to Pleistocene (~2.5 MyA, [8,49–51]). A species-level dataset (hereafter, the Cereus dataset) was generated by sampling 27 representatives of eight species and two subspecies that compose a monophyletic group (clade A) within the genus *Cereus*, as well as three species from sister clades (*Cereus spegazzini*, *Cereus mirabella*, and *Cereus kroenleinii*), used as outgroups (see Reference [17]; Supplementary Table S3). Clade A is composed of closely related species distributed in the Dry Diagonal of South America (Cerrado, Caatinga,

and Chaco biomes) that diverged from early to late Pleistocene (2.1 Mya to 80 kya, [51]). We sampled a single individual from multiple populations for most species of this clade (see Supplementary Table S3 for more details).

DNA extraction was performed by using 80 to 100 mg of root or epidermis grounded with steel beads on TissueLyser II (Qiagen, Hilden, Alemanha). We performed DNA extraction from fresh and −80 °C preserved tissues and from herbarium material (see Supplementary Table S3 for more details). Genomic DNA was extracted by using a high-salt CTAB protocol slightly modified from Inglis et al. [52]. DNA integrity, quality, and quantity were checked by using 1% agarose gel Nanodrop absorbance ratios and Qubit dsDNA High Sensitivity (Thermo Scientific, Waltham, MA, USA), respectively.

Library preparation was performed by RAPiD Genomics for Illumina sequencing, utilizing their high-throughput workflow with proprietary chemistry. Briefly, DNA was sheared to a mean fragment length of 400 bp, and fragments were end-repaired, followed by incorporation of unique dual-indexed Illumina adaptors and PCR enrichment. Sequence capture was performed by using RAPiD Genomics proprietary chemistry and workflows. Briefly, fully constructed libraries were hybridized to 120 bp probes, and probe/DNA hybrids were captured on streptavidin beads, washed, and PCR amplified. Samples were pooled equimolar and sequenced on a NovaSeq 6000 (2 × 150 bp).

2.3. Sequencing Assembly, Alignment, and Genetic Variability of Target Loci

The raw reads were trimmed by using AdapterRemoval v2 [53], removing poor-quality reads (phred < 20) and reads smaller than 60 bp. We assembled and mapped the trimmed reads by using HybPiper pipeline v1.3 [54] across the 591 target regions. To enable efficient locus recovery, we constructed a variable reference sequence file by using sequences from multiple species across subfamily Cactoideae (Appendix B). The trimmed reads were mapped to reference sequences, using BWA [55], and then assembled into contigs with a minimum coverage of 8×, using SPADES [56]. Summary statistics of mapped reads and recovered target regions were obtained by using hybpiper_stats.py.

We detected putative paralogous regions in our dataset by using the command paralog_investigator.py on the HybPiper pipeline. Putative paralog regions were checked on gene trees inferred with IQ-TREE 1.6 [57] and by using the contigs from each sample obtained, using the paralog_retriever.py script on HybPiper. Putative paralogs shared among more than 10% of samples were removed from the final dataset. We also excluded regions with more than 70% missing data on the family-level sample.

The final dataset was automatically aligned by using MAFFT v7 [58]. Indel-rich regions were trimmed by using trimAL v1.3 [59] with the function –gt 0.7. The alignments were concatenated by using AMAS v0.94 [60]. Summary statistics (alignment length, number, and variable sites and parsimony-informative sites) were calculated by using AMAS v0.94 [60] for our two main datasets (Cactaceae and Cereus datasets).

We also explored nuclear regions shared between the Cactaceae591 and Angiosperm353 probe sets, following a similar strategy described by Siniscalchi et al. [32]. Our aim here was to test whether representatives from the same taxon and sequenced by different probe sets could be recovered as monophyletic groups. We selected representatives of major clades of Cactaceae sequenced with our probe set (obtained in this study) and with the Angiosperm353 probe set obtained by Baker et al. [25] (see Supplementary Table S4 for more details about this sampling). We assembled this dataset by using a reference sequence file of the expanded version (mega353 target file [47]) of the Angiosperms353 kit [26], implementing the pipeline of Hybpiper, as described above. The loci recovered in more than 50% of the selected samples were aligned by using MAFFT v7 [58], and indel-rich regions were trimmed by using trimAL v1.3 [59] with the function –gt 0.5. The final concatenated alignment obtained with AMAS v0.94 was used as input to infer a maximum likelihood tree on IQ-TREE 1.6 [57] with 10,000 ultrafast bootstrap (BS) replicates [61].

2.4. Phylogenetic Inference

To investigate the efficiency of Cactaceae591 coding and noncoding sequences separately for phylogenetic reconstructions in Cactaceae, we explored different data matrices within each dataset (Supplementary Table S3). For the Cactaceae dataset, we explored the phylogenetic signal, focusing only on coding regions (exonic sequences; Cactaceae-coding sub-dataset, herein) and all target nuclear regions (exons, introns, intergenic spacers, and anonymous nuclear regions; Cactaceae-all sub-dataset, herein). For the Cereus dataset, we explored a matrix including only noncoding nuclear regions (introns, intergenic spacers, and anonymous nuclear regions; Cereus-noncoding sub-dataset) and a matrix with all target nuclear regions (Cereus-all sub-dataset). Our main focus was to evaluate the phylogenetic signal of orthologous nuclear regions under different levels of selective constraints. We did not include the plastidial regions in our phylogenetic analysis, as the value of these regions for phylogenetic purposes was already tested in previous studies.

We performed phylogenetic inference by using both concatenated and coalescent approaches. Maximum likelihood (ML) trees were generated by using a concatenated matrix in IQ-TREE 1.6 [57] with 10,000 ultrafast bootstrap (BS) replicates [61], applying the -bb option. We assumed the best-fit model for each partition by using ModelFinder [62], which is available on IQ-TREE. The ML trees of Cactaceae were rooted by using *Pereskia bahiensis* Gürke *and Leunbergeria aureiflora* (F. Ritter) Lodé, while *Cereus* clade A gene trees were rooted by using *C. mirabella* N.P. Taylor and *C. kroenleinii* N.P. Taylor. Branch support of ML trees (BS) was classified into three categories: high (>95), moderate (95–70), and low (<70).

The coalescent species trees were inferred by a summary approach on ASTRAL-III v5.6 [63], using ML gene trees generated in IQ-TREE v1.6 [57], with 10,000 ultrafast bootstrap replicates [61]. Nodes in gene trees with BS <50 were collapsed. The branch supports of species trees were accessed by local posterior probabilities (LPPs), which were computed based on gene tree quartet frequencies [64]. Branch support of coalescent-based trees (LPP) was classified as high support (>0.95), moderate support (0.95–0.70), and low support (<0.70).

3. Results

3.1. Sequencing and Datasets

Approximately 96 million reads were obtained for all samples after the trimming process, averaging 1.1 million reads per sample. The mapped reads corresponded to 72.5 $\pm$ 4.2% (mean $\pm$ SD) of the total trimmed reads (Supplementary Table S5). The total target regions recovered from each sample ranged from 417 to 566 (Supplementary Table S5). The two plastid regions included in our orthologous probe set were recovered for all samples. Most of the target regions (547) were captured for more than 70% of our sampling, with slightly more regions recovered on the subfamily Cactoideae (550) compared to the subfamilies Pereskioideae (517) and Opuntioideae (429) (Supplementary Figure S1). We identified an average of 12 putative paralogs per sample (Supplementary Table S4), and 45 paralogous regions were shared with approximately 10% of our sampling. Considering the target loci identified as paralogs and loci that failed for most samples, we excluded 74 target regions from our datasets. We identified 13 target nuclear regions shared between the Cactaceae591 and Angiosperm353 probe sets according to the BLAST results (Supplementary Table S2), but only ten were shared by more than one representative sequenced with the Cactaceae591 probe set (Supplementary Table S5). The sequences overlapping in these loci varied from 75 to 666 base pairs. The phylogenetic inference recovered the monophyly of representatives of major clades of Cactaceae from different sequencing strategies (Cactaceae591 vs. Angiosperm353; Supplementary Figure S2), revealing the integration of ten target nuclear loci sequenced by the Cactaceae591 and Angiosperm353 panels.

The final matrix consisted of 512 orthologous regions containing 292 exonic, 328 intronic, and 67 anonymous sequences. These sequences resulted in a matrix of 748,848 bp with 194,892 parsimony informative sites (PIS) for the Cactaceae dataset and 864,313 bp

with 23,602 PIS for the Cereus dataset (Table 1). Anonymous regions showed a higher percentage of missing data but also presented a higher number of PIS than exonic regions (Table 1).

Table 1. Summary statistics of nucleotide variation calculated for coding, noncoding, and anonymous regions in the Cactaceae and Cereus datasets. Note: bp, base pairs; S, variable sites; PIS (%), parsimony informative sites (proportion of PIS).

Dataset	Alignment Length (pb)	% Missing Data	S	PIS (%)
Cactaceae				
Exons	258,375	12.83	100,715	54,156 (0.211)
Introns	309,041	23.28	149,244	78,597 (0.254)
anonymous regions	154,174	64.45	44,159	22,416 (0.145)
Cactaceae-coding	258,375	12.83	100,715	54,156 (0.211)
Cactaceae-all	721,59	33.52	294,118	155,169 (0.215)
Cereus Clade A				
Exons	303,688	12.67	24,729	10,278 (0.034)
Introns	424,012	5.49	42,732	18,454 (0.044)
anonymous regions	136,613	25.02	11,036	5148 (0.038)
Cereus-noncoding	560,025	15.26	53,768	23,603 (0.042)
Cereus-all	864,313	14.39	78,497	33,880 (0.039)

3.2. Phylogenetic Reconstructions

Phylogenetic inferences recovered a family-level tree congruent with the current Cactaceae systematics (Figure 1). The subfamilies and the tribes sampled were recovered as monophyletic groups, and the phylogenetic relationships were generally congruent with those inferred in previous studies [5,8,27,49]. Furthermore, most of the Cactaceae deep nodes were recovered with high support, regardless of the sub-dataset and inference method used (Figure 1 and Supplementary Figure S3). A topological incongruence was observed between phylogenetic reconstruction approaches. The coalescent-based inference recovered *L. aureiflora* as the sister of all remaining Cactaceae, followed by *P. bahiensis* (Figure 1), while the ML inference grouped both species in the same branch (Supplementary Figure S3). An additional major incongruence was observed among sub-datasets involving the position of Rhipsalideae in relation to core Cactoideae I and II. Using the Cactaceae-coding sub-dataset, the representatives of this tribe were clustered in a polytomic node with the highly supported clades of core Cactoideae I and II (Figure 1 and Supplementary Figure S3). However, by using the Cactaceae-all sub-dataset, we resolved these relationships with high support, setting the tribe Rhipsalideae as the sister of a clade clustering core, Cactoideae I and II (Figure 1 and Supplementary Figure S3).

Some shallow-level incongruences were observed between Cactaceae sub-datasets involving the relationships among the *Praecereus*, *Cipocereus*, and *Cereus* genera (Figure 1 and Supplementary Figure S3). For example, *Cipocereus* was highly supported within the clade of *Cereus* species in all inferences. A further example of shallow-level incongruence involved *Praecereus*, which was allocated as a sister clade of *Cereus* species, depending on the dataset and phylogenetic approach used (Figure 1 and Supplementary Figure S3).

The phylogenetic backbone for *Cereus* clade A recovered the same relationships estimated for this clade by using RAD-Seq data with high branch support (Figure 2; see Reference [17]). Furthermore, all conspecific samples nested together, with the exception of the representatives of *C. hexagonus* and *C. jamacaru* (Figure 2). The hypothesis of a peripatric origin of *C. insularis* from coastal populations of *C. fernambucensis* subsp. *fernambucensis* in Northeastern Brazil [51,65] was corroborated with all sub-datasets and phylogenetics methods. Only the ML tree inferred with the Cereus-all sub-dataset provided a slightly different relationship among these taxa, nesting *C. fernambucensis* subsp. *sericifer* within the clade of *C. fernambucensis* subsp. *fernambucensis*.

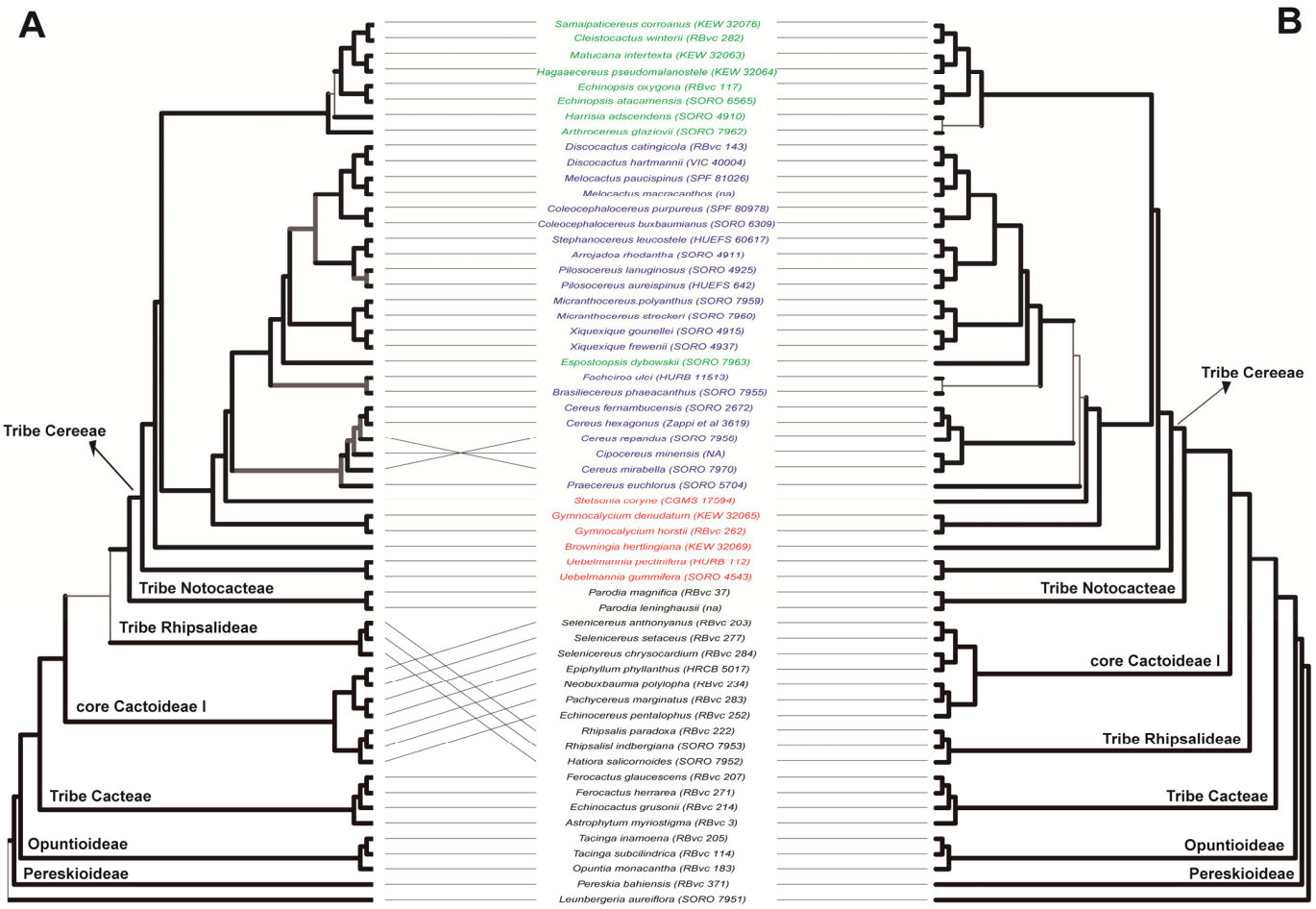


Figure 1. Comparison of coalescent-based species trees of Cactaceae estimated with Cactaceae-coding (**A**) and Cactaceae-all (**B**) sub-datasets. Cactaceae-coding and Cactaceae-all sub-datasets included 290 and 512 nuclear regions, respectively. The branches with high support (LPP > 0.95) are colored with thick dark lines, the branches with moderate support (LPP 0.95–0.70) are colored with thick gray lines, and the branches with low support (LPP < 0.70) are shown as thin light lines. Voucher number of each accession is provided in parenthesis after species names. Species from the subtribes Rebutinae, Trichocereinae, and Cereinae are colored red, green, and blue, respectively (sensu Nyffeler and Eggli [48]). LPP: local posterior probability.

We observed that, in general, the use of sub-datasets, including all target regions (both coding and noncoding), resulted in an increase in branch support to resolve relationships within both Cactaceae and Cereus clade A groups (Table 2).

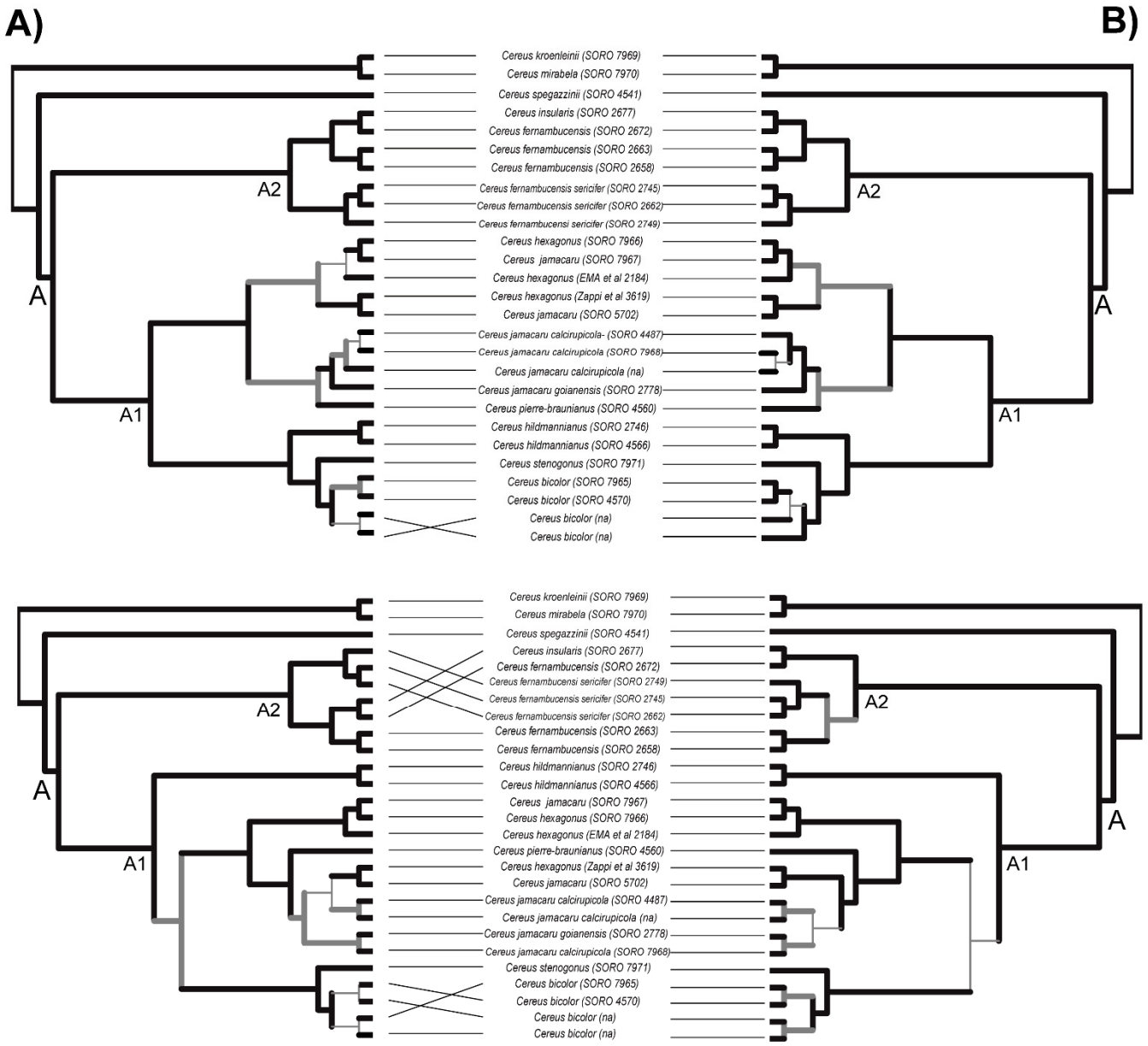


Figure 2. Comparison of phylogenies of Cereus clade A inferred with Cereus-noncoding (**A**) and Cereus-all (**B**) sub-datasets, using coalescent-based (upper) and ML concatenated (down) approaches. The branches with high support (LPP > 0.95 and BS > 95) are highlighted with thick branch lines, the branches with moderate support (LPP 0.95–0.70 and BS 95–70) are colored with thick gray lines, and the branches with low support (LPP < 0.70 and BS < 70) are shown as thin light lines. LPP, local posterior probability; BS, ultrafast bootstrap. Voucher number of each accession is provided in parenthesis after species names.

Table 2. Node support in the phylogenetic trees inferred from the Cactaceae and Cereus datasets, using different sub-datasets (coding, noncoding, and all sequences) and phylogenetic approaches. Branch support values of coalescent (LPP) and concatenated (BS) trees were classified as high (>0.95, >95), moderate (0.95–0.70, 95–70), or low (<0.70, <70), respectively.

Dataset	Total	High Support	Moderate Support	Low Support
Cactaceae-coding				
concatenated	57	51	5	1
coalescent-based	57	48	7	2
Cactaceae-all				
concatenated	57	52	3	2
coalescent-based	57	53	0	4
Cereus-noncoding				
concatenated	26	19	4	3
coalescent-based	26	19	4	3
Cereus-all				
concatenated	26	19	5	2
coalescent-based	26	21	3	2

4. Discussion

4.1. Cactaceae591: A Lineage-Specific Probe Set to Perform Evolutionary Studies in Cactus

Here, we describe Cactaceae591, a new lineage-specific probe set targeting both coding and noncoding orthologous regions in Cactaceae. Based on the large number of polymorphic sites recovered, the presence of gene regions with different levels of nucleotide variation, and the robustness of our phylogenetic inferences at both the family and species levels, we highlight the relevance of this probe set at multiple taxonomic levels within Cactaceae.

Previous evolutionary studies on recent and rapid divergent groups of angiosperms reported a robust and accurate phylogenetic reconstruction when adding information from noncoding nuclear regions [13,37–40]. Nevertheless, few target sequencing kits incorporated probes designed to capture intronic and intergenic regions (see References [40,41]). By integrating coding and noncoding sequences that are expected to exhibit different evolutionary rates, Cactaceae591 approaches the ideal characteristics of molecular markers for resolving relationships in taxa that have undergone recent radiation [66]. The increase of branch support when using both coding and noncoding data highlights the value of combining slow- and fast-evolving sequences in our phylogenetic inferences, especially at the species level. We acknowledge that our intraspecific sampling was limited to one sample per population, preventing us from testing the capability of Cactaceae591 to recover variation within populations. However, taking into account the intraspecific variation recovered in our study, we believe that the noncoding regions available in Cactaceae591 are also promising for phylogeographic and population studies on the family.

Overall, the trees inferred from the Cactaceae dataset were congruent with previous phylogenetic studies, regardless of the phylogenetic approaches (concatenation or coalescent-based reconstructions) or sub-datasets used (coding and noncoding regions). The major discordances occurred on terminal nodes and/or on short deep internodes, as is commonly seen on cactus lineages, due to the rapid and recent divergence experienced by many clades in this plant family [1,17]. It is important to highlight the importance of the implementation of coalescent-based phylogenetic methods in cacti, as there is extensive incomplete lineage sorting generating gene tree heterogeneity in this family [14], a scenario in which the concatenation approach can lead to strongly supported misleading relationships [11].

The results generated with Cactaceae591 have the potential to be integrated with phylogenomic data that are available from other studies, as Cactaceae591 shares thirteen

orthologous regions with the Angiosperm353 kit, as well as two plastid markers commonly used in phylogenetic and phylogeographic studies in cacti [2]. Recent studies have successfully used this universal probe set on Cactaceae [25,27]. However, by targeting both orthologous coding and noncoding nuclear regions, Cactaceae591 has greater potential and flexibility to infer phylogenetic relationships at both the shallow and deep levels of the evolution of Cactaceae. Hence, the combination of lineage-specific probe sets with universal panels may benefit both the cacti- and plant-specialist communities [32,67,68].

4.2. Notes on the Phylogenetic Relationships

Although our main focus was to describe Cactaceae591, some of our phylogenetic results are worth noting. Overall, the backbone phylogeny was consistent across different methods and sub-datasets, recovering the monophyly of tribes and genera of the subfamily Cactoideae. The sister relationship of *L. aureiflora* and the remaining Cactaceae taxa, followed by *P. bahiensis*, agrees with the paraphyly hypothesis for *Pereskia* s.l. [15,69] and with those inferred by using the Cactaceae120 probe set [27]. Phylogenetic inferences based on plastid DNA [4,5,49] have grouped the tribe Rhipsalideae as a member of Core Cactoideae II. Our results using all target nuclear regions resolved this tribe with high support as sister of the clade clustering core Cactoideae I and II, while using only coding nuclear regions; Rhipsalideae assumes a polytomy with these clades. Interestingly, an inconsistent relationship of Rhipsalideae with core Cactoideae was also observed by Acha and Majure [27], using the Cactaceae120 dataset. Future studies using a wider sampling of the representatives of the subfamily Cactoideae are likely required to elucidate these contentious relationships with high accuracy.

Our main sampling focused on the tribe Cereeae and its subtribes. Beyond the aware paraphyletic circumscription of subtribe Rebutinae [48], the monophyly of subtribes Cereinae and Trichocereinae is still under debate (see Reference [70]). Considering the circumscription proposed by Nyffeler and Eggli [48] for the tribe Cereeae, our results recovered the subtribe Trichocereinae as polyphyletic and the subtribes Cereinae and Rebutinae as paraphyletic groups. For instance, *Espostoopsis dybowskii* was recovered within the Cereinae clade, although this taxon is traditionally considered a member of the subtribe Trichocereinae. In the phylogenies of Ritz et al. [3], Schlumpberger and Renner [71], Lendel et al. [72], and Bombonato et al. [17], it is also possible to observe the members of these subtribes as nonmonophyletic taxa. Finally, our results found the same contentious relationships involving the genera *Cereus*, *Cipocereus*, and *Praecereus* inferred with plastid [50] and RAD-Seq data [17].

The target regions selected here were useful for distinguishing relationships within closely related species of the genus *Cereus*. The relationship within this clade consistently agreed with previous phylogenomic results [17] when applying the Cereus-all sub-dataset allied with coalescent-based inferences. The inability to completely separate the currently recognized taxonomic species of subclade *C. jamacaru* is a problem not only associated with rapid diversification of the group but also possibly with a misleading taxonomy of this genus (see Reference [73]). The potential to resolve inter- and intraspecific relationships demonstrated by this probe set may also be helpful for species delimitation purposes within this group.

5. Conclusions

The Cactaceae591 probe set showed to be a powerful strategy to resolve phylogenetic relationships at deep and shallow taxonomic levels, displaying topological congruences with previous phylogenetic studies. In addition, by targeting noncoding fast-evolving regions, Cactaceae591 potentially captures neutral variation that is useful for phylogeographic studies. The sharing of thirteen loci with the Angiosperms353 kit allows for the integration of the Cactaceae591 dataset across angiosperm phylogenetic studies.

Supplementary Materials: The supporting information can be downloaded at https://www.mdpi.com/article/10.3390/genes13040707/s1. Figure S1: Heatmap indicating 591 orthologs' recovery success per sample; scale color indicates success rate. Figure S2: Coalescent species trees estimated with ten loci shared by Cactaceae591 and Angiosperm353 probe sets. The BS values are displayed on the nodes. Species with suffix (591) were sequenced with a Cactaceae591 kit, and those with (353) were sequenced with an Angiosperm353 kit. The branches with high support (BS > 95) are colored with thick branch lines, the branches with moderate support (BS 95–70) are colored with thick gray lines, and the branches with low support (BS < 70 on ML inferences) are shown as thin light lines. Figure S3: Maximum likelihood phylogenies inferred from Cactaceae-coding (A) and Cactaceae-coding (B) datasets. The branches with high support (BS > 95) are colored with thick branch lines, the branches with moderate support (BS 95–70) are colored with thick gray lines, and the branches with low support (BS < 70 on ML inferences) are shown as thin light lines. Species colored red are from the subtribe Rebutinae, green are from the subtribe Trichocereinae, and blue are from the subtribe Cereinae (sensu Nyffeler and Eggli [48]). Table S1: Genomic data of representatives from subfamily Cactoideae used for identification of orthologous genes. Table S2: Details about the 591 orthologous regions selected in the present study, including probe number, annotation according to *Arabidopsis thaliana* and the PLAZA database, and paralog assignment for each gene. Table S3: Sampling information of Cactaceae species used in this study. RBvc: Cactário do Jardim Botânico do Rio de Janeiro; SORO: Herbário do Centro de Ciências e Tecnologias para a Sustentabilidade; HURB: Herbário do Recôncavo da Bahia; HUEFS: Herbario da Universidade Estadual de Feira de Santana; KEW: Kew DNA bank; SPF: Herbário da Universidade de São Paulo; VIC: Herbário da Universidade Federal de Viçosa. CGMS: Herbário da Fundação Universidade Federal de Mato Grosso do Sul; HRCB: Herbário Rioclarense; na: not available. Table S4: Sampling information used in analysis of shared nuclear regions with Cactaceae591 and Angiosperm353 probe set. Table S5: Summary statistics of locus recovery analysis performed on HybPiper for the sampling sequenced and mapped with Cactaceae591 probe set (see Table S3). Table S6: Length recovery of locus shared by both Cactaceae591 and Angiosperm353 probe sets obtained by HybPiper. The numbers of each column represent the locus name of the Angiosperm353 panel. We selected representatives from major clades of Cactaceae sequenced with the Cactaceae591 probe set (this study) and with the Angiosperm353 probe set (Baker et al. 2021), including representatives from closely related families of Cactaceae.

Author Contributions: Conceptualization, E.M.M., M.R.-B. and F.F.F.; data acquisition: M.R.-B. and M.C.T.; formal analysis, M.R.-B., D.T.A. and M.C.T.; funding acquisition, E.M.M.; writing—original draft, M.R.-B. and E.M.M.; writing—review and editing, M.R.-B., F.F.F., D.T.A., M.C.T. and E.M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP: 2018/03428-5 and 2019/03211-9). E.M.M. was granted a CNPq/MCTIC research productivity fellowship (303940/2019-0), M.R.-B. and M.C.T. were granted FAPESP fellowships (2018/06937-8 and 2019/11233-2), and D.T.A. was granted a fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES; Finance Code 001).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the Supplementary Materials. Accession numbers for all sequence data used to design the sequence capture array are presented in Table S1. The raw reads of target sequence capture data are available on the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA), under the BioProject accession number PRJNA812417.

Acknowledgments: We thank Gerardus Olsthoorn, Daniela Zappi, Nigel Taylor, Lidyanne Aona, Diego Gonzaga, and the Coleção de Cactos e Suculentas do Instituto de Pesquisas Jardim Botânico do Rio de Janeiro for contributing to the cactus samples used in this study.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Probe set sequences used to capture 591 target orthologous regions of Cactaceae to construct the target enrichment library.

Appendix B

Reference sequences for the 591 target orthologous regions identified from genomic and transcriptome resources detailed in Supplementary Table S1.

References

1. Guerrero, P.; Majure, L.C.; Cornejo-Romero, A.; Hernández-Hernández, T. Phylogenetic Relationships and Evolutionary Trends in the Cactus Family. *J. Hered.* **2018**, *110*, 4–21. [CrossRef] [PubMed]
2. Franco, F.F.; Amaral, D.T.; Bonatelli, I.A.S.; Romeiro-Brito, M.; Telhe, M.C.; Moraes, E.M. Evolutionary Genetics of Cacti: Research Biases, Advances and Prospects. *Genes* **2022**, *13*, 452. [CrossRef] [PubMed]
3. Ritz, C.M.; Martins, L.; Mecklenburg, R.; Goremykin, V.; Hellwig, F.H. The molecular phylogeny of Rebutia (Cactaceae) and its allies demonstrates the influence of paleogeography on the evolution of South American mountain cacti. *Am. J. Bot.* **2007**, *94*, 1321–1332. [CrossRef] [PubMed]
4. Bárcenas, R.T.; Yesson, C.; Hawkins, J.A. Molecular systematics of the Cactaceae. *Cladistics* **2011**, *27*, 470–489. [CrossRef]
5. Hernández-Hernández, T.; Hernández, H.M.; De-Nova, J.A.; Puente, R.; Eguiarte, L.E.; Magallón, S. Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales; Eudicotyledoneae). *Am. J. Bot.* **2011**, *98*, 44–61. [CrossRef]
6. Vázquez-Sánchez, M.; Terrazas, T.; Arias, S.; Ochoterena, H. Molecular phylogeny, origin and taxonomic implications of the tribe Cacteae (Cactaceae). *Syst. Biodivers.* **2013**, *11*, 103–116. [CrossRef]
7. Calvente, A.; Moraes, E.M.; Lavor, P.; Bonatelli, I.A.S.; Nacaguma, P.; Versieux, L.M.; Taylor, N.P.; Zappi, D.C. Phylogenetic analyses of Pilosocereus (Cactaceae) inferred from plastid and nuclear sequences. *Bot. J. Linn. Soc.* **2017**, *183*, 25–38. [CrossRef]
8. Hernández-Hernández, T.; Brown, J.; Schlumpberger, B.O.; Eguiarte, L.E.; Magallón, S. Beyond aridification: Multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New Phytol.* **2014**, *202*, 1382–1397. [CrossRef]
9. Maddison, W.P. Gene trees in species trees. *Syst. Biol.* **1997**, *46*, 523–536. [CrossRef]
10. Seehausen, O.; Butlin, R.; Keller, I.; Wagner, C.E.; Boughman, J.W.; Hohenlohe, P.A.; Peichel, C.L.; Saetre, G.; Bank, C.; Brännström, A.; et al. Genomics and the origin of species. *Nat. Rev. Genet.* **2014**, *15*, 176–192. [CrossRef]
11. Edwards, S.V.; Xi, Z.; Janke, A.; Faircloth, B.C.; McCormack, J.E.; Glenn, T.C.; Zhong, B.; Wu, S.; Lemmon, E.M.; Lemmon, A.R.; et al. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Mol. Phylogenet. Evol.* **2016**, *94*, 447–462. [CrossRef] [PubMed]
12. McKain, M.R.; Johnson, M.G.; Uribe-Convers, S.; Eaton, D.; Yang, Y. Practical considerations for plant phylogenomics. *Appl. Plant Sci.* **2018**, *6*, e1038. [CrossRef] [PubMed]
13. Sanderson, M.J.; Copetti, D.; Búrquez, A.; Bustamante, E.; Charboneau, J.L.; Eguiarte, L.E.; Kumar, S.; Lee, H.O.; Lee, J.; McMahon, M.; et al. Exceptional reduction of the plastid genome of saguaro cactus (Carnegiea gigantea): Loss of the ndh gene suite and inverted repeat. *Am. J. Bot.* **2015**, *102*, 1115–1127. [CrossRef] [PubMed]
14. Copetti, D.; Búrquez, A.; Bustamante, E.; Charboneau, J.L.M.; Childs, K.; Eguiarte, L.E.; Lee, S.; Liu, T.L.; McMahon, M.M.; Whiteman, N.; et al. Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 12003–12008. [CrossRef]
15. Walker, J.F.; Yang, Y.; Feng, T.; Timoneda, A.; Mikenas, J.; Hutchison, V.; Edwards, C.; Wang, N.; Ahluwalia, S.; Olivieri, J.; et al. From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. *Am. J. Bot.* **2018**, *105*, 446–462. [CrossRef]
16. Wang, N.; Yang, Y.; Moore, M.J.; Brockington, S.F.; Walker, J.F.; Brown, J.W.; Liang, B.; Feng, T.; Edwards, C.; Mikenas, J.; et al. Evolution of Portulacineae Marked by Gene Tree Conflict and Gene Family Expansion Associated with Adaptation to Harsh Environments. *Mol. Biol. Evol.* **2018**, *36*, 112–126. [CrossRef]
17. Bombonato, J.R.; Amaral, D.T.; Silva, G.A.R.; Khan, G.; Moraes, E.M.; Andrade, S.C.; Eaton, D.A.R.; Alonso, D.P.; Ribolla, P.E.M.; Taylor, N.; et al. The potential of genome-wide RAD sequences for resolving rapid radiations: A case study in Cactaceae. *Mol. Phylogenetics Evol.* **2020**, *151*, 106896. [CrossRef]
18. Merklinger, F.F.; Böhnert, T.; Arakaki, M.; Weigend, M.; Quandt, D.; Luebert, F. Quaternary diversification of a columnar cactus in the driest place on earth. *Am. J. Bot.* **2021**, *108*, 184–199. [CrossRef]
19. Köhler, M.; Reginato, M.; Souza-Chies, T.T.; Majure, L.C. Insights into Chloroplast Genome Evolution Across Opuntioideae (Cactaceae) Reveals Robust Yet Sometimes Conflicting Phylogenetic Topologies. *Front. Plant Sci.* **2020**, *11*, 729. [CrossRef]
20. Majure, L.C.; Baker, M.A.; Cloud-Hughes, M.; Salywon, A.; Neubig, K.M. Phylogenomics in Cactaceae: A case study using the chollas sensu lato (Cylindropuntieae, Opuntioideae) reveals a common pattern out of the Chihuahuan and Sonoran deserts. *Am. J. Bot.* **2019**, *106*, 1327–1345. [CrossRef]
21. Majure, L.C.; Barrios, D.; Díaz, E.; Zumwalde, B.A.; Testo, W.; Negrón-Ortíz, V. Pleistocene aridification underlies the evolutionary history of the Caribbean endemic, insular, giant Consolea (Opuntioideae). *Am. J. Bot.* **2021**, *108*, 200–215. [CrossRef] [PubMed]

22. Breslin, P.B.; Wojciechowski, M.F.; Majure, L.C. Molecular phylogeny of the Mammilloid clade (Cactaceae) resolves the monophyly of Mammillaria. *Taxon* **2021**, *70*, 308–323. [CrossRef]
23. Mamanova, L.; Coffey, A.J.; Scott, C.E.; Kozarewa, I.; Turner, E.H.; Kumar, A.; Howard, E.; Shendure, J.; Turner, D.J. Target-enrichment strategies for next-generation sequencing. *Nat. Methods* **2010**, *7*, 111–118. [CrossRef] [PubMed]
24. McDonnell, A.J.; Baker, W.J.; Dodsworth, S.; Forest, F.; Graham, S.W.; Johnson, M.G.; Pokorny, L.; Tate, J.; Wicke, S.; Wickett, N.J. Exploring Angiosperms353: Developing and applying a universal toolkit for flowering plant phylogenomics. *Appl. Plant Sci.* **2021**, *9*, e11443. [CrossRef] [PubMed]
25. Baker, W.J.; Bailey, P.; Barber, V.; Barker, A.; Bellot, S.; Bishop, D.; Botigué, L.R.; Brewer, G.; Carruthers, T.; Clarkson, J.J.; et al. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Syst. Biol.* **2022**, *71*, 301–319. [CrossRef]
26. Johnson, M.G.; Pokorny, L.; Dodsworth, S.; Botigue, L.R.; Cowan, R.S.; Devault, A.; Eiserhardt, W.L.; Epitawalage, N.; Forest, F.; Kim, J.T.; et al. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Syst. Biol.* **2019**, *68*, 594–606. [CrossRef]
27. Acha, S.; Majure, L.C. A New Approach Using Targeted Sequence Capture for Phylogenomic Studies across Cactaceae. *Genes* **2022**, *13*, 350. [CrossRef]
28. Shee, Z.Q.; Frodin, D.G.; Cámara-Leret, R.; Pokorny, L. Reconstructing the complex evolutionary history of the Papuasian Schefflera radiation through herbariomics. *Front. Plant Sci.* **2020**, *11*, 258. [CrossRef]
29. Chau, J.H.; Rahfeldt, W.A.; Olmstead, R.G. Comparison of taxon-specific versus general locus sets for targeted sequence capture in plant phylogenomics. *Appl. Plant Sci.* **2018**, *6*, e1032. [CrossRef]
30. Jantzen, J.R.; Amarasinghe, P.; Folk, R.A.; Reginato, M.; Michelangeli, F.A.; Soltis, D.E.; Cellinese, N.; Soltis, P.S. A two-tier bioinformatic pipeline to develop probes for target capture of nuclear loci with applications in Melastomataceae. *Appl. Plant Sci.* **2020**, *8*, e11345. [CrossRef]
31. Yardeni, G.; Viruel, J.; Paris, M.; Hess, J.; Crego, C.G.; de La Harpe, M.; Rivera, N.; Barfuss, M.H.J.; Till, W.; Guzmán-Jacob, V.; et al. Taxon-specific or universal? Using target capture to study the evolutionary history of rapid radiations. *Mol. Ecol. Resour.* **2022**, *22*, 927–945. [CrossRef] [PubMed]
32. Siniscalchi, C.M.; Hidalgo, O.; Palazzesi, L.; Pellicer, J.; Pokorny, L.; Maurin, O.; Leitch, I.J.; Forest, F.; Baker, W.J.; Mandel, J.R. Lineage-specific vs. universal: A comparison of the Compositae1061 and Angiosperms353 enrichment panels in the sunflower family. *Appl. Plant Sci.* **2022**, *9*, e11422. [CrossRef] [PubMed]
33. Weitemier, K.; Straub, S.C.; Cronn, R.C.; Fishbein, M.; Schmickl, R.; McDonnell, A.; Liston, A. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Appl. Plant Sci.* **2014**, *2*, 1400042. [CrossRef] [PubMed]
34. Villaverde, T.; Pokorny, L.; Olsson, S.; Rincón-Barrado, M.; Johnson, M.G.; Gardner, E.M.; Wickett, N.J.; Molero, J.; Riina, R.; Sanmartín, I. Bridging the micro-and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytol.* **2018**, *220*, 636–650. [CrossRef]
35. Sanderson, B.J.; DiFazio, S.P.; Cronk, Q.C.; Ma, T.; Olson, M.S. A targeted sequence capture array for phylogenetics and population genomics in the Salicaceae. *Appl. Plant Sci.* **2020**, *8*, e11394. [CrossRef]
36. Peakall, R.; Wong, D.C.; Phillips, R.D.; Ruibal, M.; Eyles, R.; Rodriguez-Delgado, C.; Linde, C.C. A multitiered sequence capture strategy spanning broad evolutionary scales: Application for phylogenetic and phylogeographic studies of orchids. *Mol. Ecol. Resour.* **2021**, *21*, 1118–1140. [CrossRef]
37. Larridon, I.; Villaverde, T.; Zuntini, A.R.; Pokorny, L.; Brewer, G.E.; Epitawalage, N.; Fairlie, I.; Hahn, M.; Kim, J.; Maguilla, E.; et al. Tackling rapid radiations with targeted sequencing. *Front. Plant Sci.* **2020**, *10*, 1655. [CrossRef]
38. Bagley, J.C.; Uribe-Convers, S.; Carlsen, M.M.; Muchhala, N. Utility of targeted sequence capture for phylogenomics in rapid, recent angiosperm radiations: Neotropical Burmeistera bellflowers as a case study. *Mol. Phylogenet. Evol.* **2020**, *152*, 106769. [CrossRef]
39. Thomas, A.E.; Igea, J.; Meudt, H.M.; Albach, D.C.; Lee, W.G.; Tanentzap, A.J. Using target sequence capture to improve the phylogenetic resolution of a rapid radiation in New Zealand Veronica. *Am. J. Bot.* **2021**, *108*, 1289–1306. [CrossRef]
40. de La Harpe, M.; Hess, J.; Loiseau, O.; Salamin, N.; Lexer, C.; Paris, M. A dedicated target capture approach reveals variable genetic markers across micro-and macro-evolutionary time scales in palms. *Mol. Ecol. Resour.* **2019**, *19*, 221–234. [CrossRef]
41. Christe, C.; Boluda, C.G.; Koubínová, D.; Gautier, L.; Naciri, Y. New genetic markers for Sapotaceae phylogenomics: More than 600 nuclear genes applicable from family to population levels. *Mol. Phylogenet. Evol.* **2021**, *160*, 107123. [CrossRef] [PubMed]
42. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 238. [CrossRef] [PubMed]
43. Wu, Y.; Xu, J.; Han, X.; Qiao, G.; Yang, K.; Wen, Z.; Wen, X. Comparative transcriptome analysis combining SMRT-and Illumina-based RNA-Seq identifies potential candidate genes involved in betalain biosynthesis in pitaya fruit. *Int. J. Mol. Sci.* **2020**, *21*, 3288. [CrossRef] [PubMed]
44. Cheng, C.Y.; Krishnakumar, V.; Chan, A.P.; Thibaud-Nissen, F.; Schobel, S.; Town, C.D. Araport11: A complete reannotation of the Arabidopsis thaliana reference genome. *Plant J.* **2017**, *89*, 789–804. [CrossRef] [PubMed]
45. Van Bel, M.; Silvestri, F.; Weitz, E.M.; Kreft, L.; Botzki, A.; Coppens, F.; Vandepoele, K. PLAZA 5.0: Extending the scope and power of comparative and functional genomics in plants. *Nucleic Acids Res.* **2022**, *50*, 1468–1474. [CrossRef] [PubMed]
46. Zheng, J.; Meinhardt, L.W.; Goenaga, R.; Zhang, D.; Yin, Y. The chromosome-level genome of dragon fruit reveals whole-genome duplication and chromosomal co-localization of betacyanin biosynthetic genes. *Hortic. Res.* **2021**, *8*, 1468–1474. [CrossRef]

47. McLay, T.G.; Birch, J.L.; Gunn, B.F.; Ning, W.; Tate, J.A.; Nauheimer, L.; Joyce, E.M.; Simpson, L.; Schmidt-Lebuhn, A.N.; Baker, W.J.; et al. New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Appl. Plant Sci.* **2022**, *9*, e11420. [CrossRef]
48. Nyffeler, R.; Eggli, U. A farewell to dated ideas and concepts: Molecular phylogenetics and a revised suprageneric classification of the family Cactaceae. *Schumannia* **2010**, *6*, 109–149.
49. Arakaki, M.; Christin, P.A.; Nyffeler, R.; Lendel, A.; Eggli, U.; Ogburn, R.M.; Spriggs, E.; Moore, M.J.; Edwards, E.J. Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 8379–8384. [CrossRef]
50. Franco, F.F.; Silva, G.A.R.; Moraes, E.M.; Taylor, N.; Zappi, D.C.; Jojima, C.L.; Machado, M.C. Plio-Pleistocene diversification of Cereus (Cactaceae, Cereeae) and closely allied genera. *Bot. J. Linn.* **2017**, *183*, 199–210. [CrossRef]
51. Amaral, D.T.; Minhós-Yano, I.; Oliveira, J.V.M.; Romeiro-Brito, M.; Bonatelli, I.A.S.; Taylor, N.P.; Zappi, D.C.; Moraes, E.M.; Eaton, D.A.R.; Franco, F.F. Tracking the xeric biomes of South America: The spatiotemporal diversification of Mandacaru cactus. *J. Biogeogr.* **2021**, *48*, 3085–3103. [CrossRef]
52. Inglis, P.W.; Pappas, M.D.C.R.; Resende, L.V.; Grattapaglia, D. Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS ONE* **2018**, *13*, e0206085. [CrossRef] [PubMed]
53. Schubert, M.; Lindgreen, S.; Orlando, L. AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **2016**, *9*, 88. [CrossRef] [PubMed]
54. Johnson, M.G.; Gardner, E.M.; Liu, Y.; Medina, R.; Goffinet, B.; Shaw, A.J.; Zerega, N.J.C.; Wickett, N.J. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci.* **2016**, *4*, 1600016. [CrossRef] [PubMed]
55. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef] [PubMed]
56. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef] [PubMed]
57. Nguyen, L.T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef]
58. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
59. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2019**, *25*, 1972–1973. [CrossRef]
60. Borowiec, M.L. AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ* **2016**, *25*, 1972–1973. [CrossRef]
61. Hoang, D.T.; Chernomor, O.; Von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522. [CrossRef] [PubMed]
62. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.; Von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef] [PubMed]
63. Zhang, C.; Rabiee, M.; Sayyari, E.; Mirarab, S. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* **2018**, *19*, 15–30. [CrossRef] [PubMed]
64. Sayyarim, E.; Mirarab, S. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* **2016**, *33*, 1654–1668. [CrossRef] [PubMed]
65. Franco, F.F.; Jojima, C.L.; Perez, M.F.; Zappi, D.C.; Taylor, N.; Moraes, E.M. The xeric side of the Brazilian Atlantic Forest: The forces shaping phylogeographic structure of cacti. *Ecol. Evol.* **2017**, *7*, 9281–9293. [CrossRef] [PubMed]
66. Donoghue, M.J.; Sanderson, M.J. The suitability of molecular and morphological evidence in reconstructing plant phylogeny. In *Molecular Systematics of Plants*; Soltis, P.S., Soltis, D.E., Doyle, J.J., Eds.; Springer: Boston, MA, USA, 1992; pp. 340–368.
67. Eserman, L.A.; Thomas, S.K.; Coffey, E.E.; Leebens-Mack, J.H. Target sequence capture in orchids: Developing a kit to sequence hundreds of single-copy loci. *Appl. Plant Sci.* **2021**, *9*, e11416. [CrossRef] [PubMed]
68. Hendriks, K.P.; Mandáková, T.; Hay, N.M.; Ly, E.; Hooft van Huysduynen, A.; Tamrakar, R.; Thomas, S.K.; Toro-Núñez, O.; Pires, C.; Nikolov, L.A.; et al. The best of both worlds: Combining lineage-specific and universal bait sets in target-enrichment hybridization reactions. *Appl. Plant Sci.* **2021**, *9*, e11438. [CrossRef] [PubMed]
69. Edwards, E.J.; Nyffeler, R.; Donoghue, M.J. Basal cactus phylogeny: Implications of Pereskia (Cactaceae) paraphyly for the transition to the cactus life form. *Am. J. Bot.* **2005**, *92*, 1177–1188. [CrossRef]
70. Fantinati, M.R.; Soffiatti, P.; Calvente, A. A new phylogenetic hypothesis for Cereinae (Cactaceae) points to a monophyletic subtribe. *Syst. Bot.* **2021**, *46*, 689–699. [CrossRef]
71. Schlumpberger, B.O.; Renner, S.S. Molecular phylogenetics of Echinopsis (Cactaceae): Polyphyly at all levels and convergent evolution of pollination modes and growth forms. *Am. J. Bot.* **2012**, *99*, 1335–1349. [CrossRef]

72. Lendel, A. South American Cacti in Time and Space: Studies on the Diversification of the Tribe Cereeae, with Particular Focus on Subtribe Trichocereinae (Cactaceae). Ph.D. Thesis, Mathematisch- Naturwissenschaftlichen Fakultat der Universitat Zurich, Zurich, Switzerland, 2013.
73. Taylor, N.; Zappi, D. Notes on plants called Cereus hexagonus (Cactaceae). *Bradleya* **2019**, *37*, 17–25. [CrossRef]

CHAPTER 2: Unraveling phylogenetic relationships of tribe Cereeae using target enrichment sequencing

Monique Romeiro-Brito[1], Nigel P. Taylor[2], Daniela C. Zappi[3], Milena C. Telhe[1], Fernando F. Franco[1] and Evandro M. Moraes[1]

[1] Departamento de Biologia. Centro de Ciências Humanas e Biológicas. Universidade Federal de São Carlos (UFSCar). Sorocaba, Brazil.

[2] University of Gibraltar, Gibraltar Botanic Gardens Campus, Gibraltar.

[3] Programa de Pós-Graduação em Botânica, Instituto de Ciências Biológicas Universidade de Brasília (UNB), Brasília, Brazil.

Draft manuscript
To be submitted to: Annals of Botany

Nomenclatural novelties presented here are not intentended to be validly published.

Abstract

Cactaceae are succulent plants, mostly endemic to the American continent, and one of the most endangered taxa groups in the world. Molecular phylogenies had been key to unraveling phylogenetic relationships among major cacti groups, once hampered by high levels of morphological plasticity. Phylogenetic studies using plastid markers did not provide adequate resolution for determining generic relationships within cactus groups. This is the case for the tribe Cereeae s.l., a highly diverse group from South America. To comprehensively understand suprageneric and generic relationships within the tribe Cereeae, we integrated sequence data from public genetic databases with a newly generated sequence dataset of 591 nuclear regions across representatives of this tribe. We inferred concatenated and coalescent phylogenetic trees, and compared the performance of both approaches. The definition of major suprageneric groups was mostly congruent among different datasets, but only the dataset from the 591 nuclear regions recovered a fully resolved topology at suprageneric and generic levels. We propose a new taxonomic classification within Cereeae based on well-resolved clades, including new subtribes (Aylosterinae **subtr. nov.**, Uebelmanniinae **subtr. nov.**, Gymnocalyciinae **subtr. nov.**) and revised subtribes (Trichocereinae, Rebutiinae, and Cereinae). We shed light on the mandatory use of genomic datasets allied with coalescent inference to recover and explore evolutionary patterns within the tribe Cereeae.

Keywords: Cactaceae, Cactoideae, South American cacti, systematics, phylogenomics, Cactaceae591 probe set.

2.1 Introduction

Cactus species are succulent and conspicuous plants found primarily in dry and arid regions in Neotropics. This family is one of the most diverse groups of succulent plants in Caryophyllales (Hernández-Ledesma et al. 2015), and one of the most endangered taxa in the world (Goettsch et al. 2015, Pillet et al. 2022, Amaral et al. 2022). Over the last two decades, molecular phylogenetics has been fundamental to elucidate phylogenetic relationships among major lineages of this family (Nyffeler 2002, Bárcenas et al. 2011, Hernández-Hernández et al. 2011), previously hindered by the high level of morphological convergence. Traditionally, this family is subdivided into four subfamilies, named Maihuenoideae, Pereskeoideae, Opuntioideae, and Cactoideae (Guerrero et al. 2018), the last one exhibiting the greatest species diversity and morphological variation (Anderson, 2001, Applequist and Wallace 2002, Hunt et al. 2006, Nyffeler and Eggli 2010, Hernández-Hernández et al. 2011). Regardless of these efforts, the delimitation of suprageneric lineages within the subfamily Cactoideae has long been a subject of debate, particularly for the core Cactoideae II (Guerrero et al. 2018), a group of South American cacti with unclear phylogenetic placement.

The main systematic controversy within the core Cactoideae II lies upon a well-supported clade composed of globose and columnar cacti from the traditionally circumscribe tribes Browningieae, Cereeae, and Trichocereeae (also known as BCT clade, Nyffeler 2002, Nyffeler & Eggli 2010). The relationships among these tribes remained unknown due to the non-monophyly at the tribe and genus levels (Nyffeler 2002, Ritz et al. 2007, Hernández-Hernández et al 2011, Schlumberger and Renner 2012, Franck et al. 2013, Lendel 2013, Ritz et al. 2016, Franco et al. 2017, Calvente et al. 2017, Fantinati et al. 2022). The most recent systematic classification proposed for the family broadened the circumscription of the tribe Cereeae, included all representatives from the BCT clade (tribe Cereeae *sensu lato*), and divided it into three subtribes: Cereinae, Trichocereinae, and Rebutinae (Fig. 1, Nyffeler and Eggli 2010). In spite of this taxonomic rearrangement, non-monophyletic groups remained persistent in the tribe Cereeae s.l (Schlumberger and Renner 2012, Romeiro-Brito et al. 2022).

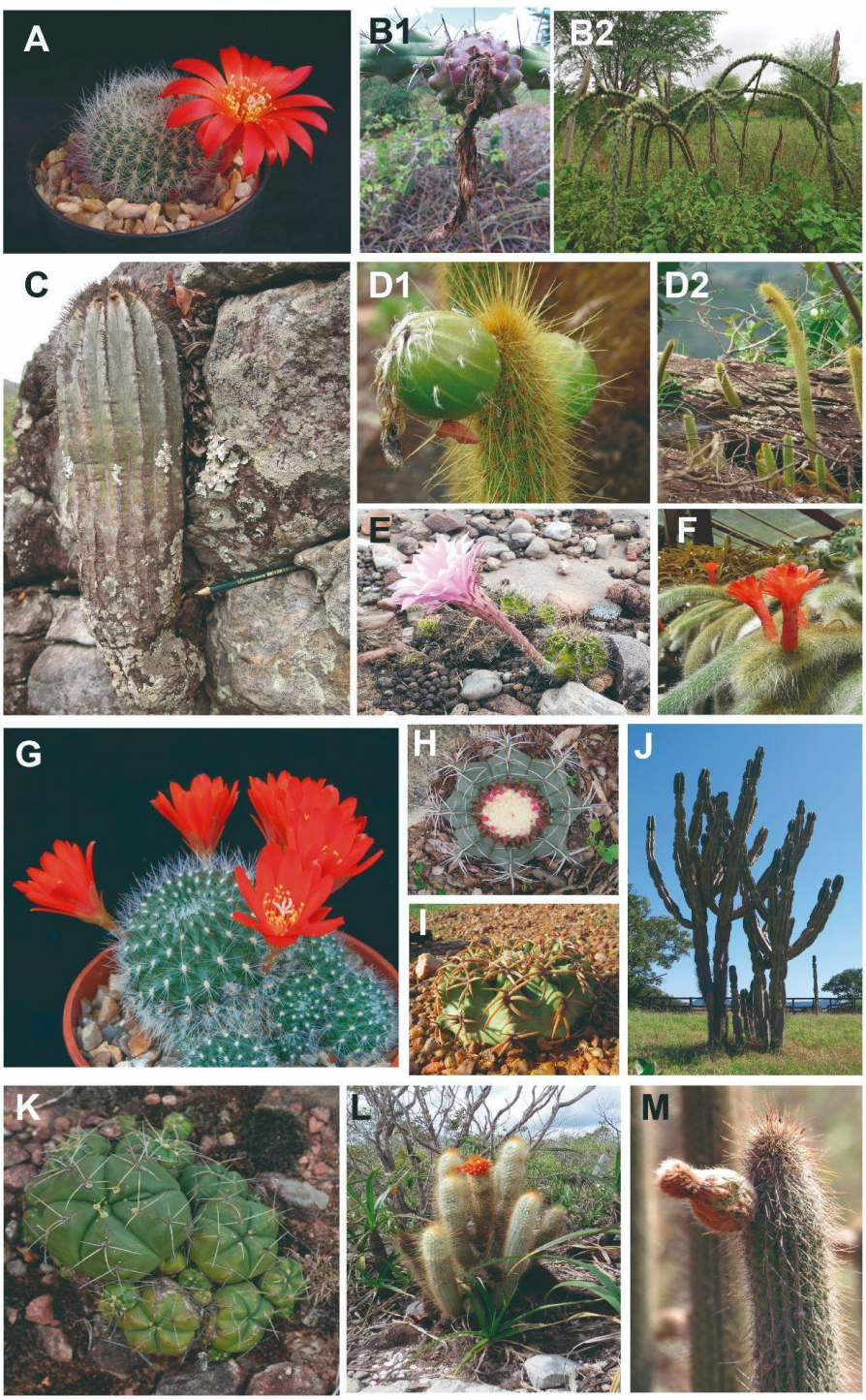


Figure 1. Representatives of subtribe Rebutiinae (A, C, G, K), Trichocereinae (B1, B2, D1, D2, E, F), and Cereinae (H, I, J, L, M) sensu Nyffeler and Eggli (2010) showing examples of growth form variation in those major groups. A) *Rebutia minuscula* K. Schum, B1 and B2) *Harrisia adscendens* (Gürke) Britton & Rose, C) *Uebelmannia pectinifera* Buining, D1 and D2) *Arthrocereus rondonianus* Backeb. & Voll, E) *Echinopsis oxygona* (Link) Zucc. ex Pfeiff., F) *Cleistocactus winteri* D.R. Hunt, G) *Aylostera fiebrigii* (Gürke) Backeb., H) *Melocactus glaucescens* Buining & Brederoo, I) *Discocactus bahiensis* Britton & Rose, J) *Cereus jamacaru* DC, K) *Gymnocalycium denudatum* (Link & Otto) Pfeiff. ex Mittler, L) *Micranthocereus auriazureus* Buining & Brederoo and M) *Facheiroa squamosa* (Gürke) P.J.Braun & Esteves. Photo credits: A and G: M. Lowry; B, D, H, I, L and M: M. C. Telhe; E and K) M. Köhler; F and J) M. Romeiro-Brito.

Few studies have investigated phylogenetic relationships using morphological or molecular data at the suprageneric level in the tribe Cereeae s.l. (Braun 1988, Taylor and Zappi, 1989, Soffiati 2003, Nyffeler and Eggli 2010, Lendel 2013, Fantinati et al. 2022). The circumscription of the subtribes within Cereeae lacks apomorphic characteristics capable of distinguishing them, particularly between the subtribes Trichocereinae and Cereinae (Taylor and Zappi 2004). While Trichocereinae traditionally included taxa with flowers and fruits with trichomes, bristles, and scales, Cereinae is composed of species with naked flowers (or with minute scales, Barthlott and Hunt 1993). Thought to be the result of convergent phenotypic evolution, these reproductive morphological characters have not been sufficient for delimiting subtribes and genera in both groups (Taylor and Zappi 1989, Schlumberger and Renner 2012, Franck et al. 2013). Besides, molecular systematic studies in this group had shown low phylogenetic support to delineate groups at generic levels (Lendel 2013, Fantinati et al. 2022). The lack of phylogenetic resolution might occur due to the recent and rapid diversification of this group during the late Miocene (Hernández-Hernández et al. 2014), but also as a consequence of poor phylogenetic signal when using few traditional molecular markers.

The use of hight-throughput sequencing has been emerging as an alternative to traditional molecular markers to study phylogenetic relationships in Cactaceae (Franco et al. 2022). So far, most phylogenomic studies in the family have focused on surveying relationships at subfamily, genus and species levels using different high-throughput sequencing approaches (RAD-seq, Bombonato et al. 2020, Amaral et al. 2021b; GBS, Breslin et al. 2021, Merklinger et al. 2021; genome skimming, Majure et al. 2019, Köhler et al. 2020; Majure et al. 2021, Majure et al. 2022, Majure et al. 2023). The recent availability of genome and transcriptome sequencing of Cactaceae representatives (Copetti et al. 2017, Walker et al. 2018, Wang et al. 2018, Amaral et al. 2021a, Chen et al. 2022) has enabled the development of lineage-specific probe sets for capturing nuclear genes in Cactaceae (Acha and Majure 2022, Romeiro-Brito et al. 2022). Indeed, the target sequencing capture approach has proven to be an effective resource for unraveling relationships at shallow and deep levels in Cactaceae (Romeiro-Brito et al. 2022) and delineating the contentious relationship of a recently diverged genus in this group (Taylor et al. in revision). Therefore, this strategy is a potential candidate to resolve phylogenetic relationships and propose taxonomic adjustments within Cereeae.

In this study, our main goal was to infer suprageneric relationships within the tribe Cereeae s.l. and generic relationships in the subtribe Cereinae, proposing a new taxonomic classification. To this end, we integrated newly generated genome-scale data gathered with Cactaceae591 target sequencing probe set and with sequence data from public databases.

Considering the long-lasting controversies in the systematics of Cactoideae (Nyffeler and Eggli, 2010), we assume as hypotheses the non-monophyly of the current circumscription of subtribes Cereinae, Rebutiinae and Trichocereinae, as well as of the current generic delimitation within the subtribe Cereinae. Furthermore, we explored the potential of concatenated and coalescent inference approaches in resolving contentious phylogenetic relationships in the subtribe Cereinae.

2.2 Material and Methods

2.2.1 Datasets and taxon sampling

In order to explore distinct gene and genome-scale datasets to estimate the relationships within the tribe Cereeae, and to cover a broad sampling of the focal clade which otherwise would be missing, we used different datasets compiled from newly generated sequences as well as publicly available data. Our datasets were named as follows: ‘Cactaceae591’, which was composed of newly generated data from more than five hundred nuclear regions obtained with Cactaceae591 target sequencing panel (Romeiro-Brito et al. 2022); 'Angiosperm353’, composed of fasta data from more than three hundred nuclear regions generated with the universal probe set Angiosperm353 (Johnson et al. 2019), which were retrieved from the Kew Tree Of Life Explorer Platform (Baker et al. 2022, https://treeoflife.kew.org/tree-of-life); and ‘gene-scale’ consisted of sequences of eight plastids (*rbcL*, *atpB-rbcL*, *trnK-matK*, *rpL16*, *petL-psbE*, *trnT-trnL*, *trnL-trnF*, *trnS-trnG*) and two nuclear (*phyC* and *ppc*) regions, which were recovered from our Cactaceae591 raw dataset and from the GenBank (National Center for Biotechnology Information, NCBI).

The Cactaceae591 dataset consisted of 29 of 48 recognized genera and 122 of 561 recognized species from the tribe Cereeae s.l. (assuming accepted genera of tribe Cereeae according to Nyffeler and Eggli et al. 2010, and genera and species count according to Korotkova et al. 2021). This sampling comprised all genera from the subtribe Cereinae (14 genera), three of seven genera from the subtribe Rebutinae, and 12 of 27 genera from the subtribe Trichocereinae (Table S1). In order to confirm the phylogenetic position in cases of monotypic genera (e.g. *Espostoopsis* Buxb. and *Leocereus* Britton & Rose) or lineages with controversial taxonomy (e.g. species from genus *Micranthocereus* Backeb.), we sampled two specimens from different geographic records or different collectors.

The Angiosperm353 dataset included 43 of 48 recognized genera and 53 of 561 recognized species from the tribe Cereeae s.l.. This sampling comprised all genera from the subtribe Cereinae, six of seven genera from the subtribe Rebutinae, and 23 of 27 genera from the subtribe Trichocereinae (Table S1).

The gene-scale sampling dataset was obtained using SuperCRUNCH v1.3 pipeline (Portik and Wiens 2020) to parse, edit and generate the dataset gathered from GenBank. First, we updated the recognized species name according to synonyms from the latest checklist of Cactaceae in Korotokova et al. (2021) and then selected all sequences from the tribe Cereeae s.l., allowing up to 60% of missing data for each locus. Whenever a taxon presented multiple entries for a given locus, we prioritized selecting the longest sequence or, when possible, the one recovered from our Cactaceae591 dataset (details of the final dataset in Table S2). This dataset included 47 of 48 recognized genera and 177 of 561 recognized species from the tribe Cereeae s.l.. The final sampling comprised all genera from the subtribe Cereinae and Rebutinae, and only one genus (*Weberbauerocereus* Backeb.) from the subtribe Trichocereinae is missing (Table S2).

We attempted to select the same species and/or genus for our outgroup sampling in the three datasets (Table S1 and S2). The outgroup taxa included representatives of major clades of the subfamily Cactoideae (tribe Notocacteae, tribe Rhipsalideae, tribe Phyllocacteae [Core Cactoideae I], and tribe Cacteae; *sensu* Nyffeler and Eggli, 2010), representatives of the subfamilies Opuntioideae and Pereskioideae, and one representative of Portulacaceae (*Portulaca hirsutissima* Cambess.).

2.2.2 DNA extraction and target capture sequencing library of Cactaceae591 dataset

Genomic DNA for Cactaceae591 dataset was extracted from the root and epidermis from preserved at -80° C tissues or herbarium material using a high-salt CTAB protocol (modified from Martínez-González et al. (2017) and Inglis & al. (2018), and detailed in Appendix S1). We isolated and enriched 591 regions using the target capture probe-set Cactaceae591 (Romeiro-Brito et al. 2022). This probe set includes 587 genic (exonic and intronic regions) and intergenic nuclear regions and also four regions commonly used in Cactaceae studies (plastidial regions: *trnK-matK* and *rbcL*; nuclear regions: *ppc* and *phyc*). Library preparation and sequencing were performed by RAPiD Genomics LLC (Gainsville, Florida, USA) utilizing their high-throughput workflow with proprietary chemistry. Samples

were pooled in equimolar concentrations and sequenced on a NovaSeq 6000 (2 × 150 bp and 2 × 250 bp).

2.2.3 Processing raw reads of Cactaceae591 dataset

Raw reads from Cactaceae591 dataset were trimmed using AdapterRemoval v2 (Schubert and al. 2016), removing poor quality reads (phred < 20), adapters, and short reads (< 60 bp). The trimmed reads were mapped and assembled using HybPiper v2.0.3 pipeline (Johnson et al. 2016, https://github.com/mossmatters/HybPiper), using our references available in Romeiro-Brito et al. 2021). We obtained the 591 on-target regions using the DNA sequence references available in Romeiro-Brito et al. (2022), setting eight reads as the minimum coverage. The putative paralogs identified in Romeiro-Brito et al. (2022) were removed from the final dataset. We skimmed additional plastid regions commonly used for Cereeae s.l. phylogenies (*atpB-rbcl, rpL16, petL-psbE, trnT-trnL, trnL-trnF, trnS-trnG* Appendix S1) as off-target regions with HybPiper by using Cactaceae reference sequences and setting the minimum coverage reads to 4.

2.2.4 Alignment and trimming genome-scale datasets

We used MAFFT v7 (Katoh and Standley 2013) for sequence alignment using auto option and trimAL v1.3 (Capella-Gutiérrez et al. 2009) for removing gaps found in at least 40% of the sequences using gt argument and poorly aligned regions using st argument.

We removed from Cactaceae591 and Angiosperm353 data the nuclear regions with more than 60% of missing data and the outlier regions using the genesortR script (Mongiardino Koch 2021). Here, the outlier loci were represented by loci with phylogenetic metrics that deviated significantly among all other loci. This script uses phylogenetic signal (e.g. average bootstrap support) and phylogenetic biases metrics (e.g. level of saturation, and root-to-tip variance) and resumes it using a principal component analysis, accounting for the increase of phylogenetic usefulness of each locus against major phylogenetic biases. This script requires concatenated alignment, a partition file, gene trees of each locus, and a species tree. The gene trees were estimated using IQ-TREE v2 (Mihn et al. 2020) with 1,000 ultrafast bootstrap replicates (Hoang et al. 2018) and rooted using the pxrr program available in phyx (Brown et al. 2017). The species tree was summarized using ASTRAL v5.7 (Sayyari and Mirarab 2016).

Finally, we removed 5% of outlier genes according to the genesortR script. We estimated genetic diversity statistics from all datasets using AMAS (Borowiec 2016).

2.2.5 Concatenated and coalescent phylogenetic inference

We used the genome-scale datasets Cactaceae591 and Angiosperm353 to estimate phylogenetic trees using both the concatenated and coalescent approaches. For the gene-scale dataset, we implemented only the concatenated approach. The concatenated inferences were implemented in IQ-TREE v2 (Mihn et al. 2020) with 10,000 ultrafast bootstraps (UfBoot) replicates (Hoang et al. 2018) and 10,000 SH-like approximate likelihood ratio test (SH-aLRT) replicates (Guindon et al. 2010). The best substitution model for each partition was estimated by ModelFinder (Kalyaanamoorthy et al. 2017), using the -m command available on IQ-TREE. We inferred coalescent species trees using the summary approach implemented in ASTRAL-hybrid (Zhang and Mirarab 2022). The ML gene trees were generated in IQ-TREE v2 (Mihn et al. 2020) with 10,000 ultrafast bootstrap replicates (Hoang et al. 2018). We set the minimum bootstrap values of the gene trees branch support to be considered for species tree estimation as 40. The branch supports of species trees were accessed by local posterior probabilities (LPP).

We assessed gene tree conflict against species tree inferred from the Cactaceae591 dataset using PhyParts (Smith et al. 2015) and checked for gene tree incongruences using PhyPartsPieCharts (https://github.com/mossmatters/phyloscripts/tree/master/phypartspiecharts).

2.3 Results

2.3.1 Variability of genetic and genomic datasets

The efficiency of capture of Cactaceae591 regions was higher within the representatives of Cactoideae (ranging from 480 to 567 regions) compared to representatives of Opuntioideae (ranging from 383 to 404 regions), Pereskioideae (ranging from 485 to 521 regions), and Portulacaceae (137 regions) (Fig. S1). After removing regions with more than 60% of missing data, paralogs, and outlier regions, the Cactaceae591 dataset included 459 nuclear regions and the Angiosperm353 dataset included 318 nuclear regions (Table 1). The Cactaceae591 dataset presented the highest proportion of variable sites, parsimony informative sites, and branch support among the three datasets delimited in this study (Table 1). The Angiosperm353

presented the least percentage of missing data and the gene-scale dataset presented the most complete taxon sampling regarding genera and species sampling (Table 1).

Table 1. Genetic statistics of nucleotide variation calculated for the datasets used in the present study. bp: base pairs; S: variable sites, PIS (%): parsimony informative sites (proportion of PIS). Supported nodes: percentage of supported nodes from concatenated inferences for the gene-scale (ultrafast bootstrap > 95) and from concatenated and coalescent inferences for the genome-scale datasets (ultrafast bootstrap > 95/ local posterior probability > 0.8). No. of taxa includes both ingroup and outgroup sampled taxa in each dataset.

Dataset	Nº.loci	Nº.taxa	Alignment length (bp)	% missing data	S (%)	PIS (%)	Supported nodes (%)
Cactaceae591	459	146	511,090	15.75	287,596 (0.563)	171,109 (0.335)	94.1/93.5
Angiosperm353	318	84	136,589	12.63	59,975 (0.439)	27,508 (0.201)	82.3/87.1
gene-scale	10	209	9,923	34.36	3,502 (0.353)	1,901 (0.191)	63.4

2.3.2 Phylogenetic relationships among major clades of tribe Cereeae

All phylogenetic inferences recovered the tribe Cereeae as a well-supported clade and sister to the tribe Notocacteae (Fig. 2). None of the Cereeae subtribes circumscribed by Nyfeller and Eggli (2010) were recovered as monophyletic, regardless of the dataset used in this study. The genome-scale datasets (Cactaceae591 and Angiosperm353) provided a higher phylogenetic resolution along all nodes of the tribe Cereeae (Table 1, Fig. 2a and 2b, respectively). The gene-scale dataset lacked node support in some of the backbone and shallow relationships of the Cereeae phylogeny (Fig. 2c and Fig. S2). Overal, six major clades within tribe Cereeae were recovered with high resolution regardless of the datasets and phylogenetic inferences: *Uebelmannia* clade, *Aylostera* clade, Rebutiinae clade (including *Browningia* Britton & Rose, *Rebutia* K.Schum. and *Weingartia* Werderm.), Trichocereinae clade (excluding *Espostoopsis dybowskii*), *Gymnocalycium* Pfeiff. clade, and Cereinae clade (including *Stetsonia coryne* (Salm-Dyck) Britton & Rose and *Espostoopsis dybowskii*). We were not able to infer whether *Uebelmannia* Buining or *Aylostera* Speg. was the first divergent group within the tribe Cereeae, either because of low phylogenetic resolution or insufficient taxa representation of each clade (Fig. 2). However, both genome-scale datasets were

consistent in recovering the Rebutiinae clade as a sister lineage to the remaining representatives of the tribe Cereeae s.l: subtribe Trichocereinae, *Gymnocalycium* clade and subtribe Cereinae.

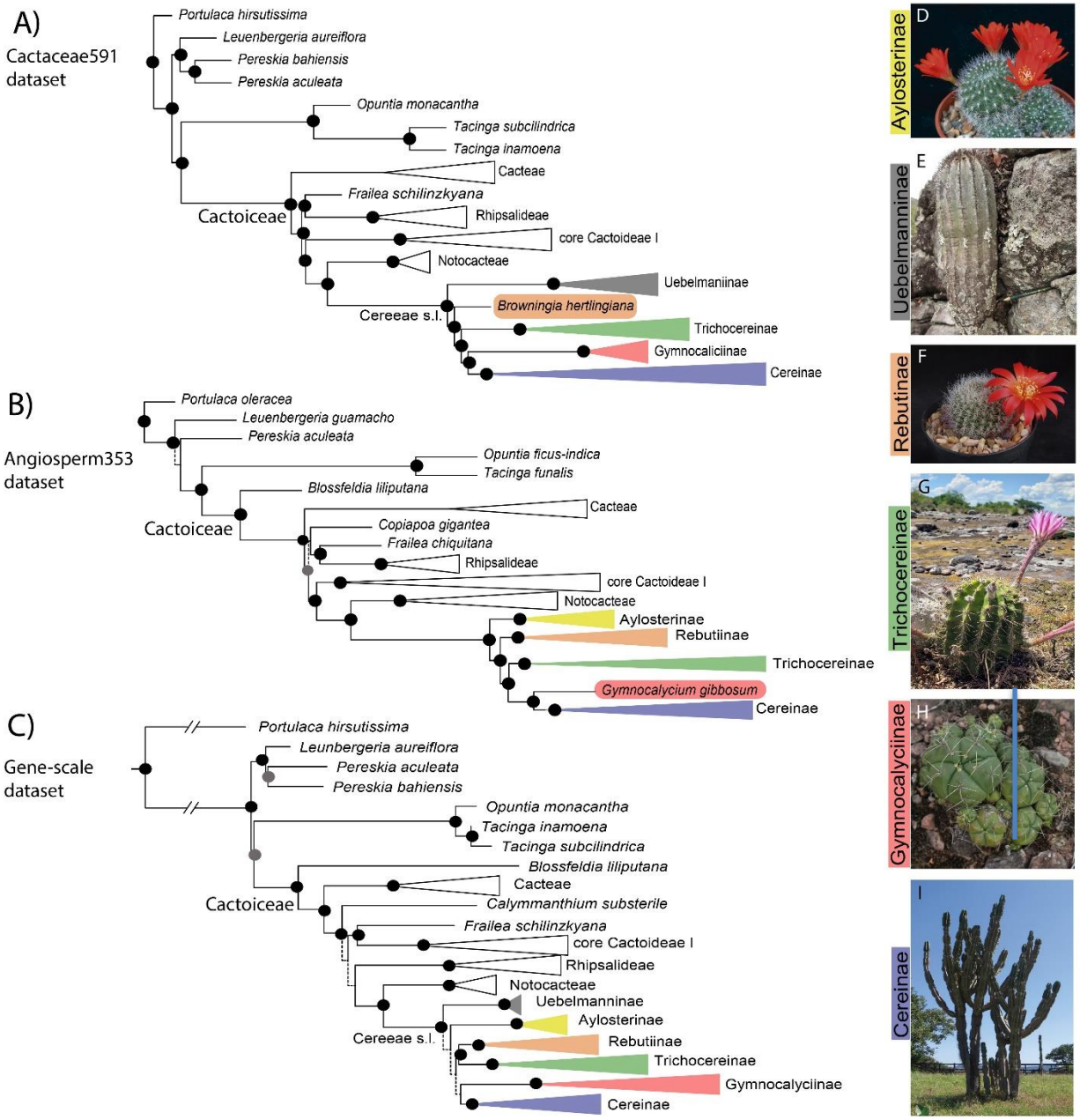


Figure 2. Phylogenetic tree reconstructions showing major clades of subfamily Cactoideae and tribe Cereeae from coalescent-based inference of Cactaceae591 dataset (A), coalescent-based inference of Angiosperm353 dataset (B) and maximum likelihood inference of gene-scale dataset (C). Highly supported branches (LPP > 0.9 for coalescent-based inference and BS/SH-aLRT > 95/80 for maximum likelihood inference) are depicted with black circles; moderately supported branches (0.9 > LPP > 0.7 for coalescent-based inference and 95 > BS and 80 < SH-aLRT for maximum likelihood) are depicted with gray circles. Low supported nodes (LPP < 0.7 in coalescent-based inference and BS/SH-aLRT < 95/80 in maximum likelihood inference) are represented by dashed branch lines.

2.3.3 Phylogenetic relationships within major clades of tribe Cereeae s.l.

Within Trichocereinae, the Cactaceae591 dataset recovered *Arthrocereus* A.Berger and *Harrisia* as the early divergent lineages of the subtribe, whereas the Angiosperm353 dataset recovered *Reicheocactus* Backeb. and *Harrisia* Britton as the early divergent lineages of this group (Fig. 3 and 4). The phylogenetic inferences of all datasets supported *Cleistocactus* Lem. s. str. clade sister to *Oreocereus* (A.Berger) Riccob clade, although the genus *Cleistocactus* was recovered as a non-monophyletic group either in genomic-scale and gene-scale datasets (Fig. 3, 4 and Fig S2). Both *Cleistocactus* and *Oreocereus* clades were sisters to *Echinopsis* s.l. clade using the genome-scale datasets, considering that in the Angiosperm353 dataset the *Echinopsis* Zucc. clade included a representative of the genus *Denmoza* Britton & Rose (Fig 3). The gene-scale dataset did not support *Echinopsis* s.l. as monophyletic, recovering some representatives closely related to the genus *Harrisia* and the remaining *Echinopsis* representatives in a clade including the genus *Denmoza* (Fig S2). Another incongruence among datasets is related to the genus *Lasiocereus* F. Ritter, which is allied to *Browningia* and *Rebutia* clade in the gene-scale dataset (Fig S2), but closely related to *Espostoa* in Angiosperm353 dataset (Fig 3).

In all analyses, the genus *Gymnocalycium* was recovered as a sister group to the subtribe Cereinae s.l, which includes *Stetsonia coryne* and *Espostoopsis dybowskii* according to all phylogenetic inferences, regardless of the dataset used. The genomic dataset resolved the generic relationships within the subtribe Cereinae (especially the Cactaceae591 dataset Fig 5), while the gene-scale dataset did not present enough phylogenetic signal to support deep and shallow nodes of subtribe Cereinae (Fig S2).

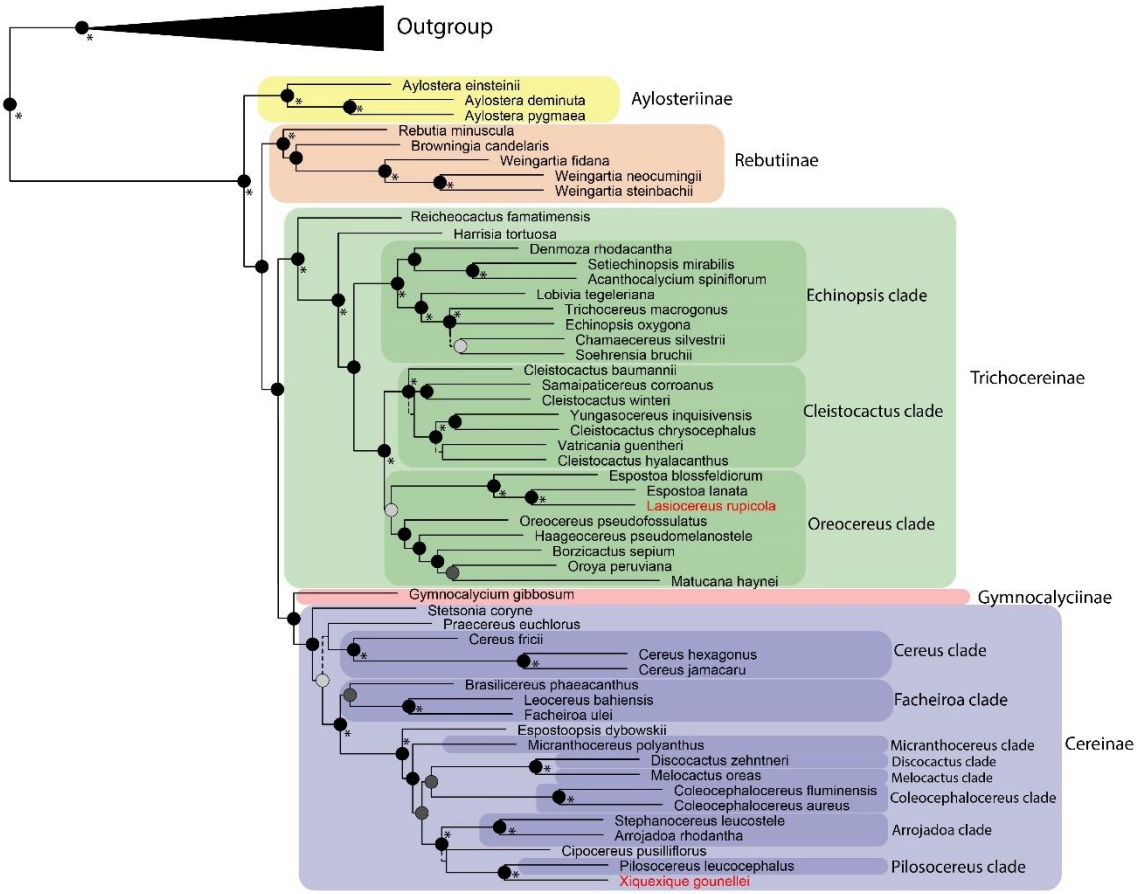


Figure 3. Phylogenetic tree reconstructions of coalescent-based inference using Angiosperm353 dataset, showing the relationships within major clades of tribe Cereeae. Highly supported branches (LPP > 0.95) are depicted with black circles in the nodes. Nodes with LPP = 1 are highlighted with an asterisk. Moderately supported branches are depicted with dark gray (0.95 > LPP > 0.8) and light gray (0.8 > LPP > 0.75) circles. Low supported nodes (LPP < 0.7) are shown with dashed lines in respective branches.

The backbone of the subtribe Cereaine showed several short internal branches (Fig 5). The first group diverging this subtribe was the genus *Stetsonia*, followed by the *Praecereus* Buxb.. *Cereus* Mill. clade (including *Cipocereus* F.Ritter), *Facheiroa* Britton & Rose clade (including *Brasilicereus* Backeb. and *Leocereus*), and the monotypic genus *Espostoopsis*. The next well-supported group comprises most of the diversity of the subtribe Cereinae.

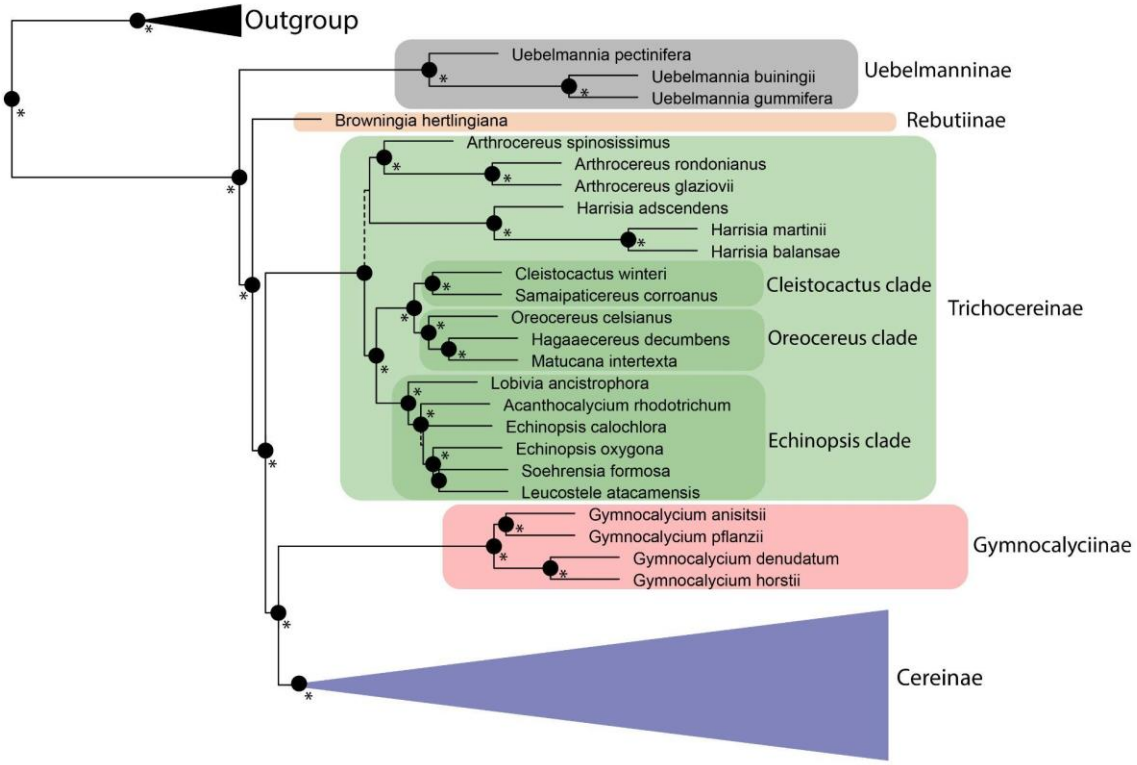


Figure 4. Phylogenetic tree reconstructions of coalescent-based inference using Cactaceae591 dataset, highlighting the relationships of early-diverging lineages of tribe Cereeae and within subtribe Trichocereinae. Highly supported branches (LPP > 0.95) are depicted with black circles in the nodes. Nodes with LPP = 1 are highlighted with an asterisk. Moderately supported branches (0.95 > LPP > 0.8) are depicted with dark gray circles. Low supported nodes (LPP < 0.7) are shown with dashed lines in respective branches.

Melocactus Link & Otto, *Discocactus* Pfeiff., *Coleocephalocereus* Backeb. and *Xiquexique* Lavor, Calvente & Versieux correspond to the few monophyletic genera within the subtribe Cereinae. We highlight the polyphyly of the genus *Micranthocereus*, whose representatives are scattered within the three main clades assigned as: *Micranthocereus* s. str. clade (which includes the type *Micranthocereus polyanthus* (Werderm.) Backeb. and its allies, *Xiquexique* and *Pilosocereus bohlei* Hofacker), *Coleocephalocereus* clade (which includes *Micranthocereus* subg. *Siccobaccatus* P.J.Braun & Esteves), and *Arrojadoa* clade (including *Arrojadoa* Britton & Rose, *Stephanocereus* A.Berger, *Cipocereus pusilliflorus* (F.Ritter) Zappi & N.P.Taylor, *Micranthocereus violaciflorus* Buining.). The position of one representative from the genus *Micranthocereus* (*Micranthocereus purpureus* (Gürke) F.Ritter) was nested within the outgroup taxa in the Angiosperm353 phylogenetic tree (Fig 3), although this same

species is recovered as closely related to *Micranthocereus* s. str. clade in the Cactaceae591 dataset (Fig. 5).

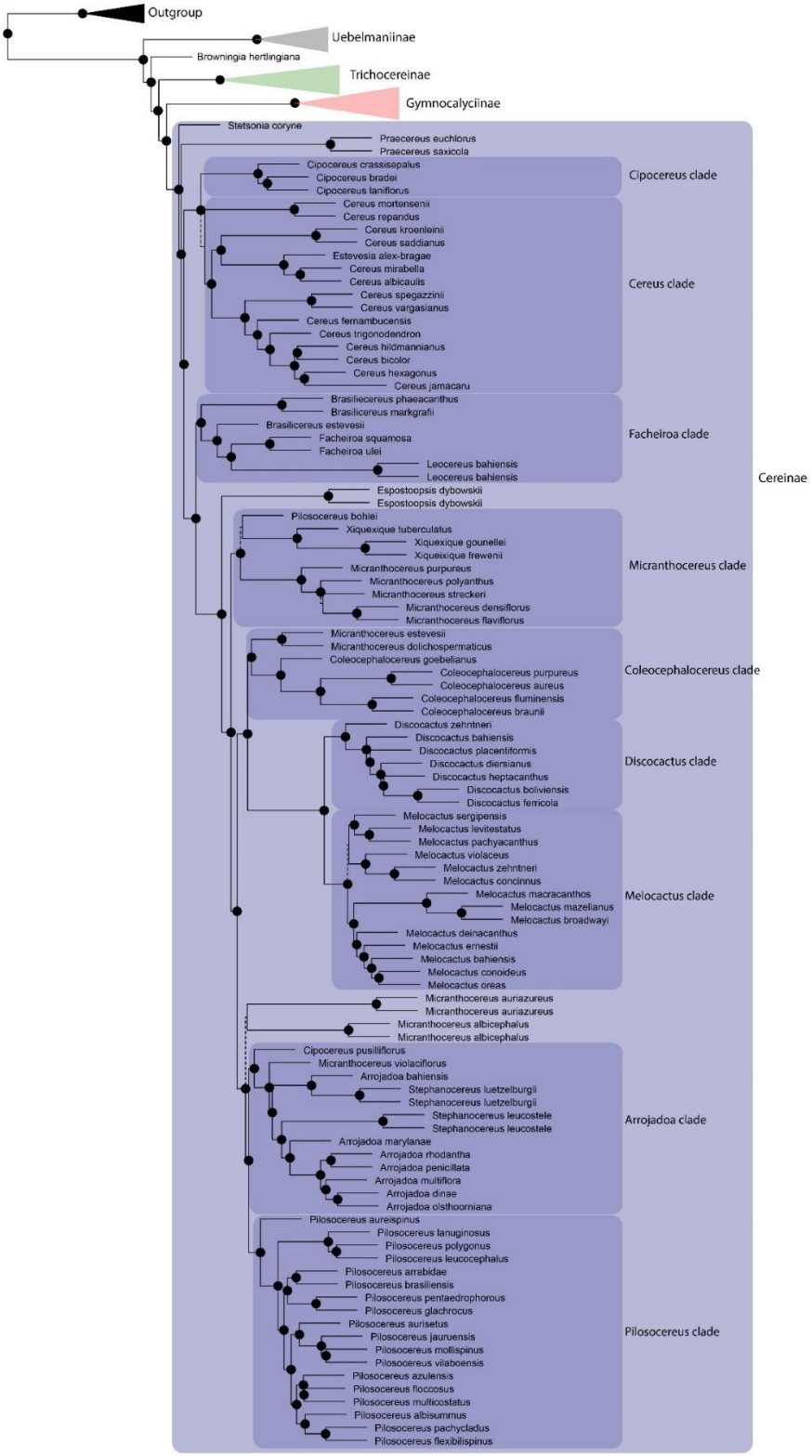


Figure 5. Phylogenetic tree reconstructions of coalescent-based inference using Cactaceae591 dataset, highlighting the relationships within subtribe Cereinae. Highly supported branches (LPP > 0.95) are depicted with black circles in the nodes. Nodes with LPP = 1 are highlighted with an asterisk. Moderately supported branches (0.95 > LPP > 0.8) are depicted with dark gray circles. Low supported nodes (LPP < 0.7) are shown with dashed lines in respective branches.

2.3.4 Incongruences among concatenated and coalescent inferences using the genomic dataset

The backbone of the phylogenies obtained from the genome-scale datasets was congruent across different phylogenetic approaches, i.e., multispecies coalescent and maximum-likelihood based on a concatenated dataset (Fig. S3 and S4), only differing in the presence of the lineages *Uebelmannia* and *Aylostera* in each genomic dataset. In shallower levels, incongruences among phylogenetic approaches occurred within the genus *Melocactus*, *Cereus*, *Micranthocereus* s.l. and *Pilosocereus* clade in Cactaceae591 dataset (Fig S3), and within *Brasilicereus* and *Cleistocactus* clade in Angiosperm353 dataset (Fig S4). Most incongruences presented high phylogenetic support in both phylogenetic approaches (represented by red lines in Fig S3 and Fig S4). Incongruences involving generic level are related to the placement of *Cipocereus pusilliflorus* (whether close with *Arrojadoa* s.l. clade or with *Micranthocereus albicephalus*) and the relationship among *Cereus* and *Cipocereus* in Cactaceae591 dataset (Fig S3).

The gene tree discordances had shown to be widespread throughout the evolutionary history of tribe Cereeae, with no dominant alternative topology among the gene trees in the Cactaceae591 dataset (Fig 6). The genera that display a higher proportion of gene tree-species tree concordant topology are *Uebelmannia*, *Harrisia*, *Gymnocalycium*, *Praecereus* and *Xiquexique* (Fig 6).

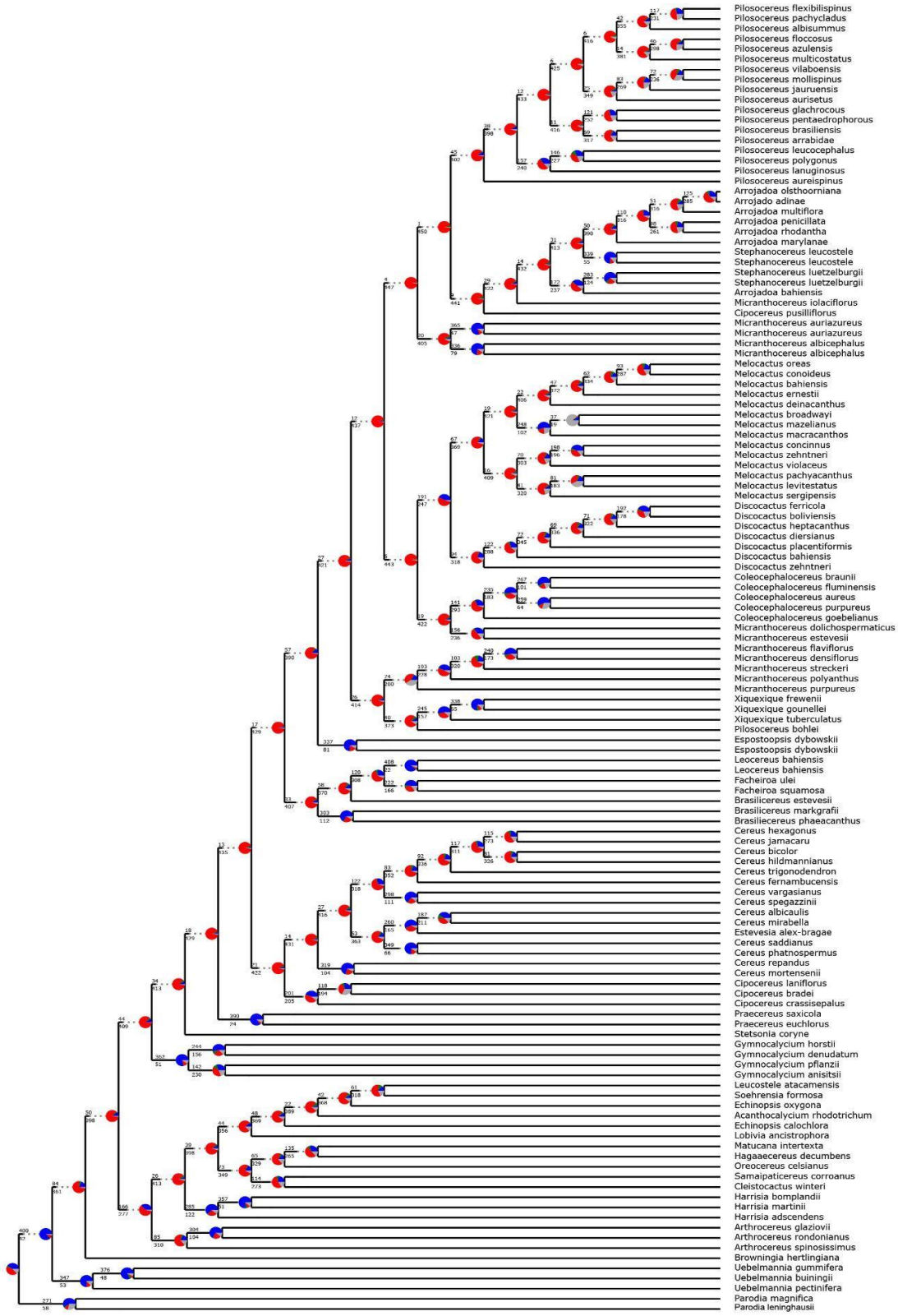


Figure 6. Gene-tree and species-tree conflicts estimated in PhyParts using coalescent-based phylogeny inferred with the Cactaceae591 dataset. The pie chart in the nodes represents the proportion of gene trees congruent with species tree in blue, genes supporting alternative dominant relationships against species tree in green, gene trees with all other incongruent relationships in red, and gene trees with no support in gray.

2.4 Discussion

The target sequencing strategies have demonstrated their effectiveness in unraveling contentious phylogenetic relationships faced in plant groups that experienced recent radiation, despite extensive patterns of morphological convergence and gene tree-species tree discordance (Lagomarsino et al. 2023). These strategies may contribute to the progress toward a robust taxonomic classification at multiple levels in Cactaceae, particularly using the customized Cactaceae591 probe set (Romeiro-Brito et al. 2022). This panel has proven to recover well-supported topologies within tribe Cereeae, even within challenging groups, as the relationships within the genera *Cereus* and *Cipocereus* (Taylor et al., under review), and may recover the position of orphan or *incertae sedis* lineages as well. For instance, both genomic datasets surprisingly revealed a close and well-supported relationship between *Frailea* Britton & Rose (an orphan genus of tiny globose cacti) and tribe Rhipsalideae (group of epiphytic cacti) (Fig 2a and 2b). This example shows that robust relationships provided by the Cactaceae591 panel can also shed light on the relationship of orphan lineages within subfamily Cactoideae and also may help to understand the evolutionary trend of convergent growth forms across Cactaceae.

This study explored comprehensive genetic dataset information, from a few molecular markers to hundreds of orthologous loci, and across contrasting phylogenetic inferences. In the meantime, this is the first phylogenomic analysis comprising most of the representatives of the tribe Cereeae, resulting in a higher topology resolution. The identification of the same major clades in our inferences and those previously identified in early phylogenetic studies (Ritz et al. 2007, Schlumberger and Renner 2012, Lendel 2013) supports the need for delimitation of these subtribes and genera in Cereeae. Moreover, the well-supported topology obtained here will be a solid basis for exploring the evolutionary and biogeographical patterns within this group.

2.4.1 Major groups of tribe Cereeae s.l.

The monophyly of the tribe Cereeae s.l. and the close relationship with the tribe Notocacteae, as evidenced by the present study, agrees with previous taxonomic (Nyffeler and Eggli 2010) and molecular phylogenetic studies in this group (Nyffeler 2002, Hernández-Hernández et al. 2011). Likewise, the non-monophyly of Cereeae subtribes was previously observed in molecular phylogenetic studies within this group (Ritz et al. 2007, Bárcenas et al.

2011, Schlumberger and Renner 2012, Bombonato et al. 2020, Romeiro-Brito et al. 2022). Although the monophyly of the subtribes Cereinae and Trichocereinae has been previously suggested by molecular phylogenetic studies (Lendel 2013, Fantinati et al. 2021), the genera *Espostoopsis* and *Stetsonia* (putatively representatives of the subtribes Trichocereinae and Rebutinae, respectively) were not included to appropriately assess their circumscription. Here, both genera were sampled and recovered with high support within the subtribe Cereinae, suggesting the inclusion of these taxa in an expanded delimitation of this subtribe. Thereby, the circumscription of the subtribe Trichocereinae should exclude the genus *Espostoopsis*, while the circumscription of the subtribe Rebutiinae should exclude the genus *Stetsonia.*

We identified six well-supported major clades across all phylogenetic inferences, indicating the existence of six subtribes within Cereeae. These subtribes consist of three monogeneric lineages and three clades with high support: *Uebelmannia*, *Aylostera*, *Gymnocalycium*, Rebutiinae s. str. (excluding *Uebelmannia*, *Aylostera*, *Gymnocalycium*, and *Stetsonia*), Trichocereinae s. str. (excluding *Espostoopsis*), and Cereinae s. l (including *Espostoopsis* and *Stetsonia*). As observed in previous phylogenetic studies (Ritz et al. 2007, Demaio et al. 2011, Mosti et al. 2011), the lineages were recovered as well-supported and cohesive groups in all inferences of the present study. Considering that their relationship among the major groups of the tribe Cereeae is now clarified using genome-scale datasets, these subgroups should be finally recognized as independent lineages within Cereeae s.l.

The earliest divergent lineage within Cereeae s.l. remains unknown. Previous phylogenetic studies have pointed to *Uebelmannia* as the first divergent lineage (Ritz et al. 2007, Hernández-Hernández et al. 2011, Lendel 2013, Romeiro-Brito et al. 2022), although these studies did not include representatives from *Aylostera,* or they lacked phylogenetic support. Both issues were also the main concern in the present study. In order to investigate this matter within the tribe Cereeae, both lineages must be included in future phylogenomic studies.

2.4.2 Generic relationships of subtribe Cereinae

To date, the Cactaceae591 dataset includes the most comprehensive sampling of the subtribe Cereinae in phylogenetic studies, including all genera and nearly half of its species diversity. To identify generic relationships and revise taxonomic classifications within this subtribe, it is essential to sample this subtribe broadly in light of its diversity and taxonomic complexity.

The present study corroborates the inclusion of two genera first placed in the subtribes Trichocereinae (*Espostoopsis dybowskii*) and Rebutinae (*Stetsonia coryne*) in subtribe Cereinae (Nyffeler and Eggli 2010). All phylogenetic inferences in this study are in line with previous phylogenetic studies of the tribe Cereeae s.l., which grouped all taxa of subtribe Cereinae in the same clade (Ritz et al. 2007, Schlumpberger and Renner 2012, Lendel 2013, Bombonato et al. 2020, Romeiro-Brito et al. 2022).

The early divergent lineages of the subtribe Cereinae consist of genera with few representatives, such as *Stetsonia* and *Praecereus*, followed by a clade that contains the majority of the subtribe's diversity. The relationships of its major clade are partially congruent with the findings of Fantinati et al. (2021). Both studies confirm the non-monophyly of *Arrojadoa*, *Micranthocereus*, *Pilosocereus*, and *Cipocereus*, and the close relationship between the following genera: 1) *Cereus* and *Cipocereus* s.s., 2) *Arrojadoa* and *Stephanocereus*, and 3) *Micranthocereus* s.str. and *Xiquexique*. On the other hand, our phylogenomic inferences disagreed with the close relationship between *Facheiroa* and *Arrojadoa*, and the non-monophyly of *Cereus*, *Discocactus*, and *Coleocephalocereus*, as recovered by Fantinati et al. (2021).

The controversial relationship between *Cereus* and *Cipocereus* was first documented by phylogenetic inferences with few molecular markers (Franco et al. 2017). Later, the use of genome-scale datasets allied with coalescent phylogenetic inferences was decisive in untangling the relationship between these genera (Fig 2, Bombonato et al. 2020, Taylor et al. in review). Bombonato et al. (2020) reported a high degree of incongruence among the RAD-Seq gene trees and species tree inference in the *Cereus* and *Cipocereus* topology, which was attributed to the rapid and recent divergence leading to an extensive incomplete lineage sorting within this group. Our phylogenetic analysis based on the Cactaceae591 dataset supported the monophyly of *Cereus* and *Cipocereus*. However, we also observed a significant proportion of incongruences among gene trees and species trees associated with these genera, particularly at the ancestral node of *Cereus*.

The *Micranthocereus* s. str. clade comprises *Xiquexique* and *Pilosocereus bohlei*. Although *P. bohlei* shares many characteristics with *Pilosocereus* species, it has long been recovered outside the *Pilosocereus* s. str. group (Calvente et al. 2017, Lavor et al. 2020, Fantinati et al. 2021, Romeiro-Brito et al. 2023). The close relationship of these taxa within the *Micranthocereus* s. str. clade gives insight into common characteristics shared among them, for example, the branching pattern occurring only at the base, hairy areoles, and curved hypanthium (Hunt et al. 2006). *P. bohlei* is somewhat similar to *Xiquexique gounellei*, both

presenting a broader and hairy areole, but sharing the branching pattern of *Micranthocereus* species. The close relationship between *Micranthocereus* s.s., *Xiquexique* and *P. bohlei,* and the latter possibly being an intermediate taxa among this clade is an intriguing result and should be further investigated in future phylogenomic analyses.

Three closely related genera (*Facheiroa*, *Brasilicereus*, and *Leocereus*) were previously placed within the tribes Trichocereeae and Browningeae (Taylor and Zappi 2004) due to the presence of scales and hair-spines in the pericarpel of the flower. Early molecular phylogenies inferences with these genera (Sofiatti 2003) suggested its position within the tribe Cereeae (Sofiatti 2003). In the present study, all genera were recovered in the same clade regardless of the dataset and phylogenetic inference. These results suggest that the similarity of floral traits used to delineate major groups in Cactoideae may at least partially result from convergent evolution, particularly within the BCT clade (Guerrero et al. 2018).

The *Melocactus* clade clustered genera into two distinct groups, one including *Melocactus, Discocactus,* and the other comprising *Coleocephalocereus,* and *Micranthocereus* subg. *Siccobaccatus.* The former group consists of globose plants with apical cephalia, while the latter includes columnar plants with lateral cephalia. The close relationship between the genera *Melocactus* and *Discocactus* has been previously confirmed by molecular phylogenetic studies (Hernández-Hernández et al. 2011, Santos 2013, Silva et al. 2017). However, the present study is the first to mention the close relationship between *M.* subg. *Siccobaccatus* and *Colecephalocereus*. Indeed, the species from *M.* subg. *Siccobaccatus* share characteristics with most *Colecephalocereus* species, including a columnar, erect, and solitary growth habit, sunken lateral cephalia, and association with rock outcrops. Based on these findings, we suggested the inclusion of the former to be recognized as a new subgenus of *Coleocephalocereus*.

Our findings support the monophyly of *Pilosocereus*, excluding *Pilosocereus bohlei*, a taxon previously allocated outside the genus *Pilosocereus* (Calvente et al. 2016, Lavor et al. 2020, Romeiro-Brito et al. 2023). We recovered incongruences among different phylogenetic inference approaches within the genus *Pilosocereus*, a pattern that was previously observed in this group using multilocus dataset (Romeiro-Brito et al. 2023), which indicates a complex scenario of diversification within this group. Nonetheless, our results recovered a similar topology obtained in previous phylogenetic studies (Lavor et al. 2020, Romeiro-Brito et al. 2023), but with higher node support within this group (Fig 4). These results emphasize the importance of a genome-scale dataset to obtain a comprehensive understanding of the evolutionary history of *Pilosocereus*.

The well-supported *Arrojadoa* clade is composed of *Arrojadoa*, *Stephanocereus, Cipocereus pussiliflorus*, and *Micranthocereus violaciflorus.* The early divergent lineages of *Arrojadoa* consist of *Cipocereus pussiliflorus*, and *Micranthocereus violaciflorus,* both species lacking cephalia. The following clade comprises *Arrojadoa* and *Stephanocereus* species, which is splitted into two groups: one composed of species bearing ring-cephalia (most *Arrojadoa* species and *Stephanocereus leucostele*) and other lacking or with primitive cephalia (*Arrojadoa bahiensis* and *Stephanocereus luetzelburgii*). Previous phylogenetic studies on this group (Soffiati 2003 and Fantinati et al. 2021) have suggested a close relationship between *Cipocereus pusilliflorus* and *Stephanocereus* to *Arrojadoa* species, but the positioning of *Micranthocereus violaciflorus* near *Arrojadoa* is a novelty. These results highlight the need for recognizing an expanded *Arrojadoa* that includes *Cipocereus pusilliflorus, Micranthocereus violaciflorus* and *Stephanocereus* species.

2.4.3 Generic relationships of subtribe Trichocereinae

Though the non-monophyly of subtribes is widespread in Cereeae s.l. (Guerrero et al. 2018), many phylogenetic studies using few molecular markers have recovered Trichocereinae as a monophyletic subtribe (Hernández-Hernández et al. 2011, Schlumpberger and Renner 2012, Lendel 2013). Here we highlight an updated circumscription of subtribe Trichocereinae, excluding *Espostoopsis dyboswkii*. Although this taxon resembles *Espostoa,* suggesting a close relationship with the subtribe Trichocereinae, its naked flower tube corroborates its inclusion in the subtribe Cereinae (Taylor and Zappi 2004). To this degree, flower tube characteristics are also variable within subtribes of Cereeae and should be treated with caution in taxonomic classifications of these groups.

According to gene-scale dataset phylogenetic inferences, the genus *Reicheocactus* was the earliest lineage to diverge within the subtribe Trichocereinae, followed by the genus *Arthrocereus*. We were unable to confirm this relationship using the genome-scale datasets because one of those genera was missing in each dataset. Nonetheless, the phylogenetic inferences with Angiosperm353 and Cactaceae591 datasets recovered *Reicheocactus* and *Arthrocereus* as early divergent lineages in subtribe Trichocereinae, respectively, somehow corroborating gene-scale phylogenetic results. The next divergent clade is commonly recovered by the genomic dataset, placing the genus *Harrisia* as a sister group to all the remaining representatives of Trichocereinae. The genus *Harrisia* was consistently identified as closely related to either the *Cleistocactus* clade (Franck et al. 2013) or the *Echinopsis* s.s. clade

(Schlumberger and Renner 2012, Lendel 2013). However, the present study suggests that *Harrisia* may not be closely related to either of these clades. The inclusion of *Reicheocactus*, *Arthrocereus*, and *Harrisia* in future phylogenomic studies may shed light on the *Reicheocactus* as the first divergent group of Trichocereinae and a closer relationship between *Arthrocereus* and *Harrisia.* The latter two genera are the only members of the subtribe Trichocereinae occurring outside the Andean region, spanning across Western and Eastern Brazil (*Arthrocereus*) to the Caribbean region (*Harrisia*).

Echinopsis was recovered as a polyphyletic group in our gene-scale phylogenetic tree, corroborating previous phylogenetic studies (Schlumpberger and Renner 2012, Lendel 2013). However, all trees inferred from Cactaceae591 and Angiosperm353 datasets recovered the monophyly of this genus, including *Denmoza* and excluding *Reicheocatus*. The recent taxonomic checklist from Korotkova et al. (2021) segregated *Echinopsis* into multiple genera. We recommend the use of genome-scale datasets with a comprehensive taxonomic sampling of this group in order to test its monophyly as once defined by Hunt et al. (2006).

The remaining complex groups of Trichocereinae mostly agree with previous plastid phylogenetic inferences, consisting of the *Cleistocactus* clade and *Oreocereus* clade (Schlumberger and Renner 2012, Lendel 2013). The *Cleistocactus* clade includes species pollinated by bats (*Samaipaticereus* and *Yungasocereus*), hummingbirds (*Cleistocactus* and *Cephalocleistocactus*), or both of them (*Vatricania guenteri*). The *Oreocereus* clade, as recovered from genomic datasets, comprises the genera *Borzicactus*, *Espostoa*, *Mila*, *Oreocereus*, *Oroya*, *Haageocereus*, *Matucana,* and *Lasiocereus*. The position of *Lasiocereus* among Trichocereinae in the Angiosperm353 tree should be taken with caution, considering that previous phylogenetic studies placed *Lasiocereus* species within the subtribe Rebutinae (Fig S3, Ritz et al. 2007, Schlumpberger and Renner 2012, Lendel 2013).

2.4.4 Concatenated versus coalescent inference approaches in face of extensive gene tree and species tree discordance

Resolving the phylogenetic relationships among representatives of the subtribe Cereinae has been a persistent challenge for many decades (Taylor and Zappi 1989), primarily due to the absence of apomorphic characteristics (Taylor and Zappi 2004) and the low phylogenetic resolution (e.g. Franco et al. 2017, Calvente et al. 2017, Fantinati et al. 2021). The use of genome-scale datasets to resolve relationships at multiple levels is not only a trending practice in the studies of Cactaceae groups that underwent rapid and recent divergence

(Franco et al. 2022), but a mandatory approach to investigating evolutionary and diversification histories of its tribes. Moreover, the extensive level of gene tree-species tree discordance in Cactaceae also highlighted the usefulness of coalescent inference approaches in this group, which may overperform concatenated inference approaches on contentious nodes (Bombonato et al. 2020; Romeiro-Brito et al. 2022; Romeiro-Brito et al. 2023).

Despite the extensive gene tree discordance, the Cereeae topology is stable using different genomic resources and phylogenetic approaches. Some minor discordances at terminal branches were found among concatenated and coalescent inferences, including *Cereus, Pilosocereus* and *Melocactus* clade in subtribe Cereinae, and *Cleistocactus* clade in subtribe Trichocereinae. The incomplete lineage sorting has been discussed as the main source of gene-tree and species tree discordance in cacti (Copetti et al. 2017, Wang et al. 2018, Romeiro-Brito et al. 2022, Romeiro-Brito et al. 2023), though its rapid diversification may also lead to short internal branches generating an 'anomaly zone' (Bombonato et al. 2020). The effects of hybridization events have not been thoroughly investigated in tribe Cereeae s.l. (e.g. Perez et al. 2016, Khan et al. 2020), especially the ancient hybridization events. The expansion of aridity during the Miocene and Pliocene (Arakaki et al. 2011) may have promoted the rapid diversification within this group, but would also have led to the connection and hybridization events at the generic level. Investigating other sources of discordance, like past ecological dynamics and ancient hybridization events, may help to understand the diversification of Cereeae and give a hint on the controversial and low resolution relationships remaining within this group.

2.5 Taxonomic synopsis of Cereeae

Species are listed in cases where the circumscription of the genus or subgenus is being changed significantly from that in the standard work by Hunt et al. (2006, 2013). Newly published names are indicated in **bold** typeface.

Tribe Cereeae Salm-Dyck (as 'Cereastreae')

Superficially like Pachycereeae, but pericarpel, hypanthial tube, and pericarp of unripe fruit usually lacking areoles and stiff spines (if spiny, then plant very slender, not pachycaul). Type: *Cereus* Mill. Comprising the following six subtribes:

1. Subtribe **Uebelmanniinae** N.P.Taylor, **subtr. nov.**

Globular to shortly columnar, many ribbed / tuberculate-ribbed cacti. Stems unbranched, with internal mucilage ducts; epidermis roughened, bearing waxy scales (cf. *Copiapoa*), grey-green to reddish; areoles on mature stems with long hairs and short straight spines. Cephalium lacking. Flowers small, apical, diurnal, pericarpel and very short hypanthial tube with bract-scales woolly in their axils, perianth-segments expanding, green to yellow, stigma-lobes few. Fruit scarcely fleshy, more or less naked but with the hairy perianth remains attached at apex, reddish. Seeds few, medium-sized, testa smooth.

Type and only genus: *Uebelmannia* Buining (3 spp.).

Distribution:

2. Subtribe **Aylosterinae** N.P.Taylor, **subtr. nov.**

Dwarf globular, simple or caespitose, sometimes semi-geophytic; stems with ± spiralled tubercles; spines short, but often dense; cephalium absent. Flowers diurnal, from the sides or base of the stem, shortly funnelform, pericarpel and tube with bract-scales, the style and tube ± fused in the lower half or more. Fruits and seeds small.

Type and only genus: *Aylostera* Spegazzini (incl. *Mediolobivia* Backeb., *Digitorebutia* Donald) (c. 11 spp.).

3. Subtribe Rebutiinae Donald (incl. Browningieae F.Buxb., Krainz, Die Kakteen, Lfg 33: CIV/1 (1966)).

Tall columnar, ribbed pachycaul cacti, branched and treelike, rarely single-stemmed, or dwarf globular-stemmed and simple or clustering (caespitose). Stems 7–many-ribbed or with spiralled tubercles, when ribbed these mostly low and rounded. Areoles spiny on juvenile plants, later sometimes spineless on fertile stems. Cephalium lacking (but cf. *Browningia columnaris* F.Ritter). Flowers borne laterally to basally, nocturnal and whitish or diurnal and brightly coloured, shortly tubular to infundibuliform, pericarpel and hypanthial tube bearing small to large and often overlapping bract-scales, sometimes these shortly hairy, persisting on the fleshy indehiscent fruit. Seeds small, very numerous, testa variously ornamented.

Type: *Rebutia* K.Schum. (3 spp.)

Browningia Britt. & Rose (8 spp.)

Weingartia Werderm. (incl. *Sulcorebutia* Backeb., *Cintia* Knize & Riha) (c. 16 spp.).

4. Subtribe **Gymnocalyciinae** N.P.Taylor, **subtr. nov.**

Dwarf to medium-sized globular to discoid cacti; ribs few to many, mostly low, often tuberculate. Areoles spiny or sometimes almost spineless. Cephalium lacking. Flowers diurnal, from near the stem apex or from the ‘shoulder’ of the stem, pericarpel and short-to-long

hypanthial tube scaly but otherwise naked, perianth variously coloured. Fruit dehiscent, revealing the funicular pulp in which the seeds of diverse testa morphology are embedded.

Type and only genus: *Gymnocalycium* Mittler (c. 65 spp.)

5. Subtribe Trichocereinae F.Buxb.

Stems of diverse size and form, from depressed globose to tall columnar, unbranched to treelike, many-ribbed. Areoles usually spiny. Cephalium lacking or occasionally present, then lateral. Flowers of diverse size and shape, from small to very large, nocturnal or diurnal, pericarpel and hypanthial tube clothed in discrete scales bearing abundant hair spines (woolly hairs) in their axils, perianth expanded or segments remaining erect to incurved; stamens often inserted in the tube in two series; stigma-lobes many. Fruits scaly-hairy, mostly dehiscent to reveal the funicular pulp. Seeds medium-sized, testa smooth to tuberculate.

Type: *Trichocereus* Britt. & Rose (= *Echinopsis* Zucc.)

Genera:

Mila Britt. & Rose

Pygmaeocereus Johns. & Backeb.

Haageocereus Backeb.

Espostoa Britt. & Rose (incl. *Thrixanthocereus* Backeb.)

Rauhocereus Backeb.

Weberbauerocereus Backeb.

Cleistocactus Lem. (incl. *Samaipaticereus* Cárd., *Yungasocereus* F.Ritter, *Vatricania* Backeb.)

Borzicactus Riccob.

Oreocereus (A.Berger) Riccob.

Matucana Britt. & Rose

Oroya Britt. & Rose

Harrisia Britt.

Arthrocereus A.Berger

Echinopsis Zucc. (incl. *Denmoza* Britt. & Rose)

Reicheocactus Backeb.

Incertae sedis: *Lasiocereus* F.Ritter (cf. *Espostoa*).

6. Subtribe Cereinae Britt. & Rose

Habit diverse as in Trichocereinae. Lateral or terminal cephalia developed in many taxa. Flowers of diverse size and shape, from small to very large, but sometimes very small (e.g.

Melocactus), mostly lacking areoles and hairs, bract-scales often inconspicuous or widely spaced. Fruits scaly or more often naked, dehiscent or indehiscent; seeds mostly small.

Type: *Cereus* Mill.

Genera:

Stetsonia Britt. & Rose (1 sp.)

Praecereus F.Buxb. (2 spp.)

Cipocereus F.Ritter (5 spp.) Type:

1. *C. pleurocarpus* F.Ritter

2. *C. minensis* (Werderm.) F.Ritter;

3. *C. bradei* (Backeb. & Voll) Zappi & N.P.Taylor

4. *C. laniflorus* N.P.Taylor & Zappi

5. *C. crassisepalus* (Buin. & Brederoo) Zappi & N.P.Taylor

Cereus Mill. Type: *C. hexagonus* (L.) Mill. (c. 33 spp.)

C. subg. *Oblongicarpi* (Croizat) D.R.Hunt & N.P.Taylor. (5 spp.) Type:

1. *C. repandus* (L.) Mill.;

2. *C. fricii* Backeb.

3. *C. horrispinus* Backeb.

4. *C. mortensenii* Croizat.

5. *Cereus serruliflorus* Haw (*C. haitiensis* A.R.Franck, nom. illeg.).

C. subg. *Mirabella* (F.Ritter) N.P.Taylor (incl. *Mirabella* F.Ritter, *Estevesia* P.J.Braun).

Type:

6. *C. mirabella* N.P.Taylor (*Mirabella minensis* F.Ritter);

7. *C. albicaulis* (Britt. & Rose) Luetzelb.

Incertae sedis:

8. *C. saddianus* (Rizzini & Matos-F.) P.J.Braun

9. *C. phatnospermus* K.Schum. (incl. *C. kroenleinii* N.P.Taylor)

10. *C. aethiops* Haw.

11. *C. adelmarii* (Rizzini & Mattos-F.) P.J.Braun

12. *C. estevesii* P.J.Braun (cf. *C. albicaulis*).

C. subg. *Cereus* (incl. *C.* subg. *Ebneria* (Backeb.) D.Hunt)

13. *C. spegazzinii* F.A.C.Weber

14. *C. vargasianus* Cárd.

15. *C. trigonodendron* Ule

16. *C. pierre-braunianu*s E.Esteves-Pereira

17. *C. calcirupicola* F.Ritter

18. *C. lepidotus* Salm-Dyck

19. *C. gerardi* N.P.Taylor

20. *C. jamacaru* DC.

21. *C. ingens* N.P.Taylor & M.C.Machado

22. *C. sericifer* (F.Ritter) P.J.Braun

23. *C. pentagonus* (L.) Haw. (*C. fernambucensis* Lem.)

24. *C. insularis* Hemsl.

25. *C. bicolor* Rizz. & Mattos-F.

26. *C. hildmannianus* K.Schum.

27. *C. stenogonus* K.Schum.

Incertae sedis:

28. *C. cochabambensis* Cárd.

29. *C. huilunchu* Cárd.

30. *C. hankeanus* K.Schum.

31. *C. lamprospermus* K.Schum.

32. *C. lanosus* (F.Ritter) P.J.Braun

33. *C. braunii* Cárd.

Micranthocereus Backeb. (incl. *Caerulocereus* Guiggi). (7 spp.) Type:

1. *M. polyanthus* (Werderm.) Backeb.;

2. *M. flaviflorus* Buin. & Brederoo

3. *M. alvinii* (Hofacker & M.C.Machado) N.P.Taylor & M.Lowry

4. *M. streckeri* Van Heek. & Van Criek.

5. *M. hofackerianus* (P.J.Braun & E.Esteves-Pereira) M.C.Machado

6. *M. purpureus* (Gürke) F.Ritter.

Xiquexique Lavor & Calvente (*Pilosocereus* subg. *Gounellea* Zappi). (3 spp.) Type:

1. *X. gounellei* (F.A.C.Weber) Lavor & Calvente;

2.***Xiquexique bohlei*** (Hofacker) N.P.Taylor, **comb. nov.** Basionym: *Pilosocereus bohlei* Hofacker, Kakt. and. Sukk. 52: 253–257 (2001).

3. X. tuberculatus (Werderm.) Lavor & Calvente

4. *X. frewenii* (Zappi & N.P.Taylor) Lavor & Calvente.

Arrojadoa Britt. & Rose (incl. *Stephanocereus* A.Berger, *Floribunda* F.Ritter, *Pierrebraunia* E.Esteves-Pereira, *Arrojadoopsis* Guiggi). (13 spp.)

Revised description: dwarf to medium-tall (to 4m) cylindrical-stemmed cacti, erect to decumbent, shrub like or sometimes solitary columnar; subterranean stem base and rootstock often tuberous, vascular cylinder woody. Stems 7-many ribbed, ribs low, rounded, never acute, axes sometimes segmented, and then interrupted by ring cephalia. Areoles always bearing long hairs, at least when young, always spiny. Flowers mostly from or from near the apex of stem-segments, often from terminal bristly cephalia, small to medium-sized (2–10cm), shortly tubular, pericarpel very small, it and hypanthial tube almost naked or with few inconspicuous bract-scales, with relatively small, mostly scarcely expanded perianth-segments, diurnal or nocturnal, humming-bird or bat syndrome, reddish pink to magenta, or bicoloured with paler to whitish inner segments, or greenish white. Fruits mostly indehiscent or opening by a basal pore, fleshy, never dry when ripe, variously coloured, perianth remains persistent, blackish. Seeds mostly small, 1–2 mm. Seedlings, where known, globular at first, only later becoming elongate-cylindric. Type:

1. *A. rhodantha* (Gürke) Britt. & Rose;

2. ***Arrojadoa pusilliflora*** (F.Ritter) N.P.Taylor, **comb. nov.** Basionym: *Floribunda pusilliflora* F.Ritter, Kakt. Südamer. 1: 58–60 (1979).

3. ***Arrojadoa violaciflora*** (Buining & Brederoo) N.P.Taylor, **comb. nov.** Basionym: *Micranthocereus violaciflorus* Buin., Kakt. and. Sukk. 20: 129–130 (1969).

4. *A. bahiensis* (P.J.Braun & E.Esteves-Pereira) N.P.Taylor & Eggli

5. ***Arrojadoa luetzelburgii*** (Vaupel) N.P.Taylor, **comb. nov.** Basionym: *Cereus luetzelburgii* Vaupel, Zeitschr. Sukkulentenk. 1: 57 (1923).

6. *A. marylaniae* Soares-F. & M.C.Machado

7. ***Arrojadoa leucostele*** (Guerke) N.P.Taylor, **comb. nov.** Basionym: *Cereus leucostele* Gürke, Monatsschr. Kakt.-Kunde 18: 53 (1908).

8. *A. dinae* Buin. & Brederoo

9. *A. albiflora* Buin. & Brederoo

10. *A. eriocaulis* Buin. & Brederoo

11. *A. multiflora* F.Ritter

12. *A. olsthoorniana* Hofacker & M.C.Machado

13. *A. penicillata* (Gürke) Britt. & Rose.

Facheiroa Britt. & Rose (incl. *Zehntnerella* Britt. & Rose, *Leocereus* Britt. & Rose, *Brasilicereus* Backeb., *Bragaia* P.J.Braun). (7 spp.)

Revised description: medium to tall cylindric cacti, erect and self-supporting or slender and sometimes leaning on surrounding vegetation, shrubby to treelike, sparsely to many-branched,

vascular cylinder woody. Stems unsegmented, 7-many-ribbed, ribs low, rounded, never acute. Areoles always spiny, mostly lacking long-hairs. Flowers lateral, never terminal, sometimes borne from a bristly/woolly lateral cephalium, shortly tubular, to 7.5×7.5 cm, pericarpel and hypanthial tube scaly, woolly or spiny, never naked, perianth small and hardly expanded or as broad as the flower is long, diurnal to nocturnal, bird, bat or moth syndrome, reddish, green or white. Fruit indehiscent or disintegrating when mature, scaly or with deciduous spiny areoles, fleshy. Seeds small to medium, 1–2.5 mm. Seedlings, where known, globular at first, later elongating. Type:

1. *F. ulei* (Gürke) Werderm.

2. *F. cephaliomelana* Buining & Brederoo

3. *F. squamosa* (Gürke) P.J.Braun & E.Esteves-Pereira

4. ***Facheiroa phaeacantha*** (Gürke) N.P.Taylor, **comb. nov.** Basionym: *Cereus phaeacanthus* Gürke, Monatsschr. Kakt.-Kunde 18: 57 (1908).

5. ***Facheiroa markgrafii*** (Backeb. & Voll) N.P.Taylor, **comb. nov**. Basionym: *Brasilicereus markgrafii* Backeb. & Voll, Arch. Jard. Bot. Rio de Janeiro 9: 155 (1949, publ. 1950).

6. ***Facheiroa bragaia*** N.P.Taylor, **nom. nov.** Replaced synonym: *Bragaia estevesii* Hofacker & P.J.Braun, Kakt. and. Sukk. 60(12): 328 (2009), non *Facheiroa estevesii* P.J.Braun (= *F. cephaliomena* subsp. *estevesii* (P.J.Braun) N.P.Taylor & Zappi).

7. ***Facheiroa bahiensis*** (Britt. & Rose) N.P.Taylor, **comb. nov.** Basionym: *Leocereus bahiensis* Britt. & Rose, Cact. 2: 108 (1920).

Melocactus Link & Otto, nom. cons. (c. 40 spp.) Type: *Cactus melocactus* L., typ. cons.

Discocactus Pfeiff. (13 spp.) Type: D. insignis Pfeiff. (= *D. placentiformis* (Lehm.) K.Schum.).

Coleocephalocereus Backeb. (incl. *Buiningia* F.Buxb., *Siccobaccatus* P.J.Braun & Esteves-Pereira, *Mariottia* Guiggi). (9 spp.) Type: *C. fluminensis* (Miq.) Backeb.

C. subg. *Coleocephalocereus*

1. *C. fluminensis* (Miq.) Backeb.

2. *C. decumbens* F.Ritter

3. *C. pluricostatus* Buin. & Brederoo

4. *C. buxbaumianus* Buin.

C. subg. *Simplex* N.P.Taylor. Type and only species:

5. *C. goebelianus* (Vaupel) Buin.

C. subg. *Buiningia* (F.Buxb.) P.J.Braun. Type: *Buiningia* brevicylindrica Buin. (= *C. aureus* F.Ritter)

6. *C. aureus* F.Ritter

7. *C. purpureus* (Buin. & Brederoo) F.Ritter.

Coleocephalocereus subg. ***Siccobaccatus*** (P.J.Braun & E.Esteves-Pereira) N.P.Taylor, **comb. nov.** Basionym: *Siccobaccatus* P.J.Braun & E.Esteves-Pereira, Succulenta (NL) 69: 7 (1990). (2 spp.) Type: *S. dolichospermaticus* (Buin. & Brederoo) P.J.Braun & E.Esteves-Pereira.

Differs from other subgenera of Coleocephalocereus in fruits which are dry and disintegrate at maturity; seeds elongate, testa ± smooth (wind dispersed).

8. ***Coleocephalocereus dolichospermaticus*** (Buin. & Brederoo) N.P.Taylor, *comb. nov.* Basionym: *Austrocephalocereus dolichospermaticus* Buin. & Brederoo, Kakt. and. Sukk. 25: 76–79 (1974).

9. ***Coleocephalocereus neoestevesii*** N.P.Taylor, **nom. nov.** Replaced synonym: *Austrocephalocereus estevesii* Buin. & Brederoo, Cact. Succ. J. (US) 47: 267 (1975), non *Coleocephalocereus estevesii* L.Diers (= *C. buxbaumianus* subsp. *flavisetus* (F.Ritter) N.P.Taylor & Zappi).

Pilosocereus Byles & G.Rowley (excl. *P.* subg. *Gounellea* Zappi, *Caerulocereus* Guiggi). (c. 60 spp.) Type: *P. leucocephalus* (Poselger) Byles & G.Rowley.

Incertae sedis within Cereinae:

1. *Micranthocereus albicephalus* (Buin. & Brederoo) F.Ritter

2. *M. auriazureus* Buin. & Brederoo.

2.6 Acknowledgments

We thank Gerardus Olsthoorn, Dr. Lidyanne Y. S. Aona, Dr. Diego R. Gonzaga, and the Coleção de Cactos e Suculentas do Instituto de Pesquisas Jardim Botânico do Rio de Janeiro for contributing to the cactus samples used in this study. We thank Dr. Juliana F. Martinez and Msc. Heidi S. M. Utsunomiya for technical support. We also thank Dr. Matias Köhler and Dr. Martin Lowry for sharing photos from representatives of tribe Cereeae and Dr. Danilo T. Amaral and Dr. Isabel A. S. Bonatelli for valuable comments and discussions that improved this manuscript.

2.7 Funding

This study was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP: 2018/03428-5, 2019/03211-9). M.R.B. and M.C.T. were granted by FAPESP fellowships (2018/06937-8 and 2019/11233-2).

2.8 References

Acha S. and Majure, L. C. (2022). A New Approach Using Targeted Sequence Capture for Phylogenomic Studies across Cactaceae. Genes, 13, 350.

Amaral D. T., Bombonato J. R., da Silva Andrade S. C., Moraes E. M. and Franco F. F. (2021a). The genome of a thorny species: Comparative genomic analysis among South and North American Cactaceae. Planta, 254, 1-7.

Amaral D. T., Minhós-Yano I., Oliveira J. V. M., Romeiro-Brito M., Bonatelli I. A. S., Taylor N. P. et al. (2021b). Tracking the xeric biomes of South America: The spatiotemporal diversification of Mandacaru cactus. Journal of Biogeography, 48, 3085-3103.

Amaral D. T., Bonatelli I. A., Romeiro-Brito M., Moraes E. M. and Franco, F. F. (2022). Spatial patterns of evolutionary diversity in Cactaceae show low ecological representation within protected areas. Biological Conservation, 273, 109677.

Anderson E. F. 2001 .The cactus family. Timber Press, Portland, Oregon, USA.

Applequist W. L. and Wallace, R. S. (2002). Deletions in the plastid trnT–trnL intergenic spacer define clades within Cactaceae subfamily Cactoideae. Plant Systematics and Evolution, 231, 153-162.

Baker W. J., Bailey P., Barber V., Barker A., Bellot S., Bishop D. et al. (2022). A comprehensive phylogenomic platform for exploring the angiosperm tree of life. Systematic Biology, 71, 301-319.

Bárcenas R. T., Yesson C. and Hawkins, J. A. (2011). Molecular systematics of the Cactaceae. Cladistics, 27, 470-489.

Barthlott W. and D. Hunt. 1993. Cactaceae. Pp. 161–197 in The Families and Genera of Vascular Plants,vol.2, eds.K.Kubitzki, J. G. Rohwer, and V. Bittrich. Berlin: Springer-Verlag

Bombonato J. R., Amaral D. T., Silva G. A. R., Khan G., Moraes E. M., da Silva Andrade S. C. et al (2020). The potential of genome-wide RAD sequences for resolving rapid radiations: A case study in Cactaceae. Molecular Phylogenetics and Evolution, 151, 106896.

Borowiec M.L. 2016. AMAS: A fast tool for alignment manipulation and computing of summary statistics. PeerJ, 25, 1972–1973.

Breslin P. B., Wojciechowski M. F. and Majure, L. C. (2021). Molecular phylogeny of the Mammilloid clade (Cactaceae) resolves the monophyly of *Mammillaria*. Taxon, 70, 308-323.

Brown J. W., Walker J. F. and Smith, S. A. (2017). Phyx: phylogenetic tools for unix. *Bioinformatics*, *33*, 1886-1888.

Calvente A., Moraes E. M., Lavor P., Bonatelli I. A., Nacaguma P., Versieux L. M. et al. (2017). Phylogenetic analyses of *Pilosocereus* (Cactaceae) inferred from plastid and nuclear sequences. Botanical Journal of the Linnean Society, 183, 25-38.

Capella-Gutiérrez S., Silla-Martínez J. M. and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics, 25, 1972-1973.

Chen C., Li F., Xie F., Chen J., Hua Q., Chen J. et al. (2022). Pitaya Genome and Multiomics Database (PGMD): A Comprehensive and Integrative Resource of Selenicereus undatus. Genes, 13, 745.

Copetti D., Búrquez A., Bustamante E., Charboneau J. L., Childs K. L., Eguiarte L. E. et al. (2017). Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. Proceedings of the National Academy of Sciences, 114, 12003-12008.

Demaio P. H., Barfuss M. H., Kiesling R., Till W. and Chiapella, J. O. (2011). Molecular phylogeny of *Gymnocalycium* (Cactaceae): assessment of alternative infrageneric systems, a new subgenus, and trends in the evolution of the genus. American Journal of Botany, 98, 1841-1854.

Fantinati M. R., Soffiatti P. and Calvente, A. (2021). A new phylogenetic hypothesis for Cereinae (Cactaceae) points to a monophyletic subtribe. Systematic Botany, 46, 689-699.

Franck A. R., Cochrane B. J. and Garey J. R. (2013). Phylogeny, biogeography, and infrageneric classification of *Harrisia* (Cactaceae). Systematic Botany, 38, 210-223

Franco F. F., Silva G. A. R., Moraes E. M., Taylor N., Zappi D. C., Jojima C. L. and Machado, M. C. (2017). Plio-Pleistocene diversification of *Cereus* (Cactaceae, Cereeae) and closely allied genera. Botanical Journal of the Linnean Society, 183, 199-210.

Franco F. F., Amaral D. T., Bonatelli I. A., Romeiro-Brito M., Telhe M. C. and Moraes, E. M. (2022). Evolutionary genetics of cacti: research biases, advances and prospects. Genes, 13, 452.

Guerrero P. C., Majure L. C., Cornejo-Romero A. and Hernández-Hernández, T. (2018). Phylogenetic relationships and evolutionary trends in the cactus family. Journal of Heredity, 110, 4-21.

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. Systematic Biology, 59, 307-321.

Hernández-Hernández T., Hernández H. M., De-Nova J. A., Puente R., Eguiarte L. E. and Magallón S. (2011). Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). American Journal of Botany, 98, 44-61.

Hernández-Hernández 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 phytologist, 202, 1382-1397.

Hernández-Ledesma P., Berendsohn W. G., Borsch T., Von Mering S., Akhani H., Arias S. et al. (2015). A taxonomic backbone for the global synthesis of species diversity in the angiosperm order Caryophyllales. Willdenowia, 45, 281-383.

Hoang D. T., Chernomor O., Von Haeseler A., Minh B. Q. and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Molecular Biology and Evolution, 35, 518-522.

Hunt D. , N. P. Taylor and G. Charles . 2006 . The new cactus lexcon. dh Books, Milborne Port, UK.

Inglis P. W., Pappas M. D. C. R., Resende L. V. and Grattapaglia, D. (2018). Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. PloS one, 13, e0206085.

Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J. et al. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Applications in plant sciences, 4, 1600016.

Kalyaanamoorthy S., Minh B. Q., Wong T. K., Von Haeseler A. and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nature methods, 14, 587-589.

Katoh, K. and Standley D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution, 30, 772-780.

Köhler M., Reginato M., Souza-Chies T. T. and Majure L. C. (2020). Insights into chloroplast genome evolution across Opuntioideae (Cactaceae) reveals robust yet sometimes conflicting phylogenetic topologies. Frontiers in Plant Science, 11, 729.

Korotkova N., Aquino D., Arias S., Eggli U., Franck A., Gómez-Hinostrosa C. et al. (2021). Cactaceae at Caryophyllales. org–a dynamic online species-level taxonomic backbone for the family. Willdenowia, 51, 251-270.

Lagomarsino L. P., Frankel L., Uribe-Convers S., Antonelli A. and Muchhala, N. (2022). Increased resolution in the face of conflict: phylogenomics of the Neotropical bellflowers (Campanulaceae: Lobelioideae), a rapid plant radiation. Annals of Botany, 129, 723-736.

Lendel A. (2013). South American Cacti in time and space: studies on the diversification of the tribe Cereeae, with particular focus on subtribe Trichocereinae (Cactaceae). PhD Thesis, University of Zurich, Switzerland.

Majure L. C., Baker M. A., Cloud-Hughes M., Salywon A. and Neubig K. M. (2019). Phylogenomics in Cactaceae: A case study using the chollas sensu lato (Cylindropuntieae, Opuntioideae) reveals a common pattern out of the Chihuahuan and Sonoran deserts. American Journal of Botany, 106, 1327-1345.

Majure L. C., Barrios D., Díaz E., Zumwalde B. A., Testo W. and Negrón-Ortíz, V. (2021). Pleistocene aridification underlies the evolutionary history of the Caribbean endemic, insular, giant *Consolea* (Opuntioideae). American Journal of Botany, 108, 200-215.

Majure L. C., Barrios D., Díaz E., Bacci L. F. and Piñeyro, Y. E. (2022). Phylogenomics of the Caribbean melocacti: Cryptic species and multiple invasions. Taxon, 71, 993-1012.

Majure L. C., Achá S., Baker M. A., Puente-Martínez R., Köhler M. and Fehlberg, S. (2023). Phylogenomics of One of the World's Most Intriguing Groups of CAM Plants, the Opuntioids (Opuntioideae: Cactaceae): Adaptation to Tropical Dry Forests Helped Drive Prominent Morphological Features in the Clade. Diversity, 15, 570.

Martínez-González C. R., Ramírez-Mendoza R., Jiménez-Ramírez J., Gallegos-Vázquez C. and Luna-Vega, I. (2017). Improved method for genomic DNA extraction for *Opuntia* Mill.(Cactaceae). Plant Methods, 13, 1-10.

Merklinger F. F., Böhnert T., Arakaki M., Weigend M., Quandt D. and Luebert, F. (2021). Quaternary diversification of a columnar cactus in the driest place on earth. American Journal of Botany, 108, 184-199.

Minh B. Q., Schmidt H. A., Chernomor O., Schrempf D., Woodhams M. D., Von Haeseler A. and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution, 37, 1530-1534.

Mongiardino Koch N. (2021). Phylogenomic subsampling and the search for phylogenetically reliable loci. Molecular Biology and Evolution, 38, 4025-4038.

Mosti S., Bandara N. L. and Papini, A. (2011). Further insights and new combinations in *Aylostera* (Cactaceae) based on molecular and morphological data. Pakistan Journal of Botany, 43, 2769-2785.

Nyffeler R. (2002). Phylogenetic relationships in the cactus family (Cactaceae) based on evidence from trnK/matK and trnL-trnF sequences. American Journal of Botany, 89, 312-326.

Nyffeler R. and Eggli, U. (2010). A farewell to dated ideas and concepts: molecular phylogenetics and a revised suprageneric classification of the family Cactaceae. Schumannia, 6, 109-149.

Braun, P. J. (1988). On the taxonomy of Brazilian Cereeae (Cactaceae). Bradleya, 1988, 85-99.

Pillet M., Goettsch B., Merow C., Maitner B., Feng X., Roehrdanz P. R. and Enquist, B. J. (2022). Elevated extinction risk of cacti under climate change. Nature plants, 8, 366-372.

Portik, D. M. and Wiens, J. J. (2020). SuperCRUNCH: A bioinformatics toolkit for creating and manipulating supermatrices and other large phylogenetic datasets. Methods in Ecology and Evolution, 11, 763-772.

Ritz C. M., Martins L., Mecklenburg R., Goremykin V. and Hellwig F. H. (2007). The molecular phylogeny of *Rebutia* (Cactaceae) and its allies demonstrates the influence of paleogeography on the evolution of South American mountain cacti. American Journal of Botany, 94, 1321-1332.

Ritz C. M., Fickenscher K., Föller J., Herrmann K., Mecklenburg R., Wahl, R. (2016). Molecular phylogenetic relationships of the Andean genus Aylostera Speg.(Cactaceae, Trichocereeae), a new classification and a morphological identification key. Plant Systematics and Evolution, 302, 763-780.

Romeiro-Brito M., Telhe M. C., Amaral D. T., Franco F. F. and Moraes E. M. (2022). A target Capture Probe Set Useful for Deep-and Shallow-Level Phylogenetic Studies in Cactaceae. Genes, 13, 707.

Santos, M. R. 2013. Filogenia Molecular, Taxonomia, Biogeografia e Conservacao de Discocactus Pfeiff. Ph.D. thesis. Vicosa, Minas Gerais: Universidade Federal de Vicosa

Sayyari, E.and Mirarab, S. (2016). Fast coalescent-based computation of local branch support from quartet frequencies. Molecular Biology and Evolution, 33, 1654-1668.

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. American Journal of Botany, 99(8), 1335-1349.

Schubert M., Lindgreen S. and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Research Notes, 9, 1-7.

Smith S. A., Moore M. J., Brown J. W. and Yang, Y. (2015). Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. BMC Evolutionary Biology, 15, 1-15.

Soffiatti P. 2003. Estudos Anatâmicos e Filogenéticos na Tribo Cereeae, Cactaceae.Ph.D. thesis. São Paulo: Universidade de São Paulo.

Taylor N. P. and Zappi, D. C. (1989). An alternative view of generic delimitation and relationships in tribe Cereeae (Cactaceae). Bradleya, 1989, 13-40.

Taylor, N. P. and D. C. Zappi. 2004. Cacti of Eastern Brazil. Kew, England: Royal Botanic Gardens.

Walker J. F., Yang Y., Feng T., Timoneda A., Mikenas J., Hutchison V. et al. (2018). From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. American Journal of Botany, 105, 446-462.

Wang N., Yang Y., Moore M. J., Brockington S. F., Walker J. F., Brown J. W. et al. (2019). Evolution of Portulacineae marked by gene tree conflict and gene family expansion associated with adaptation to harsh environments. Molecular Biology and Evolution, 36, 112-126.

Zhang C. and Mirarab S. (2022). Weighting by gene tree uncertainty improves accuracy of quartet-based species trees. Molecular Biology and Evolution, 39, msac215.

2.9 Supplementary Material

Appendix S1. DNA extraction protocol adapted from Inglis et al. 2018 and Martinez-Gonzalez et al. 2017. Distinct steps from Inglis et al. 2018 are highlighted in bold.

Before starting

- Preheat water bath/incubator at 65 ºC
- Warm up 3X CTAB (lysis buffer) at 65 °C for at least 15 minutes before use.
- The Sorbitol (wash buffer) should be stored at 4 ºC, while etanol 70% and isopropanol should be stored at -20 ºC.

- After macerating the tissue plant material, add to the powder material 1.5 mL of sorbitol (wash buffer) and 1% (15 µl) of 2-Mercaptoetanol.
- Centrifuge the material above at 2,500 x g for five minutes at room temperature.
- Discard the supernatant liquid.
- Repeat the sorbitol wash if the supernatant from step 2 and 3 is highly turbid or viscous. One sorbitol wash cycle is usually sufficient but this step can be repeated up to three times.
- Add 700 µl pre-warmed at 65 º C 3% CTAB (lysis buffer) to the sample tubes, **5 µl RNAase A (10 mg/ml),** and 1% of (7 µl) of 2-Mercaptoetanol.
- Incubate samples at 65°C for 60 minutes in a water bath, mixing by inversion every ten minutes.
- Remove from the water bath and cool the samples at room temperature for five minutes.
- Add CIA solution (24 ml chloroform: 1 ml isoamyl alcohol sample tubes (700 µl or fill to approximately 4/5 tube capacity).
- Shake tubes vigorously and **centrifuge for 8,000 x g for 60 minutes at 4 º C.**
- Transfer the supernatant to other tube and add 500 µl of cool isopropanol (-20ºC).
- **Centrifuge for 25 minutes at 8,000 x g at room temperature.**
- Discard carefully the supernatant and wash the pellet with 1000 µl cool ethanol 70% (-20 °C) and centrifuge for 5 min at 10,000 x g in room temperature.
- Dry open tubes at room temperature for approximately one hour
- **Resuspend pellets in 50 µl ultrapure water and incubate tubes at 60° C for 15 minutes.**

Table S1. Sampling information of Cactaceae and Portulacaceae species used for genomic dataset. Herbarium catalogs in voucher: RBvc: Cactário do Jardim Botânico do Rio de Janeiro; SORO: Herbário do Centro de Ciências e Tecnologias para a Sustentabilidade; HURB: Herbário do Recôncavo da Bahia; HUEFS: Herbario da Universidade Estadual de Feira de Santana; KEW: Kew DNA bank; SPF: Herbário da Universidade de São Paulo; UB: Herbário da Universidade de Brasília; GIB: Gibraltar Botanic Gardens; CGMS: Herbário da Fundação Universidade Federal de Mato Grosso do Sul; BHCB: Herbário da UFMG; ASE: Herbário da Universidade Federal de Sergipe; HRCB: Herbário Rioclarense; UFMT: Herbário UFMT; ZSS: Herbarium Zurich Succulent Plant Collection; EAC: Herbário Prisco Bezerra; UEC: Herbário da Universidade Estadual de Campinas; NA: not available. SRA code without numbers will be enabled upon publication of this manuscript.

Species	**Dataset**	**GenBank Access**	**Voucher**
Portulacaceae			
Portulaca hirsutissima	Cactaceae591	SRXXXXXX	UB1290672
Portulaca oleracea	Angiosperm353	EZGR	NA
Cactaceae			
Subfamily Pereskioideae			
Leuenbergeria aureiflora	Cactaceae591	SRR18315853	SORO 7951
Leuenbergeria guamacho	Angiosperm353	SRR7905854	NA
Pereskia aculeata	Cactaceae591	SRXXXXXX	RBvc 225
Pereskia aculeata	Angiosperm353	oneJP	NA
Pereskia bahiensis	S170A51	SRR18315854	RBvc 371
Subfamily Opuntioideae			
Tacinga inamoena	Cactaceae591	SRR18315855	RBvc 205
Tacinga subcylindrica	Cactaceae591	SRR18315857	RBvc 114
Tacinga funalis	Angiosperm353	ERR7618239	ZSS 86 4377
Opuntia monacantha	Cactaceae591	SRR18315858	RBvc 183
Opuntia ficus-indica	Angiosperm353	ERR7618187	ZSS B 13 0042
Subfamily Cactoideae			
Incertae sedis			
Blossfeldia liliputana	Angiosperm353	ERR7619498	ZSS
Copiapoa gigantea	Angiosperm353	ERR7619579	Larridon
Frailea schilinzkyana	Cactaceae591	SRXXXXXXX	
Frailea chiquitana	Angiosperm353	ERR7619821	ZSS 99 9359
Tribe Cacteae			
Astrophytum myriostigma	Cactaceae591	SRR18315866	RBvc 3

Species	Dataset	GenBank Access	Voucher
Astrophytum myriostigma	Angiosperm353	ERR7618193	ZSS D 89 2897
Ferocactus glaucescens	Cactaceae591	SRR18315868	RBvc 207
Ferocactus herrerae	Cactaceae591	SRR18315869	RBvc 271
Echinocactus grusonii	Cactaceae591	SRR18315870	RBvc 214
Echinocactus platyacanthus	Angiosperm353	ERR5033444	RBG 1990-3076
Thelocactus setispinus	Cactaceae591	SRXXXXXXX	RBvc 243
Thelocactus hexaedrophorus	Angiosperm353	ERR7618241	ZSS 87 1765
Tribe Phyllocacteae			
subtribe Corryocactinae			
Acanthocereus cuixmalensis	Angiosperm353	SRR7905861	NA
Pfeiffera ianthothele	Angiosperm353	ERR7619832	ZSS Hun 594
subtribe Hylocereinae			
Epiphyllum phyllanthus	Cactaceae591	SRR18315859	HRCB 5017
Epiphyllum phyllanthus	Angiosperm353	ERR7618206	ZSS 88 3000
Epiphyllum chrysocardium	Cactaceae591	SRR18315862	RBvc 284
Selenicereus setaceus	Cactaceae591	SRR18315860	RBvc 277
Selenicereus anthonyanus	Cactaceae591	SRR18315861	RBvc 203
Selenicereus grandiflorus	Angiosperm353	ERR7618234	ZSS 99 8198
Weberocereus tonduzii	Angiosperm353	ERR7618276	RBG 2007-1368
subtribe Echinocereinae			
Echinocereus pentalophus	Cactaceae591	SRXXXXXXX	RBvc 252
Echinocereus viridiflorus	Angiosperm353	ERR7618202	ZSS 10 0022
Lophocereus marginatus	Cactaceae591	SRR18315864	RBvc 283
Lophocereus marginatus	Angiosperm353	ERR7619827	ZSS 85 3304
Cephalocereus polylophus	Cactaceae591	SRR18315863	RBvc 234
Cephalocereus polylophus	Angiosperm353	ERR7618177	ZSS 94 2643
Carnegiea gigantea	Angiosperm353	GCA002740515 v1	NA
Tribe Rhipsalideae			
Lepismium cruciforme	Angiosperm353	ERR7618184	ZSS 78 1419
Hatiora salicornioides	Cactaceae591	SRR18315871	SORO 7952

Species	**Dataset**	**GenBank Access**	**Voucher**
Hatiora salicornioides	Angiosperm353	ERR7619824	ZSS 95 2343
Rhipsalis lindbergiana	Cactaceae591	SRR18315872	SORO 7953
Rhipsalis paradoxa	Cactaceae591	SRR18315873	RBvc 222
Rhipsalis teres	Angiosperm353	ERR7619464	GENT 20011223
Schlumbergera russelliana	Angiosperm353	ERR7619850	ZSS 96 1386
Tribe Notocacteae			
Eriosyce strausiana	Angiosperm353	ERR7618208	ZSS 96 1135
Parodia leninghausii	Cactaceae591	SRR18315874	SORO 7954
Parodia magnifica	Cactaceae591	SRR18315875	RBvc 37
Parodia microsperma	Angiosperm353	ERR5033447	ZSS 033070
Parodia erinacea	Angiosperm353	ERR7618189	ZSS 89 3232
Neowerdermannia vorwerkii	Angiosperm353	ERR7618223	ZSS 99 8506
Yavia cryptocarpa	Angiosperm353	ERR7619842	ZSS 99 7311
Tribe Cereeae			
Subtribe Rebutiinae			
Aylostera einsteinii	Angiosperm353	ERR7619796	ZSS 16 0289
Aylostera deminuta	Angiosperm353	ERR7619798	ZSS 92 2623
Aylostera pygmaea	Angiosperm353	ERR7619797	ZSS 99 6546
Browningia hertlingiana	Cactaceae591	SRR18315876	KEW 32069
Browningia candelaris	Angiosperm353	ERR7619443	ZSS 99 8520
Gymnocalycium anisitsii	Cactaceae591	SRRXXXXX	SORO 8154
Gymnocalycium denudatum	Cactaceae591	SRR18315877	KEW 32065
Gymnocalycium horstii	Cactaceae591	SRR18315879	RBvc 262
Gymnocalycium pflanzii	Cactaceae591	SRRXXXXX	RBvc 195
Gymnocalycium gibbosum	Angiosperm353	ERR5033445	RBG 1980-3955
Lasiocereus rupicola	Angiosperm353	ERR7619825	ZSS 10 1469
Rebutia minuscula	Angiosperm353	ERR7618231	ZSS 99 6563
Stetsonia coryne	Cactaceae591	SRR18315880	CGMS 17594
Stetsonia coryne	Angiosperm353	ERR7618237	ZSS D 82 3648
Uebelmannia buiningii	Cactaceae591	SRRXXXXX	NA
Uebelmannia gummifera subsp. *gummifera*	Cactaceae591	SRR18315881	HURB 112

Species	**Dataset**	**GenBank Access**	**Voucher**
Uebelmannia pectinifera subsp. *pectinifera*	Cactaceae591	SRR18315882	SORO 4543
Weingartia fidana	Angiosperm353	ERR7618248	ZSS 99 9330
Weingartia neocumingii	Angiosperm353	ERR7618246	ZSS 85 2958
Weingartia steinbachii	Angiosperm353	ERR7618247	ZSS 91 2484
Subtribe Cereinae			
Arrojadoa bahiensis	Cactaceae591	SRRXXXXX	SORO 8158
Arrojadoa dinae	Cactaceae591	SRRXXXXX	SORO 7981
Arrojadoa marylanae	Cactaceae591	SRRXXXXX	RBvc 101
Arrojadoa multiflora	Cactaceae591	SRRXXXXX	HURB 13627
Arrojadoa penicillata	Cactaceae591	SRRXXXXX	SORO 7998
Arrojadoa rhodantha	Cactaceae591	SRR18315883	SORO 4911
Arrojadoa rhodantha	Angiosperm353	ERR7618192	ZSS 11 0486
Arrojadoa olsthoorniana	Cactaceae591	SRRXXXXX	SORO 7986
Brasilicereus estevesii	Cactaceae591	SRRXXXXX	SORO 8014
Brasilicereus markgraffi	Cactaceae591	SRRXXXXX	HURB 10603
Brasilicereus phaeacanthus	Cactaceae591	SRR18315884	SORO 7955
Brasilicereus phaeacanthus	Angiosperm353	ERR5034691	ZSS-032985
Cereus albicaulis	Cactaceae591	SRRXXXXX	SORO 4559
Cereus bicolor	Cactaceae591	SRR18315916	NA
Cereus fernambucensis	Cactaceae591	SRR18315919	SORO 2672
Cereus fricii	Angiosperm353	ERR7619802	ZSS 931839
Cereus hexagonus	Cactaceae591	SRR18315927	NA
Cereus hexagonus	Angiosperm353	ERR7618269	RBG 2009-1967
Cereus hildmannianus	Cactaceae591	SRR18315929	SORO 2746
Cereus jamacaru	Cactaceae591	SRRXXXXX	SORO 8159
Cereus jamacaru	Angiosperm353	ERR7619801	ZSS 86 5065
Cereus phatnospermus	Cactaceae591	SRR18315900	SORO 7969
Cereus mirabella	Cactaceae591	SRR18315911	SORO 7970
Cereus mortensenii	Cactaceae591	SRRXXXXX	GIB G575
Cereus repandus	Cactaceae591	SRR18315885	SORO 7956
Cereus saddianus	Cactaceae591	SRRXXXXX	SORO 3632

Species	Dataset	GenBank Access	Voucher
Cereus spegazzinii	Cactaceae591	SRRXXXXX	SORO 4541
Cereus stenogonus	Cactaceae591	SRR18315933	SORO 7971
Cereus trigonodendron	Cactaceae591	SRRXXXXX	GIB MB046.03
Cereus vargasianus	Cactaceae591	SRRXXXXX	GIB PH1022.02
Cipocereus bradei	Cactaceae591	SRRXXXXX	SORO 4562
Cipocereus crassisepalus	Cactaceae591	SRRXXXXX	NA
Cipocereus laniflorus	Cactaceae591	SRRXXXXX	SORO 8016
Cipocereus pusilliflorus	Cactaceae591	SRRXXXXX	BHCB 21281
Cipocereus pusilliflorus	Angiosperm353	ERR7619804	ZSS 80 3255
Coleocephalocereus aureus	Cactaceae591	SRRXXXXX	HURB 13694
Coleocephalocereus aureus	Angiosperm353	ERR7618181	ZSS B 82 2541
Coleocephalocereus braunii	Cactaceae591	SRRXXXXX	UEC 202555
Coleocephalocereus fluminensis	Cactaceae591	SRRXXXXX	UEC 041393
Coleocephalocereus fluminensis	Angiosperm353	ERR7619808	ZSS B 88 1544
Coleocephalocereus goebelianus	Cactaceae591	SRRXXXXX	HURB 13631
Coleocephalocereus purpureus	Cactaceae591	SRR18315888	SPF 80978
Discocactus bahiensis	Cactaceae591	SRRXXXXX	SORO 8004
Discocactus boliviensis	Cactaceae591	SRRXXXXX	SORO8027
Discocactus diersianus	Cactaceae591	SRRXXXXX	SORO5738
Discocactus ferricola	Cactaceae591	SRRXXXXX	SORO8003
Discocactus heptacanthus var. *cangaensis*	Cactaceae591	SRRXXXXX	SORO7996
Discocactus placentiformis	Cactaceae591	SRRXXXXX	HRCB 13930
Discocactus zehntneri	Cactaceae591	SRRXXXXX	SORO5739
Discocactus zehntneri	Angiosperm353	ERR7619845	ZSS 10 1400
Estevesia alex-bragae	Cactaceae591	SRRXXXXX	SORO 8022
Facheiroa squamosa	Cactaceae591	SRRXXXXX	UB1239283
Facheiroa ulei	Cactaceae591	SRR18315892	HURB 11513
Facheiroa ulei	Angiosperm353	ERR7618183	ZSS 10 1088
Leocereus bahiensis	Cactaceae591	SRRXXXXX	NA
Leocereus bahiensis	Cactaceae591	SRRXXXXX	HURB 7494
Leocereus bahiensis	Angiosperm353	ERR7618214	ZSS D 88 3184

Species	Dataset	GenBank Access	Voucher
Melocactus bahiensis	Cactaceae591	SRRXXXXX	HUEFS 84848
Melocactus broadwayi	Cactaceae591	SRRXXXXX	Anterberger, 64
Melocactus concinnus	Cactaceae591	SRRXXXXX	NA
Melocactus conoideus	Cactaceae591	SRRXXXXX	NA
Melocactus deinacanthus	Cactaceae591	SRRXXXXX	SPF 80989
Melocactus ernestii	Cactaceae591	SRRXXXXX	SORO 5735
Melocactus ferreophilus	Cactaceae591	SRRXXXXX	HUEFS 168189
Melocactus pachyacanthus ssp. *pachyacanthus*	Cactaceae591	SRRXXXXX	SORO 8002
Melocactus levitestatus	Cactaceae591	SRRXXXXX	SORO7982
Melocactus macracanthos	Cactaceae591	SRR18315893	SORO7958
Melocactus mazelianus	Cactaceae591	SRRXXXXX	U.1180141
Melocactus oreas	Cactaceae591	SRRXXXXX	RBvc 135
Melocactus oreas	Angiosperm353	ERR7619847	ZSS E 10 1071
Melocactus sergipensis	Cactaceae591	SRRXXXXX	ASE 33139
Melocactus violaceus	Cactaceae591	SRRXXXXX	EAC 52277
Melocactus zehntneri	Cactaceae591	SRRXXXXX	HUEFS 226476
Micranthocereus albicephalus	Cactaceae591	SRRXXXXX	RBvc 163
Micranthocereus albicephalus	Cactaceae591	SRRXXXXX	SORO 7985
Micranthocereus auriazureus	Cactaceae591	SRRXXXXX	SPF 60294
Micranthocereus auriazureus	Cactaceae591	SRRXXXXX	HURB 13686
Micranthocereus dolichospermaticus	Cactaceae591	SRRXXXXX	UFMT 24569
Micranthocereus dolichospermaticus	Angiosperm353	ERR7619848	ZSS 10 1484
Micranthocereus estevesii	Cactaceae591	SRRXXXXX	SORO 4906
Micranthocereus flaviflorus	Cactaceae591	SRRXXXXX	SORO 7974
Micranthocereus flaviflorus subsp. *densiflorus*	Cactaceae591	SRRXXXXX	HUEFS 73449
Micranthocereus polyanthus	Cactaceae591	SRR18315895	SORO 7959
Micranthocereus polyanthus	Angiosperm353	ERR7618220	ZSS C 10 1079
Micranthocereus purpureus	Cactaceae591	SRR18315896	SORO 7973
Micranthocereus purpureus	Angiosperm353	ERR7618273	RBG 1977-907
Micranthocereus streckeri	Cactaceae591	SRR18315896	SORO 7960
Micranthocereus violaciflorus	Cactaceae591	SRRXXXXX	SORO 7972

Species	Dataset	GenBank Access	Voucher
Pilosocereus albisummus	Cactaceae591	SRRXXXXX	SORO 4530
Pilosocereus arrabidae	Cactaceae591	SRRXXXXX	SORO 2656
Pilosocereus aureispinus	Cactaceae591	SRR18315897	HUFS 642
Pilosocereus aurisetus	Cactaceae591	SRRXXXXX	SORO 4929
Pilosocereus azulensis	Cactaceae591	SRRXXXXX	SORO 4531
Pilosocereus bohlei	Cactaceae591	SRRXXXXX	SORO 3000
Pilosocereus brasiliensis	Cactaceae591	SRRXXXXX	SORO 4912
Pilosocereus flexibilispinus	Cactaceae591	SRRXXXXX	SORO 4935
Pilosocereus floccosus	Cactaceae591	SRRXXXXX	SORO 4558
Pilosocereus glachrocus	Cactaceae591	SRRXXXXX	SORO 4536
Pilosocereus jauruensis	Cactaceae591	SRRXXXXX	SORO 2646
Pilosocereus lanuginosus	Cactaceae591	SRR18315899	SORO 4925
Pilosocereus leucocephalus	Cactaceae591	SRRXXXXX	ZSS B 32820
Pilosocereus leucocephalus	Angiosperm353	ERR7618228	ZSS 95 1447
Pilosocereus mollispinus	Cactaceae591	SRRXXXXX	SORO 8015
Pilosocereus multicostatus	Cactaceae591	SRRXXXXX	SORO 2649
Pilosocereus pachycladus	Cactaceae591	SRRXXXXX	SORO 4913
Pilosocereus polygonus	Cactaceae591	SRRXXXXX	ZSS B 32823
Pilosocereus vilaboensis	Cactaceae591	SRRXXXXX	SORO 3001
Pilosocereus pentaedrophorus	Cactaceae591	SRRXXXXX	
Praecereus euchlorus	Cactaceae591	SRR18315901	SORO 5704
Praecereus euchlorus	Angiosperm353	ERR7620900	ZSS 89 2295
Praecereus saxicola	Cactaceae591	SRRXXXXX	SORO 5703
Stephanocereus leucostele	Cactaceae591	SRR18315902	SORO 7961
Stephanocereus leucostele	Cactaceae591	SRRXXXXX	SORO 7990
Stephanocereus leucostele	Angiosperm353	ERR7618236	ZSS 99 7350
Stephanocereus luetzelburgii	Cactaceae591	SRRXXXXX	HUEFS 107361
Stephanocereus luetzelburgii	Cactaceae591	SRRXXXXX	UB 1324887
Xiquexique frewenii	Cactaceae591	SRR18315903	SORO 4937
Xiquexique gounellei	Cactaceae591	SRR18315904	SORO 4934
Xiquexique gounellei	Angiosperm353	ERR7618275	RBG 2002-3078

Species	Dataset	GenBank Access	Voucher
Xiquexique gounellei subsp. *zehntneri*	Cactaceae591	SRRXXXXX	SORO 4915
Xiquexique tuberculatus	Cactaceae591	SRRXXXXX	SORO 4930
Subtribe Trichocereinae			
Acanthocalycium rhodotrichum	Cactaceae591	SRRXXXXX	SORO 8021
Acanthocalycium spiniflorum	Angiosperm353	ERR7618203	ZSS 10 0439
Arthrocereus glaziovii	Cactaceae591	SRRXXXXX	RB 757784
Arthrocereus rondonianus	Cactaceae591	SRRXXXXX	RBvc 281
Arthrocereus spinosissimus	Cactaceae591	SRRXXXXX	RB 755039
Borzicactus sepium	Angiosperm353	ERR7619442	ZSS 89 1518
Chamaecereus silvestrii	Angiosperm353	ERR7619447	ZSS 2 85 1839
Cleistocactus baumannii	Angiosperm353	ERR7619805	ZSS A 99 9185
Cleistocactus chrysocephalus	Angiosperm353	ERR7618198	ZSS 1B 85 1468
Cleistocactus hyalacanthus	Angiosperm353	ERR7618199	RBG 1999-3627
Cleistocactus winteri	Cactaceae591	SRR18315906	RBvc 282
Cleistocactus winteri	Angiosperm353	ERR7619806	ZSS 80 3277
Denmoza rhodacantha	Angiosperm353	ERR7618203	ZSS 10 0439
Echinopsis oxygona	Cactaceae591	SRR18315908	RBvc 117
Echinopsis oxygona	Angiosperm353	ERR7619448	ZSS A1 98 1499
Echinopsis calochlora	Cactaceae591	SRRXXXXX	SORO 8026
Espostoa blossfeldiorum	Angiosperm353	ERR7619817	ZSS A 99 5825
Espostoa lanata	Angiosperm353	ERR7619816	ZSS A 11 0609
Espostoopsis dybowskii	Cactaceae591	SRR18315909	SORO 7963
Espostoopsis dybowskii	Cactaceae591	SRRXXXXX	KEW 32075
Espostoopsis dybowskii	Angiosperm353	ERR7619818	ZSS 90 4640
Haageocereus fascicularis	Angiosperm353	ERR7618245	ZSS 11 0755
Haageocereus pseudomelanostele	Angiosperm353	ERR7619449	ZSS 88 1991
Haageocereus decumbens	Cactaceae591	SRR18315910	KEW 32064
Harrisia adscendens	Cactaceae591	SRR18315912	SORO 4910
Harrisia bonplandii	Cactaceae591	SRRXXXXX	SORO 8013
Harrisia martinii	Cactaceae591	SRRXXXXX	KEW 32078
Harrisia tortuosa	Angiosperm353	ERR7618272	RBG 1985-2774

Species	Dataset	GenBank Access	Voucher
Leucostele atacamensis	Cactaceae591	SRRXXXXX	SORO 6565
Lobivia ancistrophora	Cactaceae591	SRRXXXXX	RBvc 264
Lobivia tegeleriana	Angiosperm353	ERR7619812	ZSS 93 2403
Matucana haynei	Angiosperm353	ERR5033441	ZSS-033049
Matucana intertexta	Cactaceae591	SRR18315913	KEW 32063
Oreocereus celsianus	Cactaceae591	SRRXXXXX	KEW 32067
Oreocereus pseudofossulatus	Angiosperm353	ERR7618178	ZSS 99 6008
Oroya peruviana	Angiosperm353	ERR7618224	ZSS 92 2847
Reicheocactus famatimensis	Angiosperm353	ERR7618204	ZSS A 92 1806
Samaipaticereus corroanus	Cactaceae591	SRR18315914	KEW 32076
Samaipaticereus corroanus	Angiosperm353	ERR7618232	ZSS 90 3741
Setiechinopsis mirabilis	Angiosperm353	ERR7619813	ZSS HUN 403
Soenrensia formosa	Cactaceae591	SRRXXXXX	KEW 32077
Soehrensia bruchii	Angiosperm353	ERR7619814	ZSS A 77 1213
Trichocereus macrogonus	Angiosperm353	ERR7618205	ZSS 14 0002
Vatricania guentheri	Angiosperm353	ERR7618244	ZSS 94 2200
Yungasocereus inquisivensis	Angiosperm353	ERR7619843	ZSS 96 2070

Table S2. Sequence matrix of species and regions included in the gene-scale dataset. GenBank and SRA accession numbers are from previous and newly generated studies. SRA code without numbers will be enabled upon publication of this manuscript

	trnL trnF	*trnS-trnG*	*trnK-matK*	*atpb-rbcl*	*trnT-trnL*	*petL-psbE*	*rbcl*	*rpl16*	*phyC*	*ppc*
Subtribe Cereinae										
Arrojadoa bahiensis	SRRXXXXX	**MW546673.1**	SRRXXXXX	SRRXXXXX	**MW546627.1**	MW546779.1	SRRXXXXX		SRRXXXXX	SRRXXXXX
Arrojadoa dinae	SRRXXXXX	MW546674.1	SRRXXXXX	SRRXXXXX	MW546628.1	MW546780.1	SRRXXXXX	MW546714.1	SRRXXXXX	SRRXXXXX
Arrojadoa marylaniae			SRRXXXXX	SRRXXXXX			SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX
Arrojadoa multiflora			SRRXXXXX	SRRXXXXX			SRRXXXXX		SRRXXXXX	SRRXXXXX
Arrojadoa olsthoorniana	SRRXXXXX		SRRXXXXX	SRRXXXXX	SRRXXXXX		SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX
Arrojadoa penicillata		MW546675.1	SRRXXXXX	SRRXXXXX	MW546630.1	MW546782.1	SRRXXXXX	MW546715.1	SRRXXXXX	SRRXXXXX
Arrojadoa rhodantha (T)	**HM041230.1**	MW546676.1	SRR18315883	SRR18315883	SRR18256121	MW546783.1	SRR18315883	MW546716.1	SRR18315883	SRR18315883
Brasilicereus markgrafii		MW546677.1	SRRXXXXX	SRRXXXXX	MW546632.1	MW546784.1	SRRXXXXX	MW546717.1	SRRXXXXX	SRRXXXXX
Cereus albicaulis	KR998121.1	KP017441.1	SRRXXXXX	SRRXXXXX	MW546633.1	MW546785.1	SRRXXXXX	KR998129.1	SRRXXXXX	SRRXXXXX
Cereus bicolor		KP017426.1	SRR18315916	SRR18315916	KR998126.1		SRR18315916		SRR18315916	SRR18315916
Cereus fernambucensis	KT164949.1	KT164934.1	SRR18315919	SRR18315919	KR998125.1	MW546786.1	SRR18315919	MW546718.1	SRR18315919	SRR18315919
Cereus hexagonus		KP017429.1	SRR18315927	SRR18315927			SRR18315927		SRR18315927	SRR18315927
Cereus hildmannianus	HM041240.1	MW546680.1	SRR18315929	SRR18315929		MW546787.1	SRR18315929	HM041395.1	SRR18315929	SRR18315929
Cereus jamacaru	SRR	KP017419.1	SRRXXXXX		SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX
Cereus mirabella		KP017443.1	SRR18315911	SRR18315911	KR998124.1	MW546790.1	SRR18315911	KR998130.1	SRR18315911	SRR18315911
Cereus repandus		KP017427.1	SRR18315885	SRR18315885			SRR18315885		SRR18315885	SRR18315885
Cereus saddianus	KR998120.1	KP017445.1	SRRXXXXX			KR998112.1	SRRXXXXX	KR998131.1	SRRXXXXX	SRRXXXXX
Cereus trigonodendron	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX		SRRXXXXX		SRRXXXXX	SRRXXXXX

	trnL trnF	*trnS-trnG*	*trnK-matK*	*atpb-rbcl*	*trnT-trnL*	*petL-psbE*	*rbcl*	*rpl16*	*phyC*	*ppc*
Cereus vargasianus	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX
Cipocereus bradei		MW546685.1	SRRXXXXX	SRRXXXXX	MW546640.1	MW546792.1	SRRXXXXX	MW546720.1	SRRXXXXX	SRRXXXXX
Cipocereus crassisepalus		MW546686.1	SRRXXXXX	SRRXXXXX	MW546641.1	MW546793.1	SRRXXXXX	MW546721.1	SRRXXXXX	SRRXXXXX
Cipocereus pussiliflorus			SRRXXXXX	SRRXXXXX	MW546643.1	MW546795.1	SRRXXXXX	MW546722.1	SRRXXXXX	SRRXXXXX
Coleocephalocereus aureus	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX
Coleocephalocereus fluminensis	AY015405.1		SRRXXXXX	SRRXXXXX	MW546644.1	MW546796.1	SRRXXXXX	MW546723.1	SRRXXXXX	SRRXXXXX
Coleocephalocereus goebilanus	SRRXXXXX		SRRXXXXX	SRRXXXXX	MW546645.1	MW546797.1	SRRXXXXX	MW546724.1	SRRXXXXX	SRRXXXXX
Discocactus bahiensis	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX		SRRXXXXX	SRRXXXXX
Discocactus placentiformis	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX		SRRXXXXX		SRRXXXXX	SRRXXXXX
Discocactus zehntneri	HM041255.1	SRRXXXXX	SRRXXXXX	SRRXXXXX	MW546647.1	SRRXXXXX	SRRXXXXX	MW546726.1	SRRXXXXX	SRRXXXXX
Facheiroa ulei	JQ779689.1	JQ779527.1	SRR18315892	SRR18315892	SRR18315892		SRR18315892		SRR18315892	SRR18315892
Leocereus bahiensis	JQ779672.1	JQ779510.1	SRRXXXXX	SRRXXXXX	MW546649.1	MW546801.1	SRRXXXXX	MW546728.1	SRRXXXXX	SRRXXXXX
Melocactus bahiensis		MW546693.1	SRRXXXXX	SRRXXXXX	MW546650.1		SRRXXXXX	MW546729.1	SRRXXXXX	SRRXXXXX
Melocactus broadwayi	SRR18256111	SRR18256111	SRRXXXXX	SRRXXXXX	SRR18256111	SRR18256111	SRRXXXXX	SRR18256111	SRRXXXXX	SRRXXXXX
Melocactus caroli-linnaei	SRR18256115	SRR18256115	SRRXXXXX	SRRXXXXX	SRRXXXXX		SRRXXXXX	SRRXXXXX		SRRXXXXX
Melocactus concinnus	SRR	MW546694.1	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	
Melocactus deinacanthus	SRR	SRR	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX
Melocactus ernestii		MW546695.1	SRRXXXXX	SRRXXXXX	MW546652.1	MW546804.1	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX
Melocactus intortus	HM041301.1	SRR18256140	SRRXXXXX	SRRXXXXX	SRRXXXXX		SRRXXXXX	SRR18256140		
Melocactus macracanthos	SRR18256137	SRR18256137	SRR18315893	SRR18315893	SRR18256137	SRR18256137	SRR18315893	SRR18256137	SRR18315893	SRR18315893
Melocactus mazelianus	SRR18256136		SRRXXXXX	SRRXXXXX			SRRXXXXX		SRRXXXXX	SRRXXXXX

	trnL trnF	*trnS-trnG*	*trnK-matK*	*atpb-rbcl*	*trnT-trnL*	*petL-psbE*	*rbcl*	*rpl16*	*phyC*	*ppc*
Melocactus oreas	AM502567.1	MW546697.1	SRRXXXXX	SRRXXXXX	MW546654.1	MW546806.1	SRRXXXXX	MW546732.1	SRRXXXXX	SRRXXXXX
Melocactus pachyacanthus			SRRXXXXX	SRRXXXXX	SRRXXXXX		SRRXXXXX		SRRXXXXX	SRRXXXXX
Melocactus sergipensis	SRRXXXXX		SRRXXXXX	SRRXXXXX			SRRXXXXX		SRRXXXXX	SRRXXXXX
Melocactus violaceus	SRRXXXXX		SRRXXXXX	SRRXXXXX	MW546656.1	MW546808.1	SRRXXXXX		SRRXXXXX	SRRXXXXX
Melocactus zehntneri		KX301198.1	SRRXXXXX	SRRXXXXX	KX301160.1	MW546809.1	SRRXXXXX	MW546735.1	SRRXXXXX	SRRXXXXX
Micranthocereus albicephalus	AY015402.1		SRRXXXXX	SRRXXXXX	MW546658.1	MW546810.1	SRRXXXXX	MW546736.1	SRRXXXXX	SRRXXXXX
Micranthocereus auri-azureus	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX
Micranthocereus densiflorus	JQ779686.1	JQ779524.1	SRRXXXXX	SRRXXXXX			SRRXXXXX	JQ779820.1	SRRXXXXX	SRRXXXXX
Micranthocereus estevesii		MW546701.1	SRRXXXXX	SRRXXXXX	MW546660.1	MW546812.1	SRRXXXXX	OP189046.1	SRRXXXXX	SRRXXXXX
Micranthocereus flaviflorus		MW546702.1	SRRXXXXX	SRRXXXXX	MW546661.1	MW546813.1	SRRXXXXX	MW546737.1	SRRXXXXX	SRRXXXXX
Micranthocereus polyanthus		MW546703.1	SRR18315895	SRR18315895	MW546662.1	MW546814.1	SRR18315895		SRR18315895	SRR18315895
Micranthocereus purpureus		MW546704.1	SRR18315896	SRR18315896	MW546663.1	MW546815.1	SRR18315896		SRR18315896	SRR18315896
Micranthocereus violaciflorus	SRRXXXXX		SRRXXXXX	SRRXXXXX			SRRXXXXX		SRRXXXXX	SRRXXXXX
Pilosocereus albisummus		KX301216.1	SRRXXXXX	SRRXXXXX	KX301178.1	KX301140.1	SRRXXXXX	OP189047.1	SRRXXXXX	SRRXXXXX
Pilosocereus arrabidae		KX301222.1	SRRXXXXX	SRRXXXXX	KX301184.1	KX387747.1	SRRXXXXX	OP189049.1	SRRXXXXX	SRRXXXXX
Pilosocereus aureispinus	JN035536.1	KX301201.1	SRR18315897	SRR18315897	MF694643.1	KX301123.1	SRR18315897	OP189051.1	SRR18315897	SRR18315897
Pilosocereus aurisetus	SRRXXXXX	KC779361.1	SRRXXXXX	SRRXXXXX	SRRXXXXX	MW546816.1	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX
Pilosocereus azulensis		KX301214.1	SRRXXXXX	SRRXXXXX	KX301176.1	KX301138.1	SRRXXXXX		SRRXXXXX	SRRXXXXX
Pilosocereus bohlei		KX301211.1	SRRXXXXX	SRRXXXXX	KC621218.1	KX387748.1	SRRXXXXX	OP189053.1	SRRXXXXX	SRRXXXXX
Pilosocereus brasiliensis		KX301223.1	SRRXXXXX	SRRXXXXX	KX301185.1	KX387755.1	SRRXXXXX	OP189054.1	SRRXXXXX	SRRXXXXX
Pilosocereus flexibilispinus			SRRXXXXX	SRRXXXXX		OP187280.1	SRRXXXXX	OP189066.1	SRRXXXXX	SRRXXXXX

	trnL trnF	*trnS-trnG*	*trnK-matK*	*atpb-rbcl*	*trnT-trnL*	*petL-psbE*	*rbcl*	*rpl16*	*phyC*	*ppc*
Pilosocereus floccosus		KX301220.1	SRRXXXXX	SRRXXXXX	KX301182.1	KX301144.1	SRRXXXXX	OP189067.1	SRRXXXXX	SRRXXXXX
Pilosocereus glaucochrous		KX301202.1	SRRXXXXX	SRRXXXXX	KX301164.1	KX301126.1	SRRXXXXX	OP189072.1	SRRXXXXX	SRRXXXXX
Pilosocereus jauruensis		KY559357.1	SRRXXXXX	SRRXXXXX	KC621164.1	KX387746.1	SRRXXXXX	OP189076.1	SRRXXXXX	SRRXXXXX
Pilosocereus lanuginosus		KX387791.1	SRR18315899	SRR18315899		KX387731.1	SRR18315899	OP189077.1	SRR18315899	SRR18315899
Pilosocereus leucocephalus	MZ541760.1	KX301193.1	SRRXXXXX	SRRXXXXX	KX301155.1	KX387725.1	SRRXXXXX	MZ541809.1	SRRXXXXX	SRRXXXXX
Pilosocereus multicostatus		KX387806.1	SRRXXXXX	SRRXXXXX	KX301168.1	KX387752.1	SRRXXXXX	OP189084.1	SRRXXXXX	SRRXXXXX
Pilosocereus pachycladus		KX301209.1	SRRXXXXX	SRRXXXXX	KX301159.1	KX387753.1	SRRXXXXX	JQ889309.1	SRRXXXXX	SRRXXXXX
Pilosocereus polygonus		KX387794.1	SRRXXXXX	SRRXXXXX		KX387734.1	SRRXXXXX	OP189092.1	SRRXXXXX	SRRXXXXX
Pilosocereus royenii		KX301192.1	SRRXXXXX	SRRXXXXX		KX301145.1		OP189096.1	SRRXXXXX	
Pilosocereus vilaboensis	JN035533.1	KX387800.1	SRRXXXXX		JN035605.1	KX387745.1	SRRXXXXX	OP189095.1	SRRXXXXX	SRRXXXXX
Praecereus euchlorus	JQ779666.1	JQ779504.1	SRR18315901	SRR18315901	MW546665.1	MW546817.1	SRR18315901		SRR18315901	SRR18315901
Praecereus saxicola	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX		SRRXXXXX		SRRXXXXX	SRRXXXXX
Stephanocereus leucostele	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	MW546818.1	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX
Stephanocereus luetzelburgii	SRRXXXXX	MW546707.1	SRRXXXXX	SRRXXXXX	MW546667.1	MW546819.1	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX
Xiquexique frewenii			SRR18315903	OP189129.1		OP187283.1	SRR18315903	OP189069.1	SRR18315903	SRR18315903
Xiquexique gounellei	MZ541775.1	KX387787.1	SRRXXXXX	SRRXXXXX	KX301156.1	KX387727.1	SRRXXXXX	MW546745.1	SRRXXXXX	SRRXXXXX
Xiquexique tuberculatus	SRRXXXXX	SRRXXXXX	SRRXXXXX	OP189158.1	SRRXXXXX	SRRXXXXX	SRRXXXXX	MW546746.1	SRRXXXXX	SRRXXXXX
subtribe Rebutiinae										
Aylostera deminuta	HM041349.1	JQ779521.1	HM041768.1	KU737358.1				JQ779818.1		
Aylostera pygmaea	JQ779680.1	JQ779518.1	JX683851.1	KU737383.1				JQ779816.1		
Browningia candelaris	HM041235.1		HM041656.1	AM502346.1				HM041390.1		HM041544.1

	trnL trnF	*trnS-trnG*	*trnK-matK*	*atpb-rbcl*	*trnT-trnL*	*petL-psbE*	*rbcl*	*rpl16*	*phyC*	*ppc*
Browningia hertingliana	JQ779688.1	JQ779526.1	SRR18315876	AM502347.1	AF432924.1		SRR18315876	FN673555.1	SRR18315876	SRR18315876
Gymnocalycium amerhauseri	FR667113.1		FR821475.1		FR667113.1	FR848285.1				
Gymnocalycium anisitsii	AM502401.1		FR821497.1	AM502400.1		FR848281.1	SRRXXXXX		SRRXXXXX	SRRXXXXX
Gymnocalycium bodenbenderianum	FR667090.1		JX905222.1	HQ188323.1	FR667090.1	FR848257.1				
Gymnocalycium bruchii	FR667079.1		FR821441.1	HQ188307.1	FR667079.1	FR848251.1				
Gymnocalycium capillense	FR667142.1		FR821504.1	HQ188326.1	FR667142.1	FR848314.1				
Gymnocalycium denudatum	FR667087.1			SRR18315877	FR667087.1	FR848259.1	SRR18315877		SRR18315877	SRR18315877
Gymnocalycium horstii	FR667075.1		FR821437.1	SRR18315879	FR667075.1	FR848247.1	SRR18315879		SRR18315879	SRR18315879
Gymnocalycium hossei	FR667089.1		JX905238.1	HQ188318.1	FR667148.1	FR848254.1		HM041436.1		
Gymnocalycium hyptiacanthum	FR667115.1		HM041703.1		FR667115.1	FR848273.1		HM041438.1		
Gymnocalycium mesopotamicum	FR667092.1		FR821454.1	HQ188319.1	FR667092.1	FR848264.1				
Gymnocalycium mostii	FR667129.1	JQ779513.1	JX905234.1		FR667076.1	FR848303.1		JQ779811.1		
Gymnocalycium paraguayense	FR667086.1		FR821448.1	HQ188321.1	FR667086.1	FR848258.1				
Gymnocalycium pflanzii	FR667094.1		FR821456.1	SRRXXXXX	FR667094.1	FR848266.1	SRRXXXXX		SRRXXXXX	SRRXXXXX
Gymnocalycium saglionis	HM041283.1		JX683853.1	HQ188324.1	FR667073.1	FR848245.1		HM041437.1		
Gymnocalycium spegazzinii	SRR18256152	SRR18256152	SRR18256152	HQ188325.1	SRR18256152	SRR18256152	SRR18256152	SRR18256152		
Lasiocereus fulvus	JQ779687.1	JQ779525.1	JX683843.1					JQ779821.1		
Rebutia fabrisii	JQ779678.1	JQ779516.1		GU084401.1				JQ779814.1		
Rebutia minuscula	JQ779677.1	JQ779515.1		AM502470.1				JQ779813.1		
Rebutia padcayensis	JQ779676.1	JQ779514.1		AM502473.1				JQ779812.1		
Stetsonia coryne	HM041361.1	JQ779508.1			AF432939.1			HM041518.1		

	trnL trnF	*trnS-trnG*	*trnK-matK*	*atpb-rbcl*	*trnT-trnL*	*petL-psbE*	*rbcl*	*rpl16*	*phyC*	*ppc*
Weingartia arenacea	SRR18256126	JQ779522.1	SRR18256126	AM502420.1	SRR18256126	SRR18256126	SRR18256126	HM041504.1		HM041609.1
Weingartia fidana	AM502457.1		JX683868.1	AM502456.1			HQ621341.1		HQ621148.1	
Weingartia neocumingii	JQ779685.1	JQ779523.1		AM502385.1						
Uebelmannia buiningii			SRRXXXXX	SRRXXXXX			SRRXXXXX		SRRXXXXX	SRRXXXXX
Uebelmannia gummifera			SRR18315881	SRR18315881			SRR18315881		SRR18315881	SRR18315881
Uebelmannia pectinifera	HM041375.1	MW546712.1	SRR18315882		MW546672.1	MW546825.1	SRR18315882	MW546752.1	SRR18315882	SRR18315882
Subtribe Trichocereinae										
Acanthocalycium leucanthum	HM041266.1	JQ779443.1	HM041686.1					HM041420.1		
Arthrocereus glaziovii	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX
Arthrocereus rondonianus	JQ779668.1	JQ779506.1	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX
Arthrocereus spinosissimus	JQ779667.1	JQ779505.1	SRRXXXXX	SRRXXXXX			SRRXXXXX	JN166920.1	SRRXXXXX	SRRXXXXX
Borzicactus icosagonus	HM041241.1		JX683866.1					HM041396.1		HM041550.1
Borzicactus sepium	JQ779625.1	JQ779463.1	JX683852.1					KM487307.1		
Cleistocactus baumannii	JQ779662.1	JQ779500.1	JX683877.1					KM487306.1		
Cleistocactus parviflorus	HM041242.1		HM041662.1					HM041397.1		HM041551.1
Denmoza rhodacantha	HM041254.1	JQ779446.1	JX683840.1	AM502538.1				HM041408.1		HM041561.1
Espostoa lanata	JQ779616.1	JQ779454.1	JX683863.1	AM502551.1	SRR21560216	SRR21560216	SRR21560216	SRR21560216		
Espostoopsis dybowskii	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX
Echinopsis aurea	JQ779638.1	FR716745.1	FN669743.1	JN166878.1			FR853367.1	FN673649.1		
Echinopsis calochlora	JQ779647.1	JQ779485.1	SRRXXXXX	SRRXXXXX			SRRXXXXX	JQ779786.1	SRRXXXXX	SRRXXXXX
Echinopsis oxygona	JQ779645.1	JQ779483.1	SRR18315908	SRR18315908			SRR18315908	JQ779785.1	SRR18315908	SRR18315908

	trnL trnF	*trnS-trnG*	*trnK-matK*	*atpb-rbcl*	*trnT-trnL*	*petL-psbE*	*rbcl*	*rpl16*	*phyC*	*ppc*
Haageocereus decumbens	SRRXXXXX		SRRXXXXX	JQ889304.1	SRRXXXXX		SRRXXXXX	HM041439.1	SRRXXXXX	SRRXXXXX
Haageocereus kagenekii	JQ779618.1	JQ779456.1		JQ889305.1				JQ889308.1		
Harrisia adscendens		MW546709.1	SRR18315912	SRR18315912	AF432929.1	MW546821.1	SRR18315912	JN166899.1	SRR18315912	SRR18315912
Harrisia bonplandii		SRRXXXXX	SRRXXXXX	SRRXXXXX			SRRXXXXX		SRRXXXXX	SRRXXXXX
Harrisia earlei	DQ099939.1			JN166872.1		KY624768.1		DQ100008.1		
Harrisia gracilis	JQ779654.1	JQ779492.1		JX135015.1				JQ779792.1		
Harrisia martinii	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX	SRRXXXXX
Leucostele atacamensis	JQ779648.1	JQ779486.1	SRRXXXXX	SRRXXXXX			SRRXXXXX	HM041422.1	SRRXXXXX	SRRXXXXX
Leucostele terscheckii	JQ779650.1	JQ779488.1		JN166880.1				JN166904.1		
Lobivia ancistrophora	JQ779529.1	JQ779367.1	SRRXXXXX	AM502405.1			SRRXXXXX	JQ779691.1	SRRXXXXX	SRRXXXXX
Lobivia cinnabarina	JQ779552.1	JQ779390.1		AM502355.1				JQ779706.1		
Lobivia pentlandii	JQ779550.1	JQ779388.1	AY015323.1	AM502441.1				JQ779704.1		
Matucana aurantiaca	AM502564.1		FR821506.1	AM502563.1	FR667144.1	FR848316.1				
Matucana intertexta	JQ779620.1	JQ779458.1	SRR18315913	SRR18315913				JQ779764.1	SRR18315913	SRR18315913
Matucana madisoniorum	HM041299.1		HM041718.1					HM041453.1		HM041639.1
Mila caespitosa	JQ779619.1	SRR18256131	JX683872.1	SRR18256131	AF432933.1	SRR18256131	SRR18256131	HM041456.1		
Oreocereus celsianus	FR667145.1	JQ779464.1	SRRXXXXX	AM502576.1	FR667145.1	FR848317.1		HM041483.1	SRRXXXXX	SRRXXXXX
Oroya peruviana	JQ779623.1	JQ779461.1		AM502578.1				JQ779766.1		
Pygmaeocereus bylesianus	SRR18256127	JQ779453.1	SRR18256127		SRR18256127	SRR18256127	SRR18256127	SRR18256127		
Rauhocereus riosaniensis	JQ779617.1	JQ779455.1	AY015326.1					JQ779761.1		
Samaipaticereus corroanus	JQ779660.1	JQ779498.1	SRR18315914	SRR18315914	SRR18256125	SRR18256125	SRR18315914	JN166906.1	SRR18315914	SRR18315914

	trnL trnF	*trnS-trnG*	*trnK-matK*	*atpb-rbcl*	*trnT-trnL*	*petL-psbE*	*rbcl*	*rpl16*	*phyC*	*ppc*
Setiechinopsis mirabilis	JQ779607.1	JQ779445.1		AM502544.1				JQ779751.1		
Soehrensia candicans	JQ779573.1	JQ779411.1		JN166885.1	AF432941.1			JN166907.1		
Soehrensia formosa	JQ779582.1	JQ779420.1	SRRXXXXX	JN166884.1			SRRXXXXX	HM041419.1	SRRXXXXX	SRRXXXXX
Soehrensia hahniana	JQ779588.1	JQ779426.1		JN166888.1				JN166908.1		
Soehrensia schickendantzii	JQ779585.1	JQ779423.1		JN166889.1				JN166909.1		
Soehrensia vasquezii	JQ779587.1	JQ779424.1		JN166891.1				JN166913.1		
Trichocereus bridgesii	JQ779611.1	JQ779449.1		JN166894.1				JN166915.1		
Trichocereus macrogonus	JQ779612.1	JQ779450.1	HM041687.1	JN166895.1				HM041421.1		
Vatricania guentheri	JQ779657.1	JQ779495.1	JX683871.1					JQ779795.1		
Yungasocereus inquisivensis	JQ779659.1	JQ779497.1	JX683858.1	JN166898.1				JN166919.1		
Outgroups										
Astrophytum myriostigma	MK449220.1		SRR18315866	MH129814.1				MK449158.1		SRR18315866
Aztekium ritteri	MK449221.1		AY015290.1	MH129811.1		MK284039.1	MK449093.1	MK449159.1	HQ621080.1	
Blossfeldia liliputana	HM041234.1	MG263783.1	AY015284.1				AY875232.1	HM041389.1	AY875301.1	HM041543.1
Calymmanthium substerile	DQ099926.1	SRR13215330	AY015291.1	SRR13215330		SRR13215330	AY875230.1	FN673676.1	AY875314.1	
Carnegiea gigantea	HM041236.1		HM041657.1	MH129808.1		KY624795.1	KT164772.1	AY181591.1		HM041545.1
Echinocactus grusonii	HM041257.1		SRR18315870				SRR18315870	HM041411.1	SRR18315870	SRR18315870
Echinocereus pentalophus			SRXXXXXXX				SRXXXXXXX		SRXXXXXXX	SRXXXXXXX
Epiphyllum phyllanthus	HM041270.1		SRR18315859	SRR18315859			SRR18315859	LT745550.1	SRR18315859	SRR18315859
Ferocactus glaucescens			SRR18315868	SRR18315868			SRR18315868		SRR18315868	SRR18315868
Ferocactus herrarea			SRR18315869	SRR18315869			SRR18315869		SRR18315869	SRR18315869

	trnL trnF	*trnS-trnG*	*trnK-matK*	*atpb-rbcl*	*trnT-trnL*	*petL-psbE*	*rbcl*	*rpl16*	*phyC*	*ppc*
Frailea schilinzkyana			SRXXXXXXX	SRXXXXXX X			SRXXXXXXX		SRXXXXXXX	SRXXXXXXX
Hatiora salicornoides			SRR18315871	SRR18315871			SRR18315871	FN673564.1	SRR18315871	SRR18315871
Selenicereus setaceus			SRR18315860	SRR18315860			SRR18315860		SRR18315860	SRR18315860
Leunbergeria aureiflora	SRR18315853		SRR18315853	SRR18315853			SRR18315853		SRR18315853	SRR18315853
Cephalocereus polylophus	HM041307.1		SRR18315863	SRR18315863			SRR18315863	AY181597.1	SRR18315863	SRR18315863
Opuntia monacantha			SRR18315858	SRR18315858			SRR18315858		SRR18315858	SRR18315858
Lophocereus pringlei	AY181642.2		FN997217.1			KY624799.1	KF783463.1	AY181589.1	JX627714.1	KC196834.1
Lophocereus marginatus			SRR18315864				SRR18315864		SRR18315864	SRR18315864
Parodia leninghausii			SRR18315874	SRR18315874			SRR18315874		SRR18315874	SRR18315874
Parodia magnifica	HM041334.1		SRR18315875	SRR18315875			SRR18315875	HM041490.1	SRR18315875	SRR18315875
Pereskia aculeata	SRXXXXXX		SRXXXXXX	SRXXXXXX	SRXXXXXX	SRXXXXXX	SRXXXXXX	SRXXXXXX	SRXXXXXX	SRXXXXXX
Pereskia bahiensis			SRR18315854	SRR18315854		MW546823.1	SRR18315854	MW546750.1	SRR18315854	SRR18315854
Portulaca hirutissima	SRXXXXXX		SRXXXXXX	SRXXXXXX			SRXXXXXX		SRXXXXXX	
Rhipsalis baccifera	HM041350.1	KX387797.1	FN669656.1			MW546824.1	FR853282.1	FR853121.1	MW546776.1	HM041619.1
Rhipsalis lindbergiana			SRR18315872	SRR18315872			SRR18315872		SRR18315872	SRR18315872
Rhipsalis paradoxa			SRR18315873	SRR18315873			SRR18315873		SRR18315873	SRR18315873
Selenicereus anthonyanus	LT745476.1		SRR18315861	SRR18315861			SRR18315861	LT745591.1	SRR18315861	SRR18315861
Epiphyllum chrysocardium	KU598053.1		SRR18315859	SRR18315859			SRR18315859	LT745545.1	SRR18315859	SRR18315859
Tacinga inamoena			SRR18315855	SRR18315855			SRR18315855		SRR18315855	SRR18315855
Tacinga subcilindrica			SRR18315857	SRR18315857			SRR18315857		SRR18315857	SRR18315857
Thelocactus setispinus	SRXXXXXX		SRXXXXXX	SRXXXXXX	SRXXXXXX		SRXXXXXX		SRXXXXXX	SRXXXXXX

Figure S1. Heatmap indicating 591 orthologs' recovery success per sample; scale color indicates success rate.

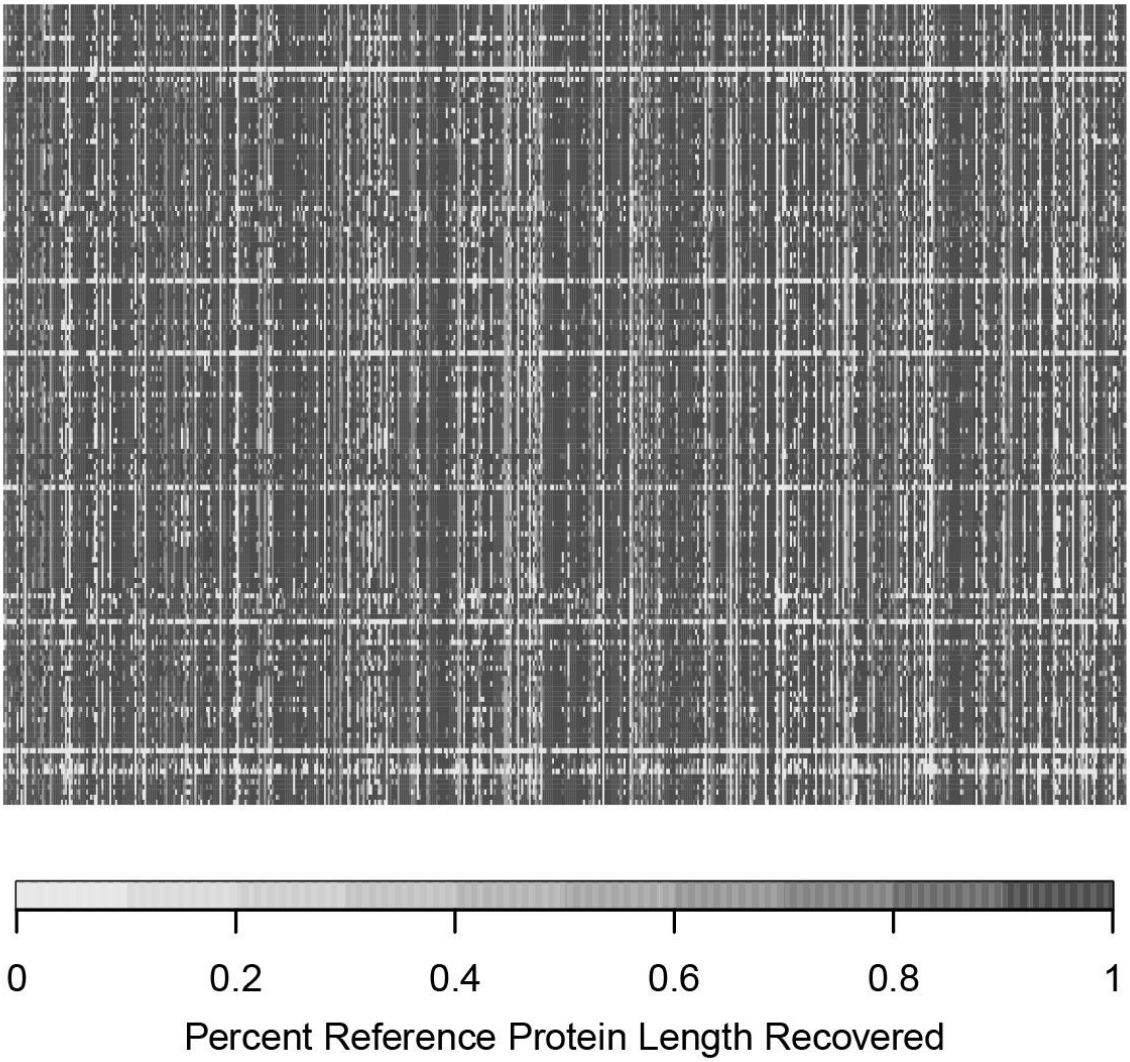


Figure S2. Maximum likelihood phylogenetic inference estimated by IQ-TREE2 using gene-scale dataset. Highly supported branches (BS/SH-aLRT > 95/80) are depicted with black circles and moderately supported branches (BS 95 and BS/SH-aLRT < 80) are depicted with gray circles. Low supported nodes are shown with dashed lines in respective branches.

Figure S3. Phylogenetic tree reconstructions of coalescent-based inference (A) and maximum likelihood (B) using the Cactaceae591 dataset. Highly supported branches (BS/SH-aLRT > 95/80 for maximum likelihood inference and LPP > 0.95 for coalescent-based inference) are depicted with black circles; moderately supported branches (95 > BS and 80 < SH-aLRT for maximum likelihood inference. and 0.95 > LPP > 0.8) are depicted with gray circles. Coalescent-based supported branches among 0.8> LPP > 0.7 are indicated by numbers below nodes. Low-supported nodes (LPP < 0.7 and BS < 95) are shown with dashed lines. Incongruences recovered among highly supported clades are represented by red lines.

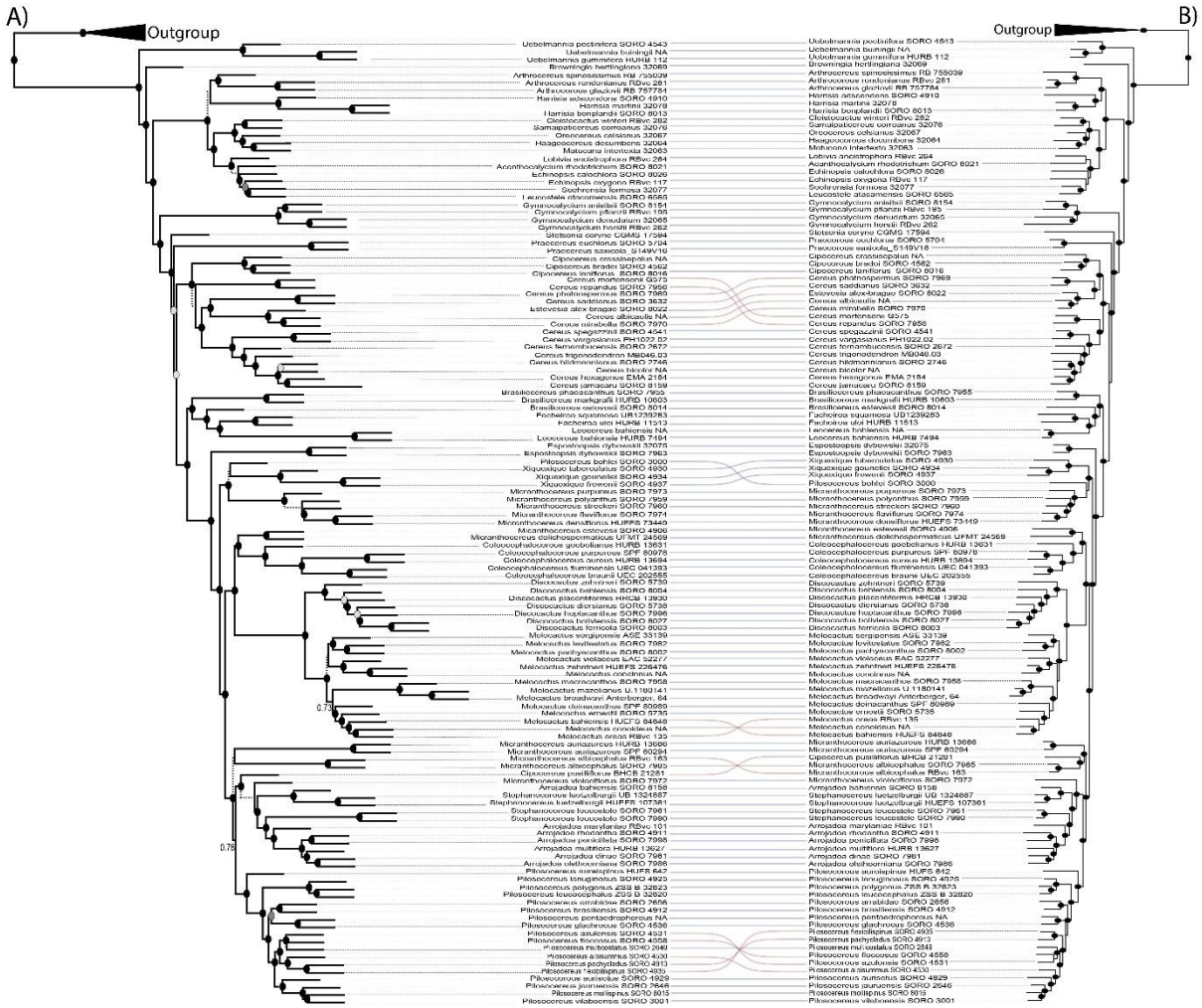
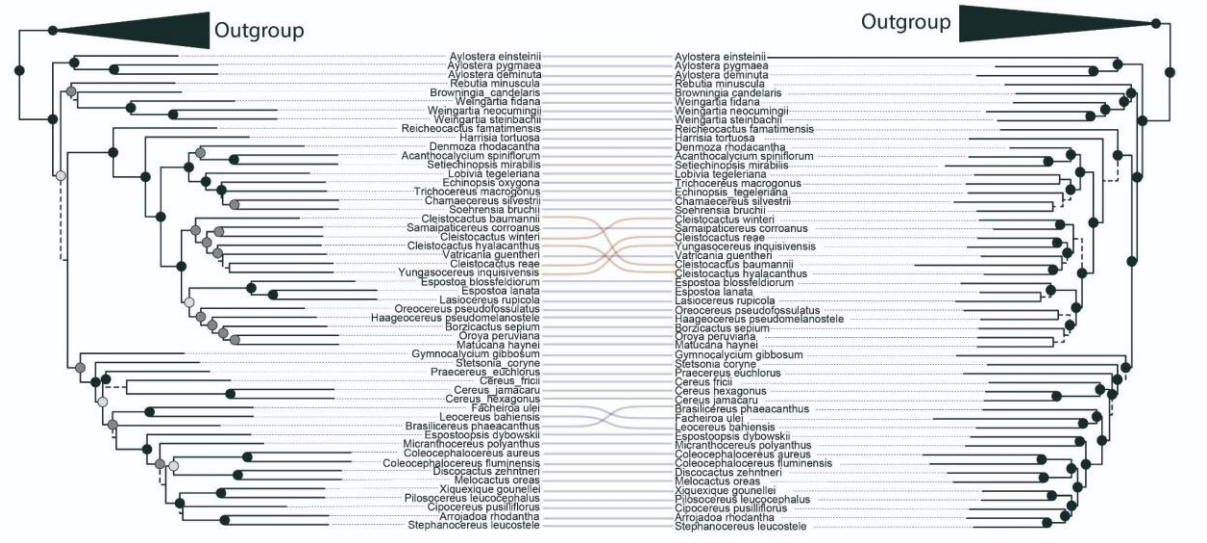


Figure S4. Phylogenetic tree reconstructions of coalescent-based inference (A) and maximum likelihood (B) using the Angiorperm353 dataset. Highly supported branches (BS/SH-aLRT > 95/80 for maximum likelihood inference and LPP > 0.95 for coalescent-based inference) are depicted with black circles; moderately supported branches (95 > BS and 80 < SH-aLRT for maximum likelihood inference. and 0.95 > LPP > 0.8) are depicted with gray circles. Coalescent-based supported branches among 0.8> LPP > 0.7 are indicated by numbers below nodes. Low-supported nodes (LPP < 0.7 and BS < 95) are shown with dashed lines. Incongruences recovered among highly supported clades are represented by red lines.



CHAPTER 3: Diversification of South American cacti: insights into diversification of arid and semiarid regions in Neotropics

Monique Romeiro-Brito[1], Milena C. Telhe[1], Flávia M. Lanna[2], Bryan C. Carstens[2], Luciano A. Digiampietri[3], Daniela C. Zappi[4], Nigel P. Taylor[5], Fernando F. Franco[1] and Evandro M. Moraes[1]

[1] Departamento de Biologia. Centro de Ciências Humanas e Biológicas. Universidade Federal de São Carlos (UFSCar). Sorocaba, Brazil.

[2] Department of Evolution, Ecology and Organismal Biology. Museum of Biological Diversity, The Ohio State University, Columbus, Ohio, USA

[3] Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo 03828-000, Brazil

[4] Programa de Pós-Graduação em Botânica, Instituto de Ciências Biológicas Universidade de Brasília (UNB), Brasília, Brazil.

[5] University of Gibraltar, Gibraltar Botanic Gardens Campus, Gibraltar.

Draft manuscript
To be submitted to: Journal of Biogeography

Abstract

The global expansion of arid and semiarid environments during the Miocene is the key biogeographic events associated with the diversification of succulent lineages around the globe. In addition to this event, life history and reproductive traits have been suggested as the key biotic drivers of the diversification of succulent lineages in the Neotropics. Here we focused on the tribe Cereeae (Cactaceae) to explore the role of biotic and abiotic traits in the diversification of succulent lineages in South America. For this, we estimated a calibrated tree using target capture sequence data to infer major biogeographic events and ancestral geographic range along this group. We also explored the contribution of biotic and abiotic traits in the diversification of this tribe using traditional diversification analysis and machine learning predictive models. The Cereeae diverged in the late Miocene, forming a widespread and continuous region from Central and Southern Andes through Espinhaço Mountain Range in eastern Brazil. Consecutive vicariant events isolated major lineages geographically in the Peruvian Andes, Southern Andes, and in the Espinhaço Mountain Range. The recolonization of eastern Brazil occurred through long-distance dispersal events in the late Miocene and Pliocene. We detected only one significant shift of diversification within this tribe, which comprises the globose cactus genera. The biotic traits associated with globose growth form and hummingbird pollination syndrome were the most important traits associated with high speciation rates in this clade. The findings of this study reveal a continuous distribution of Cereeae across the west-to-east region of South America during the late Miocene period. Following this, the vicariance events that led to the segregation of east and west lineages occurred concurrently with marine transgressions and the assembly of the Cerrado domain during the same period. Our findings suggest that morphological and ecological traits have a stronger influence on the diversification of Cereeae lineages than abiotic traits. Lastly, the results of this study emphasize the role of the Southern Andes region as a source of pre-existing dry-adapted plant lineages capable of colonizing new xeric environments towards the east during the Miocene period.

Keywords: SDTF, Cactaceae, biogeographic events, diversification rates, machine learning, abiotic and biotic traits, Dry Diagonal

3.1 Introduction

South America is home of the most diverse plant communities in the world, with its distribution being uneven across a range of diverse and contrasting biomes (Antonelli and Sanmartin 2011). Although the diversification of humid forests in this region has received the most attention, researchers have long been interested in investigating the origin and evolution of dry and arid biomes (Prado and Gibbs 1993, Prado 2000, Pennington et al. 2004, Pennington et al. 2009, and references therein). Several hypotheses have been proposed to explain the diversification of dry and arid regions in the Neotropics, such as the establishment of the Cerrado domain in the late Miocene (Simmon et al. 2009), and a widespread distribution of seasonally dry tropical forest (SDTF) during the Pleistocene (also known as the Pleistocene Arc Hypothesis, Prado and Gibbs 1993, Prado 2000). Moreover, the long-term isolation and in situ diversification of SDTF, xerophytic, and desert lineages since Neogene (Pennington et al. 2004, Pennington et al. 2010, Särkinen et al. 2012) allied with sporadic connections through long-distance dispersal (Gentry 1982) have been suggested. However, there is still limited knowledge about common patterns underlying the diversification of arid and dry regions in South America that could potentially provide the connection of dry-adapted lineages across the continent.

The South American Dry Diagonals (SADDs) comprise the dry and arid regions in South America that are embedded in two dry corridors, the eastern and western SADDs (eSADD and wSADD) (Luebert 2021). The eSADD includes open and semiarid domains, such as Caatinga, Cerrado, and Chaco (Werneck 2011), while the wSADD is composed of arid and hyperarid domains, such as Patagonia, Monte, Prepuna, Dry Puna, Atacama Desert, and Peruvian Desert (Houston and Hartley 2003, Abraham 2020). Despite the heterogeneity in climatic and edaphic conditions (Werneck 2011, Lima et al. 2018, Luebert 2021) these regions exhibit similarities in their fauna and flora (Morrone 2014, Neves et al. 2015, DRYFLOR et al. 2016, Prieto-Torres et al. 2019, Arango et al. 2020) and have been proposed as a geographic barrier that may have isolated wet forest lineages during the mid to late Miocene (Thomé et al. 2016, Ledo and Colli 2017, Thode et al. 2019, Masa-Iranzo et al. 2021, Lörch et al. 2021). Previous studies have suggested connectivity between SADD regions, assuming the Chaco and Cerrado domains as a linking route among seasonally dry tropical forest (SDTF) nuclei from wSADD to eSADD (Mogni et al. 2015). Rock outcrops and SDTF patches in Cerrado have been suggested as important stepping-stone dispersal routes and refuges for succulent plants,

particularly for those lineages restricted to eSADD (Amaral et al. 2021, Romeiro-Brito et al. 2023).

During the Miocene, the global cooling and the decline of CO_2 concentration led to the expansion of arid environments (Herbet et al. 2016). This environmental change played a crucial role in the expansion of grasslands and succulent plants on a global scale (Arakaki et al. 2011, Gutiérrez-Ortega et al. 2018). The expansion of aridity across South America was also confirmed by the decrease in sedimentation rate in the Central-Southern Andes in the late Miocene and Pliocene (Amidon et al. 2017), and was the main climatic event associated with the diversification of the biota inhabiting SADDs (Hurbath et al. 2021, Amaral et al. 2021, Griotti et al. 2023). Moreover, the geographic isolation of SDTF nuclei dates to Pliocene, resulting in *in situ* diversification promoted by topographic and edaphic heterogeneity in South America (Queiroz et al. 2017, Fernandes et al. 2022, Pérez-Escobar et al. 2022). However, no study has yet tested whether the periods of aridification would promote the formation of a recursive dry corridor connecting South American SDTF nuclei, as previously proposed in the influential article by Prado & Gibbs (1993). Changes in connectivity over time for SDTFs may have occurred through mechanisms such as range expansions (Almeida et al. 2018), jump dispersal events (Hurbath et al. 2021), and biome shifts between adjacent regions in the Miocene. These processes may have ultimately contributed to the geographic structure of SDTFs by the Pliocene (Fernandes et al. 2022).

While global aridification was the primary trigger of the diversification of dry-adapted plant lineages, different morphological strategies have played an important role in the establishment and promotion of diversity in harsh environmental conditions (Folk et al. 2020, Fagundes et al. 2020, Oliveira et al. 2020). Indeed, the combination of biotic and abiotic traits may have contributed to the rapid and recent diversification of major succulent lineages in the Neotropics (Hernández-Hernández et al. 2014). In the meantime, growth forms and reproductive traits were the main morphological and ecological drivers related to increasing diversification (e.g Hernández-Hernández et al. 2014, Tripp and Tsai 2017, Alcantara et al. 2018, Larocca et al. 2022) and biome shift among dry and semiarid regions in SDTF lineages in Neotropics (Amaral et al. 2021). The use of machine learning predictive models (e.g. random forest method) may be a potential approach to screen a wide range of traits yet to be explored in diversification shift analysis. Recent studies have shown the efficacy of these models to identify traits predicting evolutionary and ecological outcomes (Ruffley et al. 2019, Sullivan et al. 2019, Bonatelli et al. 2022, Lanna et al. 2022).

Cactus species offer a potential biological model for exploring patterns of diversification in the dry and arid environments of South America, especially those taxa belonging to the tribe Cereeae (*sensu* Romeiro-Brito et al. in prep). This tribe comprises a third of the major diversity of the subfamily Cactoideae (Hernandez-Hernandez et al. 2011) and is mainly distributed in South America, with few genera occurring in the Caribbean and Mesoamerica regions. This group has a disjunct distribution in the Southern Andes and Eastern Brazil, mainly associated with xerophytic forests and rocky outcrops in South America, and exhibits one of the highest diversification rates among Cactaceae, potentially due to climatic, morphological, and ecological factors (Hernández-Hernández et al. 2014). Therefore, screening for biotic and abiotic traits that predict the diversification of cacti from South America becomes a fundamental approach to understanding adaptive strategies of taxa inhabiting xeric formations in these regions.

Our study aimed to explore how biotic and abiotic traits contribute to the diversification in xeric habitats of South America. To achieve this, we employed a trait-dependent diversification analysis and predictive models using machine learning techniques. Our investigation centered on the cactus species from the tribe Cereeae as a biological model. Our approach required: (i) a well-supported topology of the tribe (Romeiro-Brito et al. in prep) to infer a time-calibrated tree and major biogeographical events, (ii) the identification of shifts in diversification rates and their correlation with specific biotic and abiotic traits in this group. We hypothesized that the global cooling and expansion of aridity have led to a wider distribution of this tribe and promoted connections of both SADDs in the Miocene. The isolation of xerophytic lineages in the SADDs might have occurred due to the establishment of the Cerrado domain and marine introgressions in the Miocene and Pliocene transition, reaching adjacent xerophytic forest by sporadic long-distance dispersal events. We also assume that the predictive models may show the combination of biotic and abiotic traits explaining shifts in the diversification rates in tribe Cereeae.

3.2 Material and Methods

3.2.1 Sampling and Divergence Time Estimation

We used the assembled dataset from a target capture sequencing from previous studies with tribe Cereeae (Romeiro-Brito et al. in prep). We selected a subsampling from the subtribe Cereinae aiming for a representative taxonomic and geographic sampling of each genus of this

group (Table S1). We included 15 taxa as outgroups from subfamilies Cactoideae, Opuntioideae, and Pereskioideae (Table S1).

Divergence time was estimated in MEGA X (Kumar et al. 2018), using the RelTime method (Tamura et al. 2012). This method is a computationally feasible divergence time analysis for large genomic datasets and has been shown to outperform other rapid dating methods (Barba-Montoya et al. 2021, Costa et al. 2022) or present similar results estimated by Bayesian approaches, as implemented by MCMCTree (e.g. Mello et al. 2017) and BEAST (e.g. Pie et al. 2018). Briefly, this method converts the branch lengths of a tree according to substitution sites into an ultrametric tree with branch-length relative time rates (Tamura et al. 2012).

We used as input for divergence time analysis the phylogenetic tree estimated in ASTRAL and concatenated sequences of 459 loci from the Cactaceae591 dataset in Romeiro-Brito et al (in prep). We inferred a coalescent-based phylogenetic tree using ASTRAL v5.7 (Sayyari and Mirarab 2016). The gene trees used for the inference of the ASTRAL species tree were estimated using IQ-TREE v2 (Mihn et al. 2020) with 1,000 ultrafast bootstrap replicates (Hoang et al. 2018) and rooted using the pxrr program available in phyx (Brown et al. 2017). We set the GTR+I+G as the substitution model. Due to the absence of fossil records in Cactaceae, we assumed divergence time inferred for major clades of this family by Hernández-Hernández et al. (2014). We used a normal distribution of the 95% HPD of the crown age of subfamily Cactoideae without *Blossfeldia* Werderm (21.9 - 10.9 million years ago (MyA)), setting the mean age to 15.3 MyA.

3.2.2 Biogeographic Analysis

We defined the operational biogeographic areas using *Infomap Bioregions v2* (https://www.mapequation.org/bioregions2/). This program is a clustering network algorithm that generates geographic boundaries based on species occurrence records (Edler et al. 2017). The geographic occurrences for all species from tribe Cereeae were retrieved using GBIF (http://www.gbif.org) and speciesLink (https://specieslink.net/search/) databases. We filtered geographic records using the CoordinateCleaner R package (Zizka et al. 2019) to remove duplicate, spurious, and non-terrestrial records. We mapped species occurrence in South America and inspected the coverage of geographic records collected across the known distribution for each taxa (according to Barthlott et al. 2015). We removed from the final

species occurrence dataset all geographic records outside of each species' known distribution and without a voucher specimen to verify their identity. We used the settings for minimum and maximum cell capacity among 10 to 80 geographic records, respectively, and minimum and maximum cell size among 2 to 8 degrees. The ancestral range distribution of tribe Cereeae was reconstructed using BioGeoBEARS (Matzke 2013) R package. This analysis compares and tests for the best biogeographic model (among the DIVALIKE, DIVALIKE+J, DEC, DEC+J, BAYAREALIKE, and BAYAREALIKE+J models) using the major marginal likelihood (LnL) values, p-value, and Akaike weights. For this analysis, we used the time-calibrated tree inferred in MEGAX and the biogeographical units recovered by InfoMap Bioregions.

3.2.3 Time-dependent diversification analysis

We used the time-dependent diversification analysis implemented in BAMM v 2.5 (Rabosky et al. 2014) to identify shifts in diversification rates in the tribe Cereeae and in the subtribe Cereinae, our most comprehensive sampled clade of the tribe Cereeae. The priors set in the control file were estimated using the "setBAMMpriors" function in BAMMtools R package (Rabosky et al. 2014), using the MEGAX calibrated-tree. We accounted for incomplete sampling of genera and clades using a sampling fraction based on total species number of each clade according to the species checklist of Korotkova et al. (2021). The sampling fraction is detailed in Table S2. The analysis consisted of two runs of 30 million generations of rjMCMC (reversible jump Markov Chain Monte Carlo), using four chains, and sampling a shift configuration every 5000 generations. We assessed the convergence by effective sampling size (ESS) values greater than 200 and applied a burnin of 10% to infer the best shift configuration.

3.2.4 Trait-dependent diversification analysis

Due to the uneven sampling between subtribe Cereinae and Trichocereinae in this study, we limited the trait-dependent analysis to subtribe Cereinae. We modeled the impact of traits on the diversification of subtribe Cereinae, especially the association of these traits on speciation, extinction, and transition rates using the model BiSSE available in the diversitree R package (Fitzjohn 2012). We chose three binary characters for this analysis previously associated with higher diversification rates in tribe Cereeae (Hernández-Hernández et al.

2014): presence or absence of cephalia, pollination by hummingbirds, bats, or moths, and globose and columnar growth forms (trait code detailed in Table S3).

We evaluated eight BiSSE models with increasing complexity, permitting the combination of speciation, extinction, and transition rates to vary or keep equal among the three traits (detailed in Table S4). The best-fitted model was selected based on the highest marginal likelihood value and statistical significance compared to the null model (no variation in diversification rates). We estimated the diversification and extinction parameter using a MCMC run with the best model, running 300.000 generations. We applied a burnin of 25% of this run to obtain the confidence intervals for diversification and extinction parameters.

3.2.5 Inference of biotic and abiotic traits predicting shifts in the diversification rates using machine learning

We investigated the association of biotic and abiotic traits with diversification rates in the subtribe Cereinae using supervised machine learning. We aimed to identify which traits could predict diversification rates using the random forest regression method. Random forest (Liaw and Wiener, 2002) has recently been used to build predictive models for ecological and genetic studies (Espíndola et al. 2016, Costa et al. 2018, Ruffley et al. 2019, Bonatelli et al. 2022, Lanna et al. 2022). The method offers the advantage of avoiding preconceived assumptions and has shown superior performance to models commonly used in ecological and evolutionary research (Fox et al. 2017, Schrider and Kern 2018).

Predictive analyses were performed with WEKA v3.9 (Kaufmann 2016) using the random forest regression method allied with cross-validation k-fold method. In all models, each instance was composed of species of subtribe Cereinae presented in the calibrated-tree estimated in this study. The label (response variable) is the diversification rates and the attributes (predictive variables) are the biotic and abiotic traits recovered for each species. We aimed to verify the impact of removing unimportant attributes for the model, using the selection attribute tool implemented by the method CfsSubsetEval (Hall and Smith 1997). Thus, we built two predictive models: 1) including all traits; and 2) including selected traits. We accessed the importance of each attribute using the mean decrease in the GINI index (Breiman 2002).

The diversification rates were extracted for each tip of subtribe Cereinae using the ‘getTipRates’ function in BAMMtools R package. The biotic traits of each taxa were obtained by literature review, including taxonomic descriptions (e.g. Barthlott and Hunt 2000, Taylor

and Zappi 2004, Hunt et al. 2006, Taylor and Zappi 2018) and regional floristic survey (e.g. Menezes et al. 2013, Menezes et al. 2015). The biotic database included information of 30 phenotypic traits, such as vegetative, floral, fruit, and seed information (detailed in Table S5). The abiotic traits were extracted using the final geographic records dataset obtained in the biogeographic analysis, including bioclimatic variables and solar radiation. Bioclimatic variables, solar radiation, and elevation were extracted from WorldClim v2 (Fick and Hijmans 2017) and ENVIREM (Title and Bemmels 2018). Biotic and abiotic traits used in this study are available in online supplementary material (https://figshare.com/s/15d17f5cfb7541e2c156).

3.3 Results

3.3.1 Divergence time estimation and biogeographical reconstruction

We estimated the crown age of the tribe Cereeae at 7.18 MyA (95% CI = 10.77 - 5.79 MyA, Table 1, Fig. S1), recovering the Messian stage at late Miocene as the starting age of diversification of this group. The divergence time estimated of major clades of Cereeae was recovered from the late Miocene (subtribe Cereinae) until Pliocene and Pleistocene (subtribes Uebelmaniinae, Gymnocalycinae, and Trichocereinae, Table 1, Fig. 1). Most genera of subtribe Cereinae diverged during the Pliocene (*Cereus*, *Pilosocereus*, *Coleocephalocereus*, *Arrojadoa*), reaching Pleistocene in some groups (*Discocactus* e *Melocactus*, Fig. 2). Overall, we recovered older crown ages than previous divergence time estimates in Cactoideae and tribe Cereeae (Table 1 and Fig. S1, but see results Silva et al. 2018), and in some cases, narrower confidence intervals.

Table 1. Divergence time (in MyA) estimated in previous and the present study. The divergence time corresponds to crown ages, including the mean and confidence interval (CI) of 95% in parenthesis.

Clades	**Hernández-Hernández et al. (2014)**	**Silva et al. (2018)**	**Lavor et al. (2018)**	**Amaral et al. (2021)**	**Romeiro-Brito et al. (2023)**	**This study**
Subfamíly Cactoideae[1]	15.27 (21.85-10.94)	-	-	-	-	15.97 (20.18-12.46)
Tribe Cacteae	11.94 (17.27 - 8.33)	-	-	-	-	10.53 (14.37-7.72)
Tribe Rhipsalideae	7.67 (11.82 - 4.26)	16.3 (26.3–6.02)	-	-	-	9.53 (13.8-6.58)
Tribe Cereeae[2]	6.58 (9.67-4.34)	24.5 (32.9–13.0)	-	5.16 (7.92-4.7)	7.39 (9.73-5.02)	7.18 (10.77-4.79)
Subtribe Uebelmaniinae	-	-	-	-	-	3.51 (5.62-2.19)
Subtribe Gymnocalyciinae	5.08 (7.55-3.09)	7.23 (13.1–1.64)	-	-	-	3.28 (5.59-1.93)
Subtribe Trichocereinae	6.12 (8.93-3.93)*	15.1 (19.4–8.01)	-	-	-	4.73 (7.8-2.87)
Subtribe Cereinae	-	13.4 (19.5–7.65)	-	-	-	5.58 (9.9-3.15)
Cereus+Cipocereus	-	-	-	3.78 (5.83-3.31)	-	4.92 (8.95-2.7)
Cereus	-	-	-	3.67 (5.65–3.22)	-	4.83 (8.89-2.63)
Pilosocereus	-	-	2.7 (4.71-1.17)	-	2.09 (3.16-1.31)	3.47 (7.38-1.63)
Coleocephalocereus+Discocactus+Melocactus	-	6.88 (10.2–2.40)	-	-	-	4.01 (7.92-2.03)
Discocactus+Melocactus	-	3.81 (6.77–1.07)	-	-	1.24 (2.43-0)	2.9 (5.8-1.45)
Xiquexique+Micranthocereus	-	-	-	-	-	4.06 (7.81-2.11)
Xiquexique	-	-	1.55 (3.60-0.31)	-	2.96 (4.57-1.63)	2.75 (5.63-1.34)
Arrojadoa	-	-	-	-	-	3.81 (7.87-1.8)
Facheiroa	-	-	-	-	-	4.98 (9.18-2.71)

[1] crown age after Blosffeldia split; [2] sensu Nyffeler and Eggli, 2010 * Subtribe Trichocereinae with *Espostoopsis*

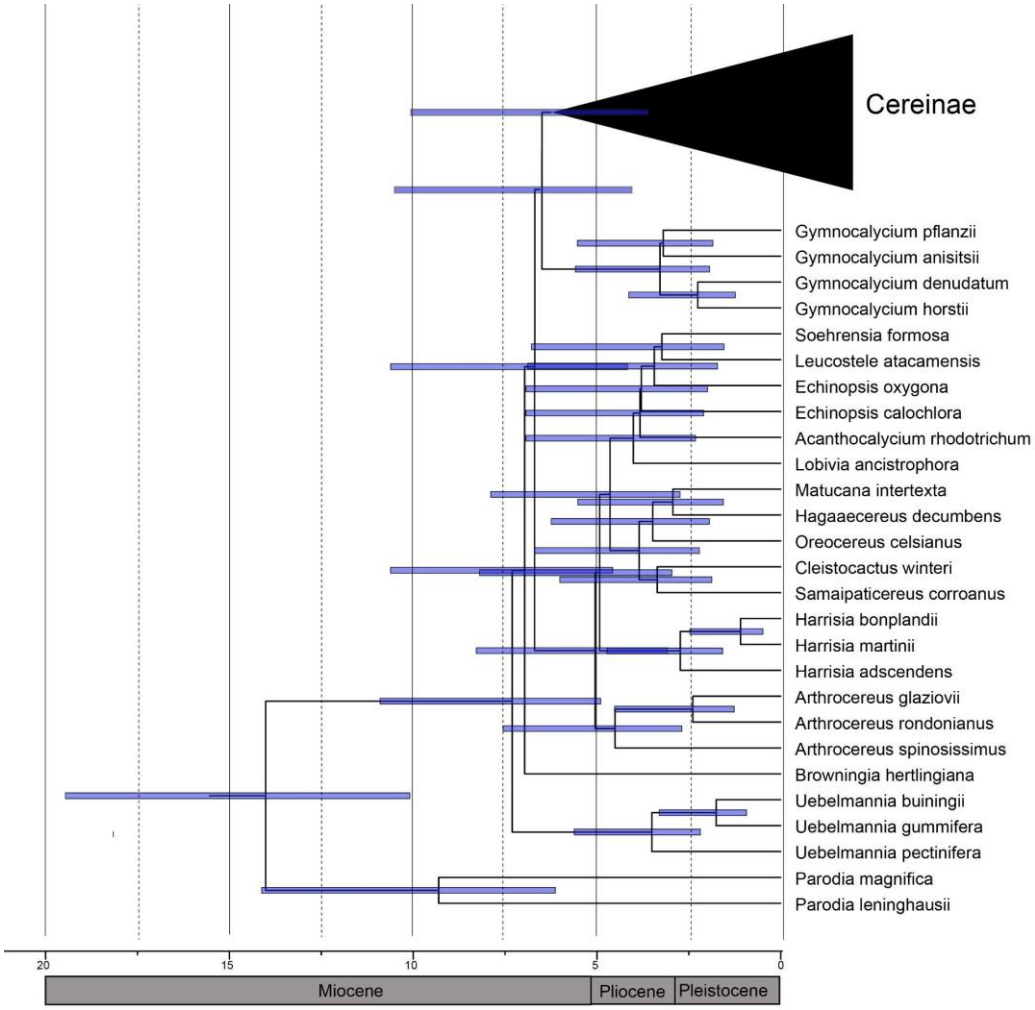


Figure 1. Time-calibrated tree of tribe Cereeae using the RelTime method in MEGAX. Node bars correspond to 95% confidence interval estimates.

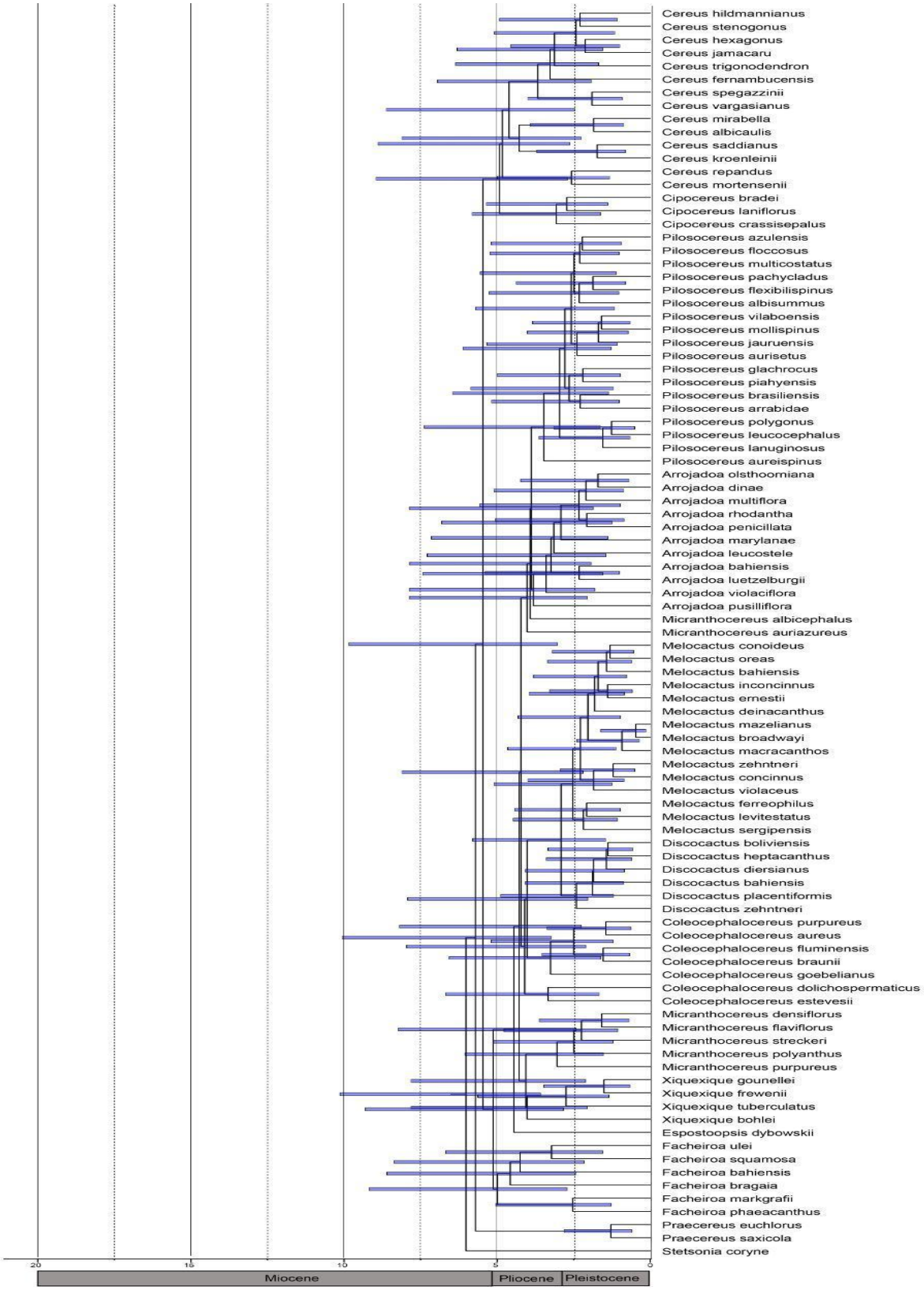


Figure 2. Time-calibrated tree of subtribe Cereinae using the RelTime method in MEGAX. Node bars correspond to 95% confidence interval estimates.

The clustering network analysis recovered 10 bioregions (Fig. 3A). The bioregions consist roughly of domains and ecoregions delimited by previous bioregionalization studies (Morrone 2014, Dinerstein et al. 2017): bio1: Caatinga and Northern Atlantic Forest; bio2 = Southern Andean region (including Chaco and Pampas domains, and Southern Andes); bio3 = Espinhaço Mountain Range; bio4 = Southern coast Atlantic Forest; bio5 = Southern interior Atlantic Forest; bio6 = Cerrado; bio7 = Mesoamerica; bio8 = Peruvian Andes; bio9 = Northern South America; bio10 = Amazon Tropical Forest.

DIVALIKE+J was the best model recovered by BioGeoBEARS (LnL = -292.7, Table S6). The ancestral range of tribe Cereeae was a widespread area comprising the Southern Andes (including Chaco), Peruvian Andes, and Espinhaço Mountain Range during the Messian stage in the late Miocene (bio3, bio8, and bio2, respectively, Fig. 3B). Each of these bioregions was isolated by successive vicariant events. The main biogeographical event responsible for the colonization of dry areas within the subtribes Cereinae and Trichocereinae was long-distance dispersal. The Southern Andes bioregion was the primary source of dispersal towards Caatinga bioregion in the subtribe Cereinae, and towards Peruvian Andes and Espinhaço Mountain Range in the subtribe Trichocereinae during the Miocene and Pliocene transition. Long-distance dispersal events also occurred during Plio-Pleistocene transition within the subtribe Cereinae from Caatinga towards surrounding dry areas in the Neotropics, such as the SDTF nuclei in Mesoamerica and the Caribbean region, dry areas within Amazonia, the Espinhaço Mountain Range, and rock outcrops in the Cerrado and Atlantic Forest. Likewise, the subtribe Tricocereinae colonized the Peruvian Andes, the Espinhaço Mountain Range, the Caatinga, and the Caribbean region.

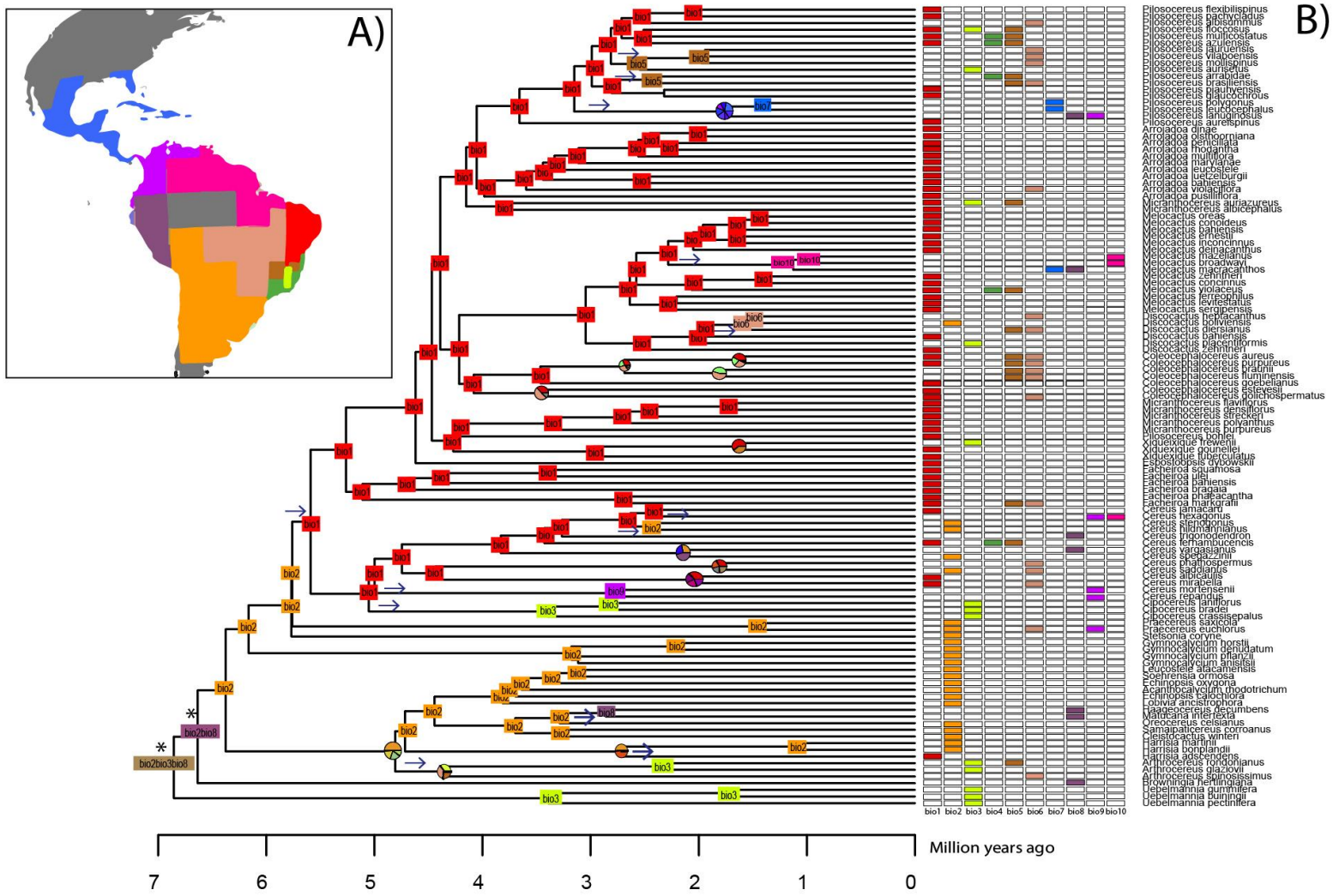


Figure 3. Biogeographic reconstruction analysis of the tribe Cereeae. A) Ten bioregions identified by clustering network analysis (Infomap Bioregions, Edler et al. 2017). B) Ancestral range distribution of tribe Cereeae reconstructed by DIVALIKE+J inferred by BioGeoBEARS. Asterisks indicate vicariance events in the backbone of the phylogeny. Blue arrows indicate long-distance dispersal events within subtribes Trichocereina and Cereinae.

3.3.2 Diversification analyses and predictive models

No significant shift in diversification rates was recovered in tribe Cereeae with high posterior probability (Fig. S2). Despite this, other shift configurations were recovered with low posterior probabilities, including two shifts in the tribe Cereeae, one after the *Uebelmannia* divergence and another after the *Browningia* divergence, and one shift in the *Melocactus+Discocactus* clade (Fig. S2). Focusing only in the subtribe Cereinae, the shift in the diversification rates in the *Melocactus+Discocactus* clade (Fig.4) was significantly recovered with the highest posterior probability. The next shift configuration with high posterior probability in subtribe Cereinae shows two shifts, one in the *Melocactus+Discocactus* clade, and the other in the Cereinae clade after *Praecereus* split.

The trait-dependent diversification analysis recovered with significant statistic values only two characteristics associated with shifts in diversification rates in subtribe Cereinae: globose growth form and hummingbird pollination syndrome (Table S4, Fig. 4). These traits coincide with pollination syndrome in the genus *Melocactus*, and with the growth form presented in *Melocactus+Discocactus* clade.

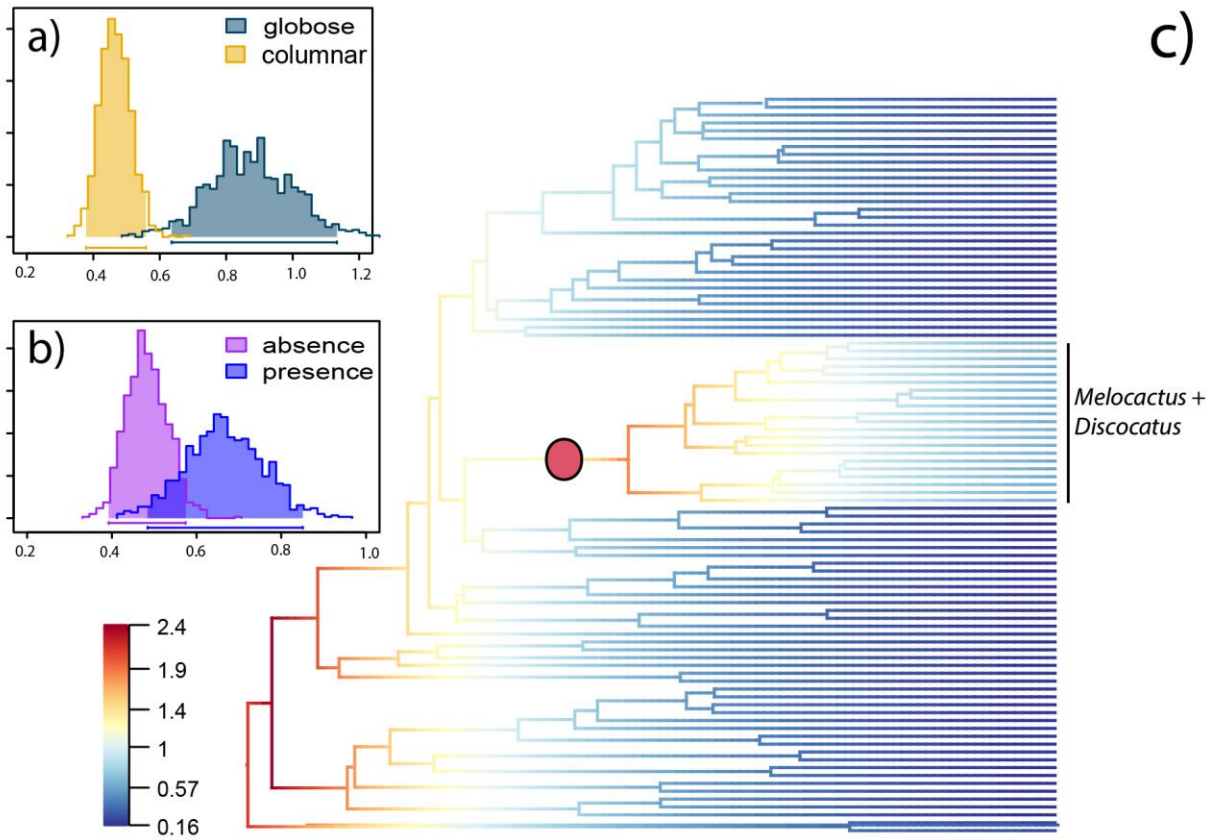


Figure 4. Diversification analysis of subtribe Cereinae. a) speciation rate related to globose and columnar growth form and b) to presence and absence of hummingbird pollinator syndrome estimated with 95% confidence interval in BISSE model. c) Model with the highest posterior probability recovered in BAMM, showing the shift in diversification rate in the *Melocactus+Discocactus* clade.

The Random Forest model estimated biotic traits related to vegetative and spine traits, including growth form, plant height, cladode diameter, presence of cephalia, spine number, and spine length as the most important attributes to predict diversification rates in the subtribe Cereinae (Table 2). Overall, biotic traits had shown to be more relevant than abiotic traits to predict diversification rates. Only one abiotic trait was ranked as the most important, involving potential evapotranspiration. Both models (with all traits and with selected traits) corroborate trait-dependent diversification analysis, retrieving growth form and pollination syndrome as the most important traits to predict diversification rates in the subtribe Cereinae.

Table 2. Ten most important attributes to predict diversification rates in subtribe Cereinae according to mean impurity decrease (GINI index) estimated with random forest regression Model for all and selected traits. Abiotic trait is highlighted in red. PET: potential evapotranspiration.

All traits		Selected traits	
Traits	**GINI**	**Traits**	**GINI**
Growth form	0.36	Growth form	0.33
Minimum diameter of the fruit	0.32	Minimum plant height	0.13
Maximum plant height	0.32	Presence of cephalium	0.11
Presence of cephalium	0.27	Maximum diameter of the cladode	0.1
Minimum diameter of the cladode	0.24	Minimum number of central spines	0.08
Maximum diameter of the cladode	0.24	Scales in floral hypanthium	0.06
Maximum length radial spines	0.22	mean PET in most humid quarter	0.05
Pollinators	0.2	Maximum length radial spines	0.05
Minimum plant height	0.2	Remnant floral persistent	0.04
Minimum number of central spines	0.18	Flowering position	0.04

3.4 Discussion

3.4.1 Diversification of tribe Cereeae during Miocene and Pliocene

Our findings on divergence time are consistent with earlier research on Cactaceae. However, our study indicates that the major lineages in this group have relatively older ages than previously reported. The tribe Cereeae diverged in the late Miocene occurring in a widespread ancestral range that connected central and Southern Andes through Southeastern Brazil in the Espinhaço Mountain Range. These results are in line with the global cooling and expansion of aridity during the late Miocene (Herbert et al. 2016) and with the rise of diversification of succulent lineages (Arakaki et al. 2011, Hernández-Hernández et al. 2014). Moreover, a broader ancestral range connecting Central and Southern Andes through Southeastern Brazil in Espinhaço Mountain Range corroborates an older connection of SDTF nuclei dated in the late Miocene (Côrtes et al. 2015, Queiroz et al. 2017, Hurbath et al. 2021), resembling the Arc Hypothesis distribution. The Arc hypothesis predicted a widespread distribution of SDTF's nuclei dated in the Pleistocene (Prado and Gibbs 1993, Prado 2000, Werneck et al. 2011), a pattern that has been confirmed for different SDTF's fauna lineages (Werneck et al. 2011, Corbett et al. 2020). The connection of SDTF nuclei in South America at

earlier and different periods suggests the existence of a recurrent, permeable corridor that formed during colder and drier periods. This corridor emerged as a result of climatic fluctuations that occurred from the late Miocene to Pleistocene climatic oscillations.

Although the connection between SDTF nuclei may persist in drier periods in Neogene and Quaternary, we only recovered a widespread connection of those regions in the backbone of the phylogeny. Successive vicariance events segregated major Cereeae lineages in Espinhaço Mountain Range, Chaco, Pampas and Southern Andes, and the Peruvian Andes during the late Miocene. The final uplift of the Andes during this period created a physical barrier between the Peruvian Andes and the Southern Andes, while marine incursions isolated the Andes regions from Southeastern Brazil (Pérez-Escobar et al. 2022, Hoorn et al. 2022). Additionally, the rise of grasslands and the development of the fire-prone Cerrado savanna, which is not suitable for succulent plant lineages, may have served as a geographical barrier for these groups (Queiroz et al. 2017). However, recent studies have identified the Cerrado, in a broad sense, as an important domain for the diversification of Cereinae lineages, such as the genera *Cereus* (Franco et al. 2017, Amaral et al. 2021) and *Pilosocereus* (Lavor et al. 2018, Romeiro-Brito et al. 2023). However, these lineages are found in rocky outcrops within the Cerrado, which act as refuges for succulent plants, protecting them from seasonal fires.

The ancestral range of the main diversity of the tribe Cereeae, including subtribes Trichocereinae, Gymnocalycinae, and Cereinae (*sensu* Romeiro-Brito in prep.), was restricted to the Southern Andes bioregion in the late Miocene. The Southern Andes bioregion is composed of the Southern Andes, Chaco and Pampas, indicating a close relationship of Chaco and Pampas from eSADD to wSADD's domains. Only after the Mio-Pliocene transition that long-distance dispersal promoted colonization of the Caatinga region from the Southern Andes. The Southern Andes bioregions englobe domains that were previously suggested by fauna, floristic and morphoclimatic similarities as linking edge to Caatinga domain, including Chaco (Ab'Saber 2000), Chiquitan and Missiones (Dryflor 2016, Queiroz et al. 2017, Prieto-Torres et al. 2019) and sub-Andean regions (Sarmiento 1975, Prieto-Torres et al. 2019). Indeed, our results highlight that long-distance dispersals during the late Miocene were the main events linking western to eastern South America STDF lineages (Queiroz et al. 2017, Almeida et al. 2018, Magalhães et al. 2019, Hurbarth et al. 2021).

Our results suggest long-term isolation and *in situ* diversification of Cereinae and Trichocereinae lineages during the Pliocene in both eSADD and wSADD, followed by long-distance dispersal events to surrounding dry areas during the Pliocene and Pleistocene. The

diversification of several genera in tribe Cereeae took place during the Plio-Pleistocene period (e.g. Franck et al. 2013, Franco et al. 2017, Amaral et al. 2020, Romeiro-Brito et al. 2023) and might have been a fundamental period of diversification of xerophytic lineages in the Neotropics (Jaramillo et al. 2020). The *in situ* diversification of those lineages during the Plio-Pleistocene might be a byproduct of a second record of aridity expansion during the late Pliocene (Amidon et al. 2018), promoting the rise of heterogeneous topographic and edaphic conditions (Fernandes et al. 2022, Pérez-Escobar et al. 2022). Moreover, the *in situ* diversification might have been fostered by climatic instability during the Pleistocene in the Neotropics (Costa et al. 2018), which also corroborates the occurrence of dispersal events in the genera *Harrisia*, *Pilosocereus*, *Cereus*, and *Melocactus* toward Caribbean and Mesoamerican regions (Franck et al. 2013, Franco et al. 2017, Amaral et al. 2020, Romeiro-Brito et al. 2023). The dispersal of succulent lineages may have been facilitated by the expansion of a dry coastal corridor (Nace et al. 2014) or central Amazonia dry corridor during the Last Glacial Maximum (Sato et al. 2021). Both corridors promoted the spread of grasslands and may have been intermittent corridors for succulent lineages during this period. These events should be considered in future niche modeling to test both corridors as connecting South America and Mesoamerica for succulent lineages.

A remarkable result of our biogeographic reconstructions relates to multiple dispersal events towards the Espinhaço Mountain Range. The Espinhaço Mountain Range, the longest pre-Cambrian orogenic belt of Brazilian territory, is predominantly characterized by the *campo rupestre* landscape, which is a unique mosaic vegetation of the old and stable montane region in southeastern Brazil (Silveira et al. 2016). This region is known to harbor several endemic plant groups (Silveira et al 2016), with the most endemic cactus representatives from the subtribe Cereinae (Zappi and Taylor 2008). The lineages of the tribe Cereeae might have reached the Espinhaço Mountain Range through different routes. While the genus *Uebelmannia* (subtribe Uebelmanniinae) was isolated from western SADD due to a vicariance event in the late Miocene, the genus *Arthrocereus* (subtribe Trichocereinae) dispersed from Southern Andes likely through a southern route, while some lineages of the subtribe Cereinae dispersed from Caatinga domain during the Plio-Pleistocene transition. Overall, our findings are consistent with previous studies that indicate recent radiation of plant lineages in *campos rupestre* (e.g. Vasconcelos et al. 2020).

3.4.2 Biotic traits as predictors of high diversification rates in dry areas

Although several lineages in the subtribe Cereinae diversified during the expansion of aridity in the late Pliocene, only the clade comprising globose cacti exhibited a shift in diversification rate. Hernández-Hernández et al. (2014) identified growth form and pollinator syndrome as the primary traits associated with high diversification rates in Cactaceae, a finding supported by both trait-dependent diversification analyses and predictive modeling results in our study. Interestingly, our predictive models indicate that biotic rather than abiotic traits are more strongly associated with diversification rates, as only one abiotic trait was found among the most important attributes in both models. This result corroborates the premise that aridity alone cannot fully account for the increase in diversification rates in Cactaceae (Hernández-Hernández et al. 2014). Similar associations between biotic and abiotic traits have been documented in other dry-adapted lineages, such as the genus *Crassula* (Lu et al. 2021), suggesting that life-history traits from such lineages are often influenced by environmental conditions (Ogburn and Edwards 2015). Nevertheless, our study emphasizes the need to explicitly distinguish between abiotic and biotic traits when analyzing diversification rates, either through trait-dependent analysis testing the role of environmental factors (e.g., Lagomarsino et al. 2016, Condamine et al. 2019, Palazzessi et al. 2022) or through the use of predictive machine learning models, as we did here.

The higher diversification rates observed in the globose Brazilian clades are an interesting finding in this study. The trait-dependent diversification and predictive models indicate that the globose growth form is associated with higher diversification rates, a result that is consistent with a previous study detecting a shift in diversification rates in this clade (Hernández-Hernández et al. 2014). Interestingly, this contradicts the pattern suggested by Hernández-Hernández et al. (2014), which proposed a higher diversification rate in columnar and arborescent growth forms within the broader context of Cactaceae. The reversive transition to a globose growth form in the Cereeae tribe may have facilitated the diversification and dispersal of these lineages in the Cerrado domain, as the compact form is believed to aid in surviving fire episodes, with globose cacti avoiding damage due to cooler soil temperatures (Taylor and Zappi, 2004). In fact, some species in this clade are encountered buried in sandy soil of the Cerrado and Espinhaço Mountain Range (e.g. *Melocactus concinnus* and most species of the genus *Discocactus*). Furthermore, the globose-growth form may be a strategy to prevent damage from desiccation and insolation due to lower surface area exposure compared to higher plants (Lu et al. 2021).

The diversification rates were also associated with hummingbird pollination, a syndrome present in the genus *Melocactus*. Many species-rich plant families are pollinated by hummingbirds (Abrahamczyk and Steudel 2022), and also present a higher diversification rate compared to clades pollinated by other groups (Givnish et al. 2014, Lagomarsino et al. 2016, Serrano-Serrano et al. 2017). However, we should take caution regarding the hummingbird pollination syndrome as the main driver of diversification in these groups. Recent studies related to the diversification of hummingbird-pollinated plant lineages have shown a more complex diversification pattern involving topographic complexity and allopatric speciation (Kessler et al. 2020). In fact, the shift of diversification rate is identified in the *Melocactus+Discocactus* clade, the first being pollinated by hummingbirds and the second by moths. Despite the main pollinators of *Melocactus* species being hummingbirds, recent studies have recovered a variety of effective pollination syndromes in this group, including pollination by lizards (Gomes et al. 2014), bees, and ants (Fagua and Ackerman 2011). Moreover, the number and specialization of pollinators of cacti groups have shown to be also related to environmental gradient conditions (Freilij et al. 2023), already suggesting a more complex diversification scenario in cacti lineages. Thus we should consider testing a broader scenario of diversification involving the variety of derived and specialized pollinators (such as birds, bats, and moths) against bees or insect pollinators in tribe Cereeae for diversification analysis in this group

3.5 Acknowledgments

We thank Gerardus Olsthoorn, Dr. Lidyanne Y. S. Aona, Dr. Diego R. Gonzaga, and the Coleção de Cactos e Suculentas do Instituto de Pesquisas Jardim Botânico do Rio de Janeiro for contributing to the cactus samples used in this study. We thank Isabel A. S. Bonatelli and Danilo T. Amaral for relevant discussions and improvements in this paper.

3.6 Funding resources

This study was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP: 2018/03428-5, 2019/03211-9). M.R.B., M.C.T. and M.K. were granted by FAPESP fellowships (2018/06937-8, 2019/11233-2 and 2022/04621-9, respectively).

3.7 Reference

Abraham E. M., Rodríguez M. D., Rubio M. C., Guida-Johnson,B., Gomez, L., and Rubio, C. (2020). Disentangling the concept of “South American Arid Diagonal”. Journal of Arid Environments, 175, 104089.

Abrahamczyk S. and Steudel B. (2022). Why are some hummingbird-pollinated plant clades so species-rich?. American Journal of Botany, 109, 1059-1062.

Ab'sáber A. N. (2000). Spaces occupied by the expansion of dry climates in south america during the quatenary ice ages. Revista do Instituto Geológico (Descontinuada), 21, 71-78.

Alcantara S., Ree R. H., and Mello-Silva R. (2018). Accelerated diversification and functional trait evolution in Velloziaceae reveal new insights into the origins of the campos rupestres’ exceptional floristic richness. Annals of Botany, 122, 165-180.

Almeida F. R., Amorim A. M. A. and van den Berg, C. (2018). Timing the origin and past connections between Andean and Atlantic Seasonally Dry Tropical Forests in South America: Insights from the biogeographical history of *Amorimia* (Malpighiaceae). Taxon, 67, 739-751.

Amaral D. T., Minhós-Yano I., Oliveira J. V. M., Romeiro-Brito M., Bonatelli I. A. S., Taylor N. P. et al. (2021). Tracking the xeric biomes of South America: The spatiotemporal diversification of Mandacaru cactus. Journal of Biogeography, 48, 3085-3103.

Amidon W. H., Fisher G. B., Burbank D. W., Ciccioli P. L., Alonso R. N., Gorin A. L et al. (2017). Mio-Pliocene aridity in the south-central Andes associated with Southern Hemisphere cold periods. Proceedings of the National Academy of Sciences, 114, 6474-6479.

Antonelli A. and Sanmartín I. (2011). Why are there so many plant species in the Neotropics?. Taxon, 60, 403-414.

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. Proceedings of the National Academy of Sciences, 108, 8379-8384.

Arango A., Villalobos F., Prieto-Torres D. A. and Guevara R. (2021). The phylogenetic diversity and structure of the seasonally dry forests in the Neotropics. Journal of Biogeography, 48, 176-186.

Barba-Montoya J., Tao Q. and Kumar, S. (2021). Assessing rapid relaxed-clock methods for Phylogenomic dating. Genome Biology and Evolution, 13, evab251.

Barthlott W., Burstedde K., Geffert J. L., Ibisch P. L., Korotkova N., Miebach A. et al. (2015). Biogeography and biodiversity of cacti. Universitätsverlag Isensee.

Barthlott W. and Hunt D. (2000). Seed diversity in the Cactaceae, subfamily Cactoideae (Succulent Plant Research vol. 5). Sherborne: David Hunt, 173.

Breiman L. (2002). Manual on setting up, using, and understanding ran- dom forests v3. 1. Berkeley, CA: Statistics Department University of California Berkeley

Brown J. W., Walker J. F. and Smith S. A. (2017). Phyx: phylogenetic tools for unix. Bioinformatics, 33, 1886-1888.

Bonatelli I. A., Gehara M., Carstens B. C., Colli G. R. and Moraes E. M. (2022). Comparative and predictive phylogeography in the South American diagonal of open formations: Unravelling the biological and environmental influences on multitaxon demography. Molecular Ecology, 31, 331-342.

Condamine F. L., Rolland J. and Morlon, H. (2019). Assessing the causes of diversification slowdowns: temperature-dependent and diversity-dependent models receive equivalent support. Ecology letters, 22(11), 1900-1912.

Corbett E. C., Bravo G. A., Schunck F., Naka L. N., Silveira L. F. and Edwards, S. V. (2020). Evidence for the Pleistocene Arc Hypothesis from genome-wide SNPs in a Neotropical dry forest specialist, the Rufous-fronted Thornbird (Furnariidae: Phacellodomus rufifrons). Molecular ecology, 29, 4457-4472.

Côrtes A. L. A., Rapini A. and Daniel, T. F. (2015). The *Tetramerium* lineage (Acanthaceae: Justicieae) does not support the Pleistocene Arc hypothesis for South American seasonally dry forests. American Journal of Botany, 102, 992-1007.

Costa G. C., Hampe A., Ledru M. P., Martinez P. A., Mazzochini G. G., Shepard D. B. et al. (2018). Biome stability in South America over the last 30 kyr: Inferences from long-term vegetation dynamics and habitat modelling. Global Ecology and Biogeography, 27, 285-297.

Costa F. P., Schrago C. G. and Mello B. (2022). Assessing the relative performance of fast molecular dating methods for phylogenomic data. BMC genomics, 23, 798.

Dryflor, Banda-R K., Delgado-Salinas A., Dexter K. G., Linares-Palomino R., Oliveira-Filho A. et al. (2016). Plant diversity patterns in neotropical dry forests and their conservation implications. Science, 353, 1383-1387.

Edler D., Guedes T., Zizka A., Rosvall M. and Antonelli A. (2017). Infomap bioregions: interactive mapping of biogeographical regions from species distributions. Systematic biology, 66, 197-204.

Espíndola A., Ruffley M., Smith M. L., Carstens B. C., Tank D. C. and Sullivan, J. (2016). Identifying cryptic diversity with predictive phylogeography. Proceedings of the Royal Society B: Biological Sciences, 283, 20161529.

Fagua J. C. and Ackerman, J. D. (2011). Consequences of floral visits by ants and invasive honeybees to the hummingbird-pollinated, Caribbean cactus *Melocactus intortus*. Plant Species Biology, 26, 193-204.

Fagundes M. V., Souza A. F. and Ganade, G. (2022). Functional traits above and below ground allow species with distinct ecological strategies to coexist in the largest seasonally dry tropical forest in the Americas. Frontiers in Forests and Global Change, 5:930099

Fernandes M. F., Cardoso D., Pennington R. T. and de Queiroz, L. P. (2022). The origins and historical assembly of the Brazilian Caatinga seasonally dry tropical forests. Frontiers in Ecology and Evolution, 10,723286.

Fick S. E. and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International journal of climatology, 37, 4302-4315.

FitzJohn R. G. (2012). Diversitree: comparative phylogenetic analyses of diversification in R. Methods in Ecology and Evolution, 3, 1084-1092.

Folk R. A., Siniscalchi C. M. and Soltis, D. E. (2020). Angiosperms at the edge: Extremity, diversity, and phylogeny. Plant, Cell and Environment, 43, 2871-2893.

Fox E. W., Hill R. A., Leibowitz S. G., Olsen A. R., Thornbrugh D. J. and Weber M. H. (2017). Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology. Environmental Monitoring and Assessment, 189, 1-20.

Franck A. R., Cochrane B. J. and Garey, J. R. (2013). Phylogeny, biogeography, and infrageneric classification of *Harrisia* (Cactaceae). Systematic Botany, 38, 210-223.

Franco F. F., Silva G. A. R., Moraes E. M., Taylor N., Zappi D. C., Jojima C. L. and Machado, M. C. (2017). Plio-Pleistocene diversification of Cereus (Cactaceae, Cereeae) and closely allied genera. Botanical Journal of the Linnean Society, 183, 199-210.

Freilij D., Larrea-Alcazar D., Lopez R. P., Velarde Simonini, F., Naoki K. and Bessega C. Environmental gradualism explains variation in pollination systems of columnar cacti: Phylogenetic and trait evolution analyses. Global Ecology and Biogeography.

Gagnon E., Ringelberg J. J., Bruneau A., Lewis G. P. and Hughes C. E. (2019). Global succulent biome phylogenetic conservatism across the pantropical *Caesalpinia* group (Leguminosae). New Phytologist, 222, 1994-2008.

Gentry A. H. (1982). Neotropical floristic diversity: phytogeographical connections between Central and South America, Pleistocene climatic fluctuations, or an accident of the Andean orogeny?. Annals of the Missouri Botanical garden, 69, 557-593.

Givnish T. J., Barfuss M. H., Van Ee B., Riina R., Schulte K., Horres R., et al. (2014). Adaptive radiation, correlated and contingent evolution, and net species diversification in Bromeliaceae. Molecular Phylogenetics and Evolution, 71, 55-78.

Gomes V. G. N., Quirino Z. G. M. and Machado I. C. (2014). Pollination and seed dispersal of *Melocactus ernestii* Vaupel subsp. *ernestii* (Cactaceae) by lizards: an example of double mutualism. Plant Biology, 16, 315-322.

Griotti M., Ceccarelli F. S. and Roig-Juñent, S. (2023). Following the aridity: Historical biogeography and diversification of the Philodromidae spider genus *Petrichus* in South America. Molecular Phylogenetics and Evolution, 180, 107684.

Guillory W. X. and Brown, J. L. (2021). A new method for integrating ecological niche modeling with phylogenetics to estimate ancestral distributions. Systematic Biology, 70, 1033-1045.

Gutiérrez-Ortega J. S., Salinas-Rodríguez M. M., Martínez J. F., Molina-Freaner F., Pérez-Farrera M. A., Vovides A. P. et al. (2018). The phylogeography of the cycad genus *Dioon* (Zamiaceae) clarifies its Cenozoic expansion and diversification in the Mexican transition zone. Annals of Botany, 121, 535-548.

Hall MA e Smith LA. 1997. Feature subset selection: a correlation based filter approach. In International Conference on Neural Information Processing and Intelligent Information Systems (p. 855-858). Berlin: Springer

Herbert T. D., Lawrence K. T., Tzanova A., Peterson L. C., Caballero-Gill R. and Kelly C. S. (2016). Late Miocene global cooling and the rise of modern ecosystems. Nature Geoscience, 9, 843-847.

Hernández-Hernández T., Hernández H. M., De-Nova J. A., Puente R., Eguiarte L. E. and Magallón, S. (2011). Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). American journal of botany, 98, 44-61.

Hernández-Hernández 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 phytologist, 202(4), 1382-1397.

Hoang D. T., Chernomor O., Von Haeseler A., Minh B. Q. and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Molecular biology and evolution, 35, 518-522.

Hoorn C., Boschman L. M., Kukla T., Sciumbata M. and Val, P. (2022). The Miocene wetland of western Amazonia and its role in Neotropical biogeography. Botanical Journal of the Linnean Society, 199, 25-35.

Hunt D. R., Taylor N. P. and Charles, G. (2006). New cactus lexicon. Royal Botanic Gardens, Kew.

Hurbath F., Stubbs R. L., Cordeiro I. and Cellinese, N. (2021). Biogeography of succulent spurges from Brazilian Seasonally Dry Tropical Forest (SDTF). Taxon, 70, 153-169.

Jaramillo C., Sepulchre P., Cardenas D., Correa-Metrio A., Moreno J. E., Trejos R. et al. (2020). Drastic vegetation change in the Guajira Peninsula (Colombia) during the Neogene. Paleoceanography and Paleoclimatology, 35, e2020PA003933.

Kaufmann M. (2016). The WEKA Workbench. Online Appendix for" Data Mining: Practical Machine Learning Tools and Techniques.

Korotkova N., Aquino D., Arias S., Eggli U., Franck A., Gómez-Hinostrosa C. et al. (2021). Cactaceae at Caryophyllales. org–a dynamic online species-level taxonomic backbone for the family. Willdenowia, 51, 251-270.

Kumar S., Stecher G., Li M., Knyaz C. and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution, 35, 1547.

Lagomarsino L. P., Condamine F. L., Antonelli A., Mulch A. and Davis, C. C. (2016). The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). New Phytologist, 210, 1430-1442.

Lanna F. M., Colli G. R., Burbrink F. T. and Carstens, B. C. (2022). Identifying traits that enable lizard adaptation to different habitats. Journal of Biogeography, 49, 104-116.

Larocca P. F., Mancio J. S., Padilha P., Mello-Silva R. and Alcantara S. (2022). Recent divergence in functional traits affects rates of speciation in the Neotropical Velloziaceae (Pandanales). Botanical Journal of the Linnean Society, 199, 144-172.

Lavor P., Calvente A., Versieux L. M. and Sanmartin, I. (2019). Bayesian spatio-temporal reconstruction reveals rapid diversification and Pleistocene range expansion in the widespread columnar cactus *Pilosocereus*. Journal of Biogeography, 46, 238-250.

Ledo R. M. D. and Colli G. R. (2017). The historical connections between the Amazon and the Atlantic Forest revisited. Journal of Biogeography, 44, 2551-2563.

Lörch M., Mutke J., Weigend M. and Luebert, F. (2021). Historical biogeography and climatic differentiation of the *Fulcaldea-Archidasyphyllum-Arnaldoa* clade of Barnadesioideae

(Asteraceae) suggest a Miocene, aridity-mediated Andean disjunction associated with climatic niche shifts. Global and Planetary Change, 201, 103495.

Lu M., Fradera-Soler M., Forest F., Barraclough T. G. and Grace, O. M. (2022). Evidence linking life-form to a major shift in diversification rate in *Crassula*. American Journal of Botany, 109, 272-290.

Luebert F. (2021). The two South American dry diagonals. Frontiers of Biogeography, 13.

Magalhaes I. L. F., Neves D. M., Santos F. R., Vidigal T. H. D. A., Brescovit A. D., Santos A. J. (2019). Phylogeny of Neotropical *Sicarius* sand spiders suggests frequent transitions from deserts to dry forests despite antique, broad-scale niche conservatism. Molecular Phylogenetics and Evolution, 140, 106569.

Masa-Iranzo I., Sanmartín I., Caruzo M. B. R. and Riina, R. (2021). Skipping the Dry Diagonal: spatio-temporal evolution of *Croton* section *Cleodora* (Euphorbiaceae) in the Neotropics. Botanical Journal of the Linnean Society, 197, 61-84.

Matzke N. 2013. BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. University of California, Berkeley, Berkeley, CA

Mello B., Tao Q., Tamura K. and Kumar, S. (2017). Fast and accurate estimates of divergence times from big data. Molecular Biology and Evolution, 34, 45-50.

Menezes M. O. T. D., Taylor N. P. and Loiola M. I. B. (2013). Flora do Ceará, Brasil: Cactaceae. Rodriguésia, 64, 757-774.

Menezes M. O., Taylor N. P., Zappi D. C. and Loiola M. I. B. (2015). Spines and ribs of *Pilosocereus arrabidae* (Lem.) Byles and GD Rowley and allies (Cactaceae): Ecologic or genetic traits?. Flora-Morphology, Distribution, Functional Ecology of Plants, 214, 44-49.

Minh B. Q., Schmidt H. A., Chernomor O., Schrempf D., Woodhams M. D., Von Haeseler A., and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular biology and evolution, 37, 1530-1534.

Mogni V. Y., Oakley L. J. and Prado D. E. (2015). The distribution of woody legumes in neotropical dry forests: the Pleistocene Arc Theory 20 years on. Edinburgh Journal of Botany, 72, 35-60.

Morrone J. J. (2014). Cladistic biogeography of the Neotropical region: identifying the main events in the diversification of the terrestrial biota. Cladistics, 30, 202-214.

Neves D. M., Dexter K. G., Pennington R. T., Bueno M. L. and Oliveira Filho A. T. (2015). Environmental and historical controls of floristic composition across the South American Dry Diagonal. Journal of Biogeography, 42, 1566-1576.

Ogburn R. M. and Edwards, E. J. (2015). Life history lability underlies rapid climate niche evolution in the angiosperm clade Montiaceae. Molecular Phylogenetics and Evolution, 92, 181-192.

Palazzesi L., Hidalgo O., Barreda V. D., Forest F. and Höhna, S. (2022). The rise of grasslands is linked to atmospheric CO2 decline in the late Palaeogene. Nature Communications, 13, 293.

Pennington R. T., Lavin M., Prado D. E., Pendry C. A., Pell S. K. and Butterworth, C. A. (2004). Historical climate change and speciation: neotropical seasonally dry forest plants show patterns of both Tertiary and Quaternary diversification. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359, 515-538.

Pennington R. T., Lavin M., and Oliveira-Filho A. (2009). Woody plant diversity, evolution, and ecology in the tropics: perspectives from seasonally dry tropical forests. Annual Review of Ecology, Evolution, and Systematics, 40, 437-457.

Pie M. R., Faircloth B. C., Ribeiro L. F., Bornschein M. R. and Mccormack, J. E. (2018). Phylogenomics of montane frogs of the Brazilian Atlantic Forest is consistent with isolation in sky islands followed by climatic stability. Biological Journal of the Linnean Society, 125, 72-82.

Prado D. E. and Gibbs P. E. (1993). Patterns of species distributions in the dry seasonal forests of South America. Annals of the Missouri Botanical Garden, 902-927.

Prado D. E. (2000). Seasonally dry forests of tropical South America: from forgotten ecosystems to a new phytogeographic unit. Edinburgh Journal of Botany, 57, 437-461.

Prieto-Torres D. A., Rojas-Soto O. R., Santiago-Alarcon D., Bonaccorso E. and Navarro-SigüEnza, A. G. (2019). Diversity, endemism, species turnover and relationships among avifauna of neotropical seasonally dry forests. Ardeola, 66(2), 257-277.

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 in Ecology and Evolution, 5, 701-707.

Ringelberg J. J., Zimmermann N. E., Weeks A., Lavin M. and Hughes C. E. (2020). Biomes as evolutionary arenas: Convergence and conservatism in the trans-continental succulent biome. Global Ecology and Biogeography, 29, 1100-1113.

Ritz C. M., Martins L., Mecklenburg R., Goremykin V. and Hellwig, F. H. (2007). The molecular phylogeny of *Rebutia* (Cactaceae) and its allies demonstrates the influence of paleogeography on the evolution of South American mountain cacti. American Journal of Botany, 94, 1321-1332.

Romeiro-Brito M., Khan G., Perez M. F., Zappi D. C., Taylor N. P., Olsthoorn G. et al. (2023). Revisiting phylogeny, systematics, and biogeography of a Pleistocene radiation. American Journal of Botany, e16134.

Ruffley M., Peterson K., Week B., Tank D. C. and Harmon, L. J. (2019). Identifying models of trait-mediated community assembly using random forests and approximate Bayesian computation. Ecology and Evolution, 9, 13218-13230.

Queiroz L.P., Cardoso D., Fernandes M.F. and Moro M.F. 2017. Diversity and evolution of flowering plants of the Caatinga domain. Pp. 23–63 in: Silva, J.M.C., Leal, I.R. & Tabarelli, M. (eds.), Caatinga: The largest Tropical Dry Forest region in South America. Cham: Springer International Publishing.

Särkinen T., Pennington R. T., Lavin M., Simon M. F. and Hughes C. E. (2012). Evolutionary islands in the Andes: persistence and isolation explain high endemism in Andean dry tropical forests. Journal of Biogeography, 39, 884-900.

Sarmiento G. (1975). The dry plant formations of South America and their floristic connections. Journal of Biogeography, 233-251.

Sato H., Kelley D. I., Mayor S. J., Martin Calvo M., Cowling S. A. and Prentice, I. C. (2021). Dry corridors opened by fire and low CO2 in Amazonian rainforest during the Last Glacial Maximum. Nature Geoscience, 14, 578-585.

Sayyari E. and Mirarab S. (2016). Fast coalescent-based computation of local branch support from quartet frequencies. Molecular biology and evolution, 33, 1654-1668.

Schrider D. R. and Kern A. D. (2018). Supervised machine learning for population genetics: a new paradigm. Trends in Genetics, 34, 301-312.

Serrano-Serrano M. L., Rolland J., Clark J. L., Salamin N. and Perret, M. (2017). Hummingbird pollination and the diversification of angiosperms: an old and successful association in Gesneriaceae. Proceedings of the Royal Society B: Biological Sciences, 284, 20162816.

Sullivan J., Smith M. L., Espíndola A., Ruffley M., Rankin A., Tank D. and Carstens, B. (2019). Integrating life history traits into predictive phylogeography. Molecular Ecology, 28, 2062-2073.

Tamura K., Battistuzzi F. U., Billing-Ross P., Murillo O., Filipski A. and Kumar S. (2012). Estimating divergence times in large molecular phylogenies. Proceedings of the National Academy of Sciences, 109, 19333-19338.

Taylor N. P. and Zappi D. C. (2004). Cacti of eastern Brazil. Royal Botanic Gardens, Kew.

Taylor N. and Zappi D. (2018). Additions and corrections to ‘Cacti of Eastern Brazil’. Bradleya, 2018, 2-21.

Thode V. A., Sanmartín I., and Lohmann L. G. (2019). Contrasting patterns of diversification between Amazonian and Atlantic forest clades of Neotropical lianas (*Amphilophium*, Bignonieae) inferred from plastid genomic data. Molecular Phylogenetics and Evolution, 133, 92-106.

Title P. O. and Bemmels J. B. (2018). ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. Ecography, 41, 291-307.

Tripp E. A. and Tsai Y. H. E. (2017). Disentangling geographical, biotic, and abiotic drivers of plant diversity in neotropical Ruellia (Acanthaceae). PLoS One, 12, e0176021.

Vasconcelos T. N., Alcantara S., Andrino C. O., Forest F., Reginato M., Simon M. F. and Pirani J. R. (2020). Fast diversification through a mosaic of evolutionary histories characterizes the endemic flora of ancient Neotropical mountains. Proceedings of the Royal Society B, 287, 20192933.

Wang N., Yang Y., Moore M. J., Brockington S. F., Walker J. F., Brown J. W. et al. (2019). Evolution of Portulacineae marked by gene tree conflict and gene family expansion associated with adaptation to harsh environments. Molecular Biology and Evolution, 36, 112-126.

Werneck F. P. (2011). The diversification of eastern South American open vegetation biomes: historical biogeography and perspectives. Quaternary Science Reviews, 30(13-14), 1630-1648.

Werneck F. P., Costa G. C., Colli G. R., Prado D. E. and Sites Jr J. W. (2011). Revisiting the historical distribution of Seasonally Dry Tropical Forests: new insights based on palaeodistribution modelling and palynological evidence. Global Ecology and Biogeography, 20, 272-288.

Zappi D. and Taylor, N. (2008). Diversidade e endemismo das Cactaceae na Cadeia do Espinhaço. Megadiversidade, 4, 111-116.

Zizka A., Silvestro D., Andermann T., Azevedo J., Duarte Ritter C., Edler D. et al. (2019). CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. Methods in Ecology and Evolution, 10, 744-751.

3.8 Supplementary Material

Table S1. Taxon sampling of Cactaceae species used for divergence time analysis, including voucher information and Sequence Read Archive code. SRA code without numbers will be enabled upon publication of this manuscript

Species	GenBank Access	Voucher
Subfamily Pereskioideae		
Leuenbergeria aureiflora	SRR18315853	SORO 7951
Subfamily Opuntioideae		
Tacinga inamoena	SRR18315855	RBvc 205
Opuntia monacantha	SRR18315858	RBvc 183
Subfamily Cactoideae		
Tribe Cacteae		
Astrophytum myriostigma	SRR18315866	RBvc 3
Ferocactus glaucescens	SRR18315868	RBvc 207
Echinocactus grusonii	SRR18315870	RBvc 214
Tribe Phyllocacteae		
subtribe Hylocereinae		
Epiphyllum phyllanthus	SRR18315859	HRCB 5017
Selenicereus setaceus	SRR18315860	RBvc 277
subtribe Echinocereinae		
Echinocereus pentalophus	SRXXXXXXX	RBvc 252
Lophocereus marginatus	SRR18315864	RBvc 283
Cephalocereus polylophus	SRR18315863	RBvc 234
Tribe Rhipsalideae		
Hatiora salicornioides	SRR18315871	SORO 7952
Rhipsalis paradoxa	SRR18315873	RBvc 222
Tribe Notocacteae		
Parodia leninghausii	SRR18315874	SORO 7954
Parodia magnifica	SRR18315875	RBvc 37
Tribe Cereeae		
subtribe Rebutiinae		
Browningia hertlingiana	SRR18315876	KEW 32069
subtribe Gymnocalyciinae		

Species	**GenBank Access**	**Voucher**
Gymnocalycium anisitsii	SRRXXXXX	SORO 8154
Gymnocalycium denudatum	SRR18315877	KEW 32065
Gymnocalycium horstii	SRR18315879	RBvc 262
Gymnocalycium pflanzii	SRRXXXXX	RBvc 195
subtribe Uebelmanniinae		
Uebelmannia buiningii	SRRXXXXX	NA
Uebelmannia gummifera subsp. gummifera	SRR18315881	HURB 112
Uebelmannia pectinifera subsp. pectinifera	SRR18315882	SORO 4543
subtribe Cereinae		
Arrojadoa bahiensis	SRRXXXXX	SORO 8158
Arrojadoa dinae	SRRXXXXX	SORO 7981
Arrojadoa marylanae	SRRXXXXX	RBvc 101
Arrojadoa multiflora	SRRXXXXX	HURB 13627
Arrojadoa penicillata	SRRXXXXX	SORO 7998
Arrojadoa rhodantha	SRR18315883	SORO 4911
Arrojadoa olsthoorniana	SRRXXXXX	SORO 7986
Brasilicereus estevesii	SRRXXXXX	SORO 8014
Brasilicereus markgraffi	SRRXXXXX	HURB 10603
Brasilicereus phaeacanthus	SRR18315884	SORO 7955
Cereus albicaulis	SRRXXXXX	SORO 4559
Cereus bicolor	SRR18315916	NA
Cereus fernambucensis	SRR18315919	SORO 2672
Cereus hexagonus	SRR18315927	NA
Cereus hildmannianus	SRR18315929	SORO 2746
Cereus jamacaru	SRRXXXXX	SORO 8159
Cereus phatnospermus	SRR18315900	SORO 7969
Cereus mirabella	SRR18315911	SORO 7970
Cereus mortensenii	SRRXXXXX	GIB G575
Cereus repandus	SRR18315885	SORO 7956
Cereus saddianus	SRRXXXXX	SORO 3632
Cereus spegazzinii	SRRXXXXX	SORO 4541
Cereus stenogonus	SRR18315933	SORO 7971
Cereus trigonodendron	SRRXXXXX	GIB MB046.03

Species	GenBank Access	Voucher
Cereus vargasianus	SRRXXXXX	GIB PH1022.02
Cipocereus bradei	SRRXXXXX	SORO 4562
Cipocereus crassisepalus	SRRXXXXX	NA
Cipocereus laniflorus	SRRXXXXX	SORO 8016
Cipocereus pusilliflorus	SRRXXXXX	BHCB 21281
Coleocephalocereus aureus	SRRXXXXX	HURB 13694
Coleocephalocereus braunii	SRRXXXXX	UEC 202555
Coleocephalocereus fluminensis	SRRXXXXX	UEC 041393
Coleocephalocereus goebelianus	SRRXXXXX	HURB 13631
Coleocephalocereus purpureus	SRR18315888	SPF 80978
Discocactus bahiensis	SRRXXXXX	SORO 8004
Discocactus boliviensis	SRRXXXXX	SORO8027
Discocactus diersianus	SRRXXXXX	SORO5738
Discocactus heptacanthus var. cangaensis	SRRXXXXX	SORO7996
Discocactus placentiformis	SRRXXXXX	HRCB 13930
Discocactus zehntneri	SRRXXXXX	SORO5739
Facheiroa squamosa	SRRXXXXX	UB1239283
Facheiroa ulei	SRR18315892	HURB 11513
Leocereus bahiensis	SRRXXXXX	NA
Leocereus bahiensis	SRRXXXXX	HURB 7494
Melocactus bahiensis	SRRXXXXX	HUEFS 84848
Melocactus broadwayi	SRRXXXXX	Anterberger, 64
Melocactus concinnus	SRRXXXXX	NA
Melocactus conoideus	SRRXXXXX	NA
Melocactus deinacanthus	SRRXXXXX	SPF 80989
Melocactus ernestii	SRRXXXXX	SORO 5735
Melocactus ferreophilus	SRRXXXXX	HUEFS 168189
Melocactus levitestatus	SRRXXXXX	SORO7982
Melocactus macracanthos	SRR18315893	SORO7958
Melocactus mazelianus	SRRXXXXX	U.1180141
Melocactus oreas	SRRXXXXX	RBvc 135
Melocactus sergipensis	SRRXXXXX	ASE 33139

Species	GenBank Access	Voucher
Melocactus violaceus	SRRXXXXX	EAC 52277
Melocactus zehntneri	SRRXXXXX	HUEFS 226476
Micranthocereus albicephalus	SRRXXXXX	RBvc 163
Micranthocereus auriazureus	SRRXXXXX	HURB 13686
Micranthocereus dolichospermaticus	SRRXXXXX	UFMT 24569
Micranthocereus estevesii	SRRXXXXX	SORO 4906
Micranthocereus flaviflorus	SRRXXXXX	SORO 7974
Micranthocereus flaviflorus subsp. densiflorus	SRRXXXXX	HUEFS 73449
Micranthocereus polyanthus	SRR18315895	SORO 7959
Micranthocereus purpureus	SRR18315896	SORO 7973
Micranthocereus streckeri	SRR18315896	SORO 7960
Micranthocereus violaciflorus	SRRXXXXX	SORO 7972
Pilosocereus albisummus	SRRXXXXX	SORO 4530
Pilosocereus arrabidae	SRRXXXXX	SORO 2656
Pilosocereus aureispinus	SRR18315897	HUFS 642
Pilosocereus aurisetus	SRRXXXXX	SORO 4929
Pilosocereus azulensis	SRRXXXXX	SORO 4531
Pilosocereus bohlei	SRRXXXXX	SORO 3000
Pilosocereus brasiliensis	SRRXXXXX	SORO 4912
Pilosocereus flexibilispinus	SRRXXXXX	SORO 4935
Pilosocereus floccosus	SRRXXXXX	SORO 4558
Pilosocereus glachrocus	SRRXXXXX	SORO 4536
Pilosocereus jauruensis	SRRXXXXX	SORO 2646
Pilosocereus lanuginosus	SRR18315899	SORO 4925
Pilosocereus leucocephalus	SRRXXXXX	ZSS B 32820
Pilosocereus mollispinus	SRRXXXXX	SORO 8015
Pilosocereus multicostatus	SRRXXXXX	SORO 2649
Pilosocereus pachycladus	SRRXXXXX	SORO 4913
Pilosocereus polygonus	SRRXXXXX	ZSS B 32823
Pilosocereus vilaboensis	SRRXXXXX	SORO 3001
Pilosocereus piauhyensis	SRRXXXXX	NA
Praecereus euchlorus	SRR18315901	SORO 5704
Praecereus saxicola	SRRXXXXX	SORO 5703

Species	**GenBank Access**	**Voucher**
Stephanocereus leucostele	SRRXXXXX	SORO 7990
Stephanocereus luetzelburgii	SRRXXXXX	UB 1324887
Stetsonia coryne	SRR18315880	CGMS 17594
Xiquexique frewenii	SRR18315903	SORO 4937
Xiquexique gounellei	SRR18315904	SORO 4934
Xiquexique tuberculatus	SRRXXXXX	SORO 4930
Subtribe Trichocereinae		
Acanthocalycium rhodotrichum	SRRXXXXX	SORO 8021
Arthrocereus glaziovii	SRRXXXXX	RB 757784
Arthrocereus rondonianus	SRRXXXXX	RBvc 281
Arthrocereus spinosissimus	SRRXXXXX	RB 755039
Cleistocactus winteri	SRR18315906	RBvc 282
Echinopsis oxygona	SRR18315908	RBvc 117
Echinopsis calochlora	SRRXXXXX	SORO 8026
Espostoopsis dybowskii	SRR18315909	SORO 7963
Haageocereus decumbens	SRR18315910	KEW 32064
Harrisia adscendens	SRR18315912	SORO 4910
Harrisia bonplandii	SRRXXXXX	SORO 8013
Harrisia martinii	SRRXXXXX	KEW 32078
Leucostele atacamensis	SRRXXXXX	SORO 6565
Lobivia ancistrophora	SRRXXXXX	RBvc 264
Matucana intertexta	SRR18315913	KEW 32063
Oreocereus celsianus	SRRXXXXX	KEW 32067
Samaipaticereus corroanus	SRR18315914	KEW 32076
Soenrensia formosa	SRRXXXXX	KEW 32077

Table S2. Sampling fraction of backbone phylogeny (first line) and clade or genera for each tip. Name of the clades in which each species are allocated: Ueb - *Uebelmannia* clade; Brown - *Browningia* clade; Arthro - *Arthrocereus* clade; Harri - *Harrisia* clade; Cleis - *Cleistocactus* clade; Oreo - *Oreocereus* clade; Echi - *Echinopsis* clade; Gymn - *Gymnocalycium* clade; Stet - *Stesonia* clade; Prae - *Praecereus* clade; Fach - *Facheiroa* clade; Espo - *Espostoopsis* clade; Micr - *Micranthocereus* clade; Xiqe - *Xiquexique* clade; Albi - *Micranthocereus albicephalus* clade; Aust - *Micranthocereus auriazureus* clade; Coleo - *Coleocephalocereus* clade; Disco - *Discocactus* clade; Melo - *Melocactus* clade; Cere - *Cereus* clade; Cipo - *Cipocereus* clade; Arro - *Arrojadoa* clade; Pilo - *Pilosocereus* clade.

Tree tips	**Clade**	**Sampling proportion Cereeae**	**Sampling proportion Cereinae**
Backbone		0.27	1
Uebelmannia pectinifera	Ueb	1	
Uebelmannia buiningii	Ueb	1	
Uebelmannia gummifera	Ueb	1	
Browningia hertlingiana	Brown	0.09	
Arthrocereus spinosissimus	Arthro	0.75	
Arthrocereus glaziovii	Arthro	0.75	
Arthrocereus rondonianus	Arthro	0.75	
Harrisia adscendens	Harri	0.16	
Harrisia bonplandii	Harri	0.16	
Harrisia martinii	Harri	0.16	
Samaipaticereus corroanus	Cleis	0.06	
Cleistocactus winteri	Cleis	0.06	
Oreocereus celsianus	Oreo	0.04	
Haageocereus decumbens	Oreo	0.04	
Matucana intertexta	Oreo	0.04	
Echinospsis ancistrophora	Echi	0.06	
Acanthocalycium rhodotrichum	Echi	0.06	
Echinopsis calochlora	Echi	0.06	
Echinopsis oxygona	Echi	0.06	
Echinopsis atacamensis	Echi	0.06	
Echinopsis formosa	Echi	0.06	
Gymnocalycium anisitsii	Gymn	0.06	
Gymnocalycium pflanzii	Gymn	0.06	
Gymnocalycium denudatum	Gymn	0.06	
Gymnocalycium horstii	Gymn	0.06	
Stetsonia coryne	Stet	1	1

Tree tips	Clade	Sampling proportion Cereeae	Sampling proportion Cereinae
Praecereus euchlorus	Prae	1	1
Praecereus saxicola	Prae	1	1
Facheiroa markgrafii	Fach	0.86	0.86
Facheiroa phaeacantha	Fach	0.86	0.86
Facheiroa bragaia	Fach	0.86	0.86
Facheiroa bahiensis	Fach	0.86	0.86
Facheiroa ulei	Fach	0.86	0.86
Facheiroa squamosa	Fach	0.86	0.86
Espostoopsis dybowskii	Espo	1	1
Pilosocereus bohlei	Micr	0.83	0.83
Xiquexique tuberculatus	Xiqe	1	1
Xiquexique gounellei	Xiqe	1	1
Xiquexique frewenii	Xiqe	1	1
Micranthocereus purpureus	Micr	0.83	0.83
Micranthocereus polyanthus	Micr	0.83	0.83
Micranthocereus streckeri	Micr	0.83	0.83
Micranthocereus densiflorus	Micr	0.83	0.83
Micranthocereus flaviflorus	Micr	0.83	0.83
Coleocephalocereus dolichospermaticus	Coleo	0.58	0.58
Coleocephalocereus neoestevesii	Coleo	0.58	0.58
Coleocephalocereus goebelianus	Coleo	0.58	0.58
Coleocephalocereus fluminensis	Coleo	0.58	0.58
Coleocephalocereus braunii	Coleo	0.58	0.58
Coleocephalocereus purpureus	Coleo	0.58	0.58
Coleocephalocereus aureus	Coleo	0.58	0.58
Discocactus zehntneri	Disco	0.43	0.43
Discocactus placentiformis	Disco	0.43	0.43
Discocactus bahiensis	Disco	0.43	0.43
Discocactus diersianus	Disco	0.43	0.43
Discocactus boliviensis	Disco	0.43	0.43
Discocactus heptacanthus	Disco	0.43	0.43
Melocactus sergipensis	Melo	0.29	0.29

Tree tips	Clade	Sampling proportion Cereeae	Sampling proportion Cereinae
Melocactus levitestatus	Melo	0.29	0.29
Melocactus ferreophilus	Melo	0.29	0.29
Melocactus violaceus	Melo	0.29	0.29
Melocactus concinnus	Melo	0.29	0.29
Melocactus zehntneri	Melo	0.29	0.29
Melocactus macracanthos	Melo	0.29	0.29
Melocactus broadwayi	Melo	0.29	0.29
Melocactus mazelianus	Melo	0.29	0.29
Melocactus deinacanthus	Melo	0.29	0.29
Melocactus inconcinnus	Melo	0.29	0.29
Melocactus ernestii	Melo	0.29	0.29
Melocactus bahiensis	Melo	0.29	0.29
Melocactus conoideus	Melo	0.29	0.29
Melocactus oreas	Melo	0.29	0.29
Micranthocereus albicephalus	Albi	1	1
Micranthocereus auriazureus	Aust	1	1
Arrojadoa pusilliflora	Arro	0.85	0.85
Arrojadoa violaciflora	Arro	0.85	0.85
Arrojadoa luetzelburgii	Arro	0.85	0.85
Arrojadoa bahiensis	Arro	0.85	0.85
Arrojadoa leucostele	Arro	0.85	0.85
Arrojadoa marylanae	Arro	0.85	0.85
Arrojadoa multiflora	Arro	0.85	0.85
Arrojadoa rhodantha	Arro	0.85	0.85
Arrojadoa penicillata	Arro	0.85	0.85
Arrojadoa olsthoorniana	Arro	0.85	0.85
Arrojadoa dinae	Arro	0.85	0.85
Pilosocereus aureispinus	Pilo	0.33	0.33
Pilosocereus lanuginosus	Pilo	0.33	0.33
Pilosocereus leucocephalus	Pilo	0.33	0.33
Pilosocereus polygonus	Pilo	0.33	0.33
Pilosocereus glaucochrous	Pilo	0.33	0.33
Pilosocereus piauhyensis	Pilo	0.33	0.33

Tree tips	Clade	Sampling proportion Cereeae	Sampling proportion Cereinae
Pilosocereus brasiliensis	Pilo	0.33	0.33
Pilosocereus arrabidae	Pilo	0.33	0.33
Pilosocereus aurisetus	Pilo	0.33	0.33
Pilosocereus mollispinus	Pilo	0.33	0.33
Pilosocereus vilaboensis	Pilo	0.33	0.33
Pilosocereus jauruensis	Pilo	0.33	0.33
Pilosocereus azulensis	Pilo	0.33	0.33
Pilosocereus multicostatus	Pilo	0.33	0.33
Pilosocereus floccosus	Pilo	0.33	0.33
Pilosocereus albisummus	Pilo	0.33	0.33
Pilosocereus pachycladus	Pilo	0.33	0.33
Pilosocereus flexibilispinus	Pilo	0.33	0.33
Cipocereus crassisepalus	Cipo	0.6	0.6
Cipocereus bradei	Cipo	0.6	0.6
Cipocereus laniflorus	Cipo	0.6	0.6
Cereus repandus	Cere	0.41	0.41
Cereus mortensenii	Cere	0.41	0.41
Cereus mirabella	Cere	0.41	0.41
Cereus albicaulis	Cere	0.41	0.41
Cereus saddianus	Cere	0.41	0.41
Cereus kroenleinii	Cere	0.41	0.41
Cereus spegazzinii	Cere	0.41	0.41
Cereus vargasianus	Cere	0.41	0.41
Cereus fernambucencis	Cere	0.41	0.41
Cereus trigonodendron	Cere	0.41	0.41
Cereus hildmannianus	Cere	0.41	0.41
Cereus stenogonus	Cere	0.41	0.41
Cereus hexagonus	Cere	0.41	0.41
Cereus jamacaru	Cere	0.41	0.41

Table S3. Trait codes used for trait-dependent diversification analysis (BiSSE model). Growth form code: 0 = globose; 1 = shrubby or arborescent; Cephalia, pollinator hummingbird, pollinator bat and pollinator moth code: 0 = absence; 1 = presence.

Taxon	Growth form	Cephalia	Pollinator hummingbird	Pollinator bat	Pollinator moth
Arrojadoa bahiensis	1	0	1	0	0
Arrojadoa dinae	1	1	1	0	0
Arrojadoa marylanae	1	1	1	0	0
Arrojadoa multiflora	1	1	1	0	0
Arrojadoa penicillata	1	1	1	0	0
Arrojadoa rhodantha	1	1	1	0	0
Arrojadoa olsthoorniana	1	1	1	0	0
Facheioroa bragaia	1	0	0	1	0
Facheioroa markgrafii	1	0	0	1	0
Facheioroa phaeacantha	1	0	0	1	0
Cereus albicaulis	1	0	0	0	1
Cereus fernambucencis	1	0	0	0	1
Cereus hexagonus	1	0	0	0	1
Cereus hildmannianus	1	0	0	0	1
Cereus jamacaru	1	0	0	0	1
Cereus phatnospermus	1	0	0	0	1
Cereus mirabella	1	0	0	0	1
Cereus mortensenii	1	1	0	0	1
Cereus repandus	1	0	0	0	1
Cereus saddianus	1	0	0	0	1
Cereus spegazzinii	1	0	0	0	1
Cereus stenogonus	1	0	0	0	1
Cereus trigonodendron	1	0	0	0	1
Cereus vargasianus	1	0	0	0	1
Cipocereus bradei	1	0	0	1	0
Cipocereus crassisepalus	1	0	0	1	0
Cipocereus laniflorus	1	0	0	1	0
Arrojadoa pusilliflora	1	0	0	1	0
Coleocephalocereus aureus	0	1	1	0	0

Taxon	Growth form	Cephalia	Pollinator hummingbird	Pollinator bat	Pollinator moth
Coleocephalocereus braunii	1	1	1	0	0
Coleocephalocereus fluminensis	1	1	0	1	0
Coleocephalocereus goebelianus	1	1	0	1	0
Coleocephalocereus purpureus	1	1	1	0	0
Discocactus bahiensis	0	1	0	0	1
Discocactus boliviensis	0	1	0	0	1
Discocactus diersianus	0	1	0	0	1
Discocactus heptacanthus	0	1	0	0	1
Discocactus placentiformis	0	1	0	0	1
Discocactus zehntneri	0	1	0	0	1
Espostoopsis dybowskii	1	1	0	1	0
Facheiroa squamosa	1	0	0	1	0
Facheiroa ulei	1	1	0	1	0
Facheiroa bahiensis	1	0	0	1	0
Melocactus bahiensis	0	1	1	0	0
Melocactus broadwayi	0	1	1	0	0
Melocactus concinnus	0	1	1	0	0
Melocactus conoideus	0	1	1	0	0
Melocactus deinacanthus	0	1	1	0	0
Melocactus ernestii	0	1	1	0	0
Melocactus ferreophilus	0	1	1	0	0
Melocactus inconcinnus	0	1	1	0	0
Melocactus levitestatus	0	1	1	0	0
Melocactus macracanthos	0	1	1	0	0
Melocactus mazelianus	0	1	1	0	0
Melocactus oreas	0	1	1	0	0
Melocactus sergipensis	0	1	1	0	0
Melocactus violaceus	0	1	1	0	0
Melocactus zehntneri	0	1	1	0	0
Micranthocereus albicephalus	1	1	0	1	0
Micranthocereus auriazureus	1	0	1	0	0

Taxon	Growth form	Cephalia	Pollinator hummingbird	Pollinator bat	Pollinator moth
Coleocephalocereus dolichospermaticus	1	1	0	1	0
Coleocephalocereus estevesii	1	1	0	1	0
Micranthocereus densiflorus	1	0	1	0	0
Micranthocereus flaviflorus	1	0	1	0	0
Micranthocereus polyanthus	1	1	1	0	0
Micranthocereus purpureus	1	1	0	1	0
Micranthocereus streckeri	1	1	1	0	0
Arrojadoa violaciflora	1	1	1	0	0
Pilosocereus albisummus	1	0	0	1	0
Pilosocereus arrabidae	1	0	0	1	0
Pilosocereus aureispinus	1	0	0	1	0
Pilosocereus aurisetus	1	0	0	1	0
Pilosocereus azulensis	1	0	0	1	0
Pilosocereus bohlei	1	0	0	1	0
Pilosocereus brasiliensis	1	0	0	1	0
Pilosocereus flexibilispinus	1	0	0	1	0
Pilosocereus floccosus	1	0	0	1	0
Xiquexique frewenii	1	0	0	1	0
Pilosocereus glaucochrous	1	0	0	1	0
Xiquexique gounellei	1	0	0	1	0
Pilosocereus jauruensis	1	0	0	1	0
Pilosocereus lanuginosus	1	0	0	1	0
Pilosocereus leucocephalus	1	0	0	1	0
Pilosocereus mollispinus	1	0	0	1	0
Pilosocereus multicostatus	1	0	0	1	0
Pilosocereus pachycladus	1	0	0	1	0
Pilosocereus piauhyensis	1	0	0	1	0
Pilosocereus polygonus	1	0	0	1	0
Xiquexique tuberculatus	1	0	0	1	0
Pilosocereus vilaboensis	1	0	0	1	0
Praecereus euchlorus	1	0	0	0	1
Praecereus saxicola	1	0	0	0	1

Taxon	Growth form	Cephalia	Pollinator hummingbird	Pollinator bat	Pollinator moth
Arrojadoa leucostele	1	1	0	1	0
Arrojadoa luetzelburgii	1	1	0	1	0
Stetsonia coryne	1	0	0	0	1

Table S4. Comparison among models estimated by BiSSE model and respective statistics estimated for characteristics with significant AIC and LRT values (i.e. when compared to the null model, with no variation in speciation, extinction, and transition rate). The parameters correspond to λ = speciation rate (events/Ma/lineage); μ = extinction rate (events/Ma/lineage); q = transition rate between states; NP, number of parameters in each model.

Models	Np	LnL	AIC	p	λ0	λ1	μ0	μ1	q01	q10
Growth form										
(λ0=λ1; μ0=μ1; q01=q10)	3	-197.0	400.0	0.00	0.508	0.508	0.000	0.000	0.009	0.009
(λ0≠λ1; μ0=μ1; q01=q10)	4	-190.2	390.4	0.67	0.878	0.447	0.000	0.000	0.008	0.008
(λ0=λ1; μ0≠μ1; q01=q10)	4	-197.0	402.0	0.00	0.508	0.508	0.000	0.000	0.008	0.008
(λ0=λ1; μ0=μ1; q01≠q10)	4	-196.5	401.0	0.00	0.508	0.508	0.000	0.000	0.000	0.011
(λ0≠λ1; μ0≠μ1; q01=q10)	5	-191.2	392.6	0.37	0.878	0.447	0.000	0.000	0.007	0.007
(λ0≠λ1; μ0=μ1; q01≠q10)	5	-190.8	391.6	1.00	0.876	0.448	0.000	0.000	0.000	0.009
(λ0=λ1; μ0≠μ1; q01≠q10)	5	-196.5	403.0	0.00	0.508	0.508	0.000	0.000	0.000	0.011
(λ0≠λ1; μ0≠μ1; q01≠q10)	6	-190.8	393.6	0.876	0.448	0.000	0.000	0.000	0.009	0.009
Pollinator - hummingbird										
(λ0=λ1; μ0=μ1; q01=q10)	3	-222.1	450.2	0.04	0.514	0.514	0.000	0.000	0.042	0.042
(λ0≠λ1; μ0=μ1; q01=q10)	4	-217.1	444.2	0.92	0.575	0.775	0.000	0.000	0.042	0.042
(λ0=λ1; μ0≠μ1; q01=q10)	4	-221.1	452.2	0.01	0.574	0.574	0.000	0.000	0.042	0.042
(λ0=λ1; μ0=μ1; q01≠q10)	4	-222.1	444.2	0.92	0.516	0.516	0.000	0.000	0.000	0.093
(λ0≠λ1; μ0≠μ1; q01=q10)	5	-221.1	452.2	0.01	0.574	0.452	0.000	0.000	0.048	0.048
(λ0≠λ1; μ0=μ1; q01≠q10)	5	-217.6	445.2	1	0.564	0.477	0.000	0.000	0.000	0.093
(λ0=λ1; μ0≠μ1; q01≠q10)	5	-218.1	446.2	0.69	0.516	0.516	0.000	0.000	0.000	0.093
(λ0≠λ1; μ0≠μ1; q01≠q10)	6	-218.0	448.0	0.600	0.372	0.000	0.000	0.000	0.090	0.090

Table S5. Description of biotic traits obtained for subtribe Cereinae taxa to build the predictive model.

Traits	Type	Description (classes or metrics)
Vegetative traits		
Growth form	Binary	Growth form (globose, columnar).
Plant height	Numeric	Minimum and maximum plant height (cm).
Cladode diameter	Numeric	Minimum and maximum cladode diameter (cm).
Nº of ribs	Numeric	Minimum and maximum number of ribs
Cephalium	Binary	cephalium (presence, absence).
Spine traits		
Nº of central spines	Numeric	Minimum and maximum number of central spines
Central spines length	Numeric	Minimum and maximum spines length (mm).
Nº of radial spines	Numeric	Minimum and maximum number of radial spines
Radial spines length	Numeric	Minimum and maximum radial spines length (mm).
Flower traits		
Flowering region	Category	Flowering region (apical, subapical, lateral,undifferentiated).
Flower length	Numeric	Minimum and maximum flower length (mm).
Flower diameter	Numeric	Minimum and maximum Flower diameter (mm).
Flower form	Binary	Flower form (funnelform, tubular, salverform, campanulate).
Anthesis	Binary	Anthesis (diurnal, crepuscular, nocturnal).
Floral hypanthium	Category	floral hypanthium (naked or with trichomes, hairy, bristly, with scales and areoles).
Fruit traits		
Fruit length	Numeric	Minimum and maximum fruit length (mm).
Fruit diameter	Numeric	Minimum and maximum fruit diameter (mm).
Fruit form	Category	Fruit form (ovoid, globose, piriform, oblong, tuberculate, clavate)
Dehiscence	Binary	Dehiscence (presence, absence).
Pulp	Binary	Pulp (succulent, dry).
Seed traits		
Seed length	Numeric	Minimum and maximum seed length (mm).
Seed diameter	Numeric	Minimum and maximum seed diameter (mm).
Seed form	Binary	Seed form (ovoid, globose, oblong, cochleariform, hat-shaped, pyriform)

Table S6. Estimates for each biogeographic model in BioGeoBEARS. Lnl: maximum likelihood, Np: number of parameters, d: dispersal parameter; e: extinction parameter; j: founder event parameter; pval: significance value; AIC: Akaike information criteria value; and AICwt: weighted-AIC value. The best model is highlighted in red.

models	LnL	Np	d	e	j	AIC	AICwt
DEC	-319.9	2	0.017	0.009	0	593.4	0
DEC+J	-293.7	3	0.011	0	0.016	643.8	1.00
DIVALIKE	-314.1	2	0.018	0	0	591.5	0
DIVALIKE+J	-292.7	3	0.014	0	0.013	632.2	1.00
BAYAREALIKE	-331.4	2	0.014	0.159	0	553.7	0
BAYAREALIKE+J	-303.8	3	0.007	0.004	0.02	666.9	0

Figure S1. Calibrated tree of tribe Cereeae with all outgroups using the RelTime method in MEGAX. Node bars correspond to 95% confidence intervals estimates.

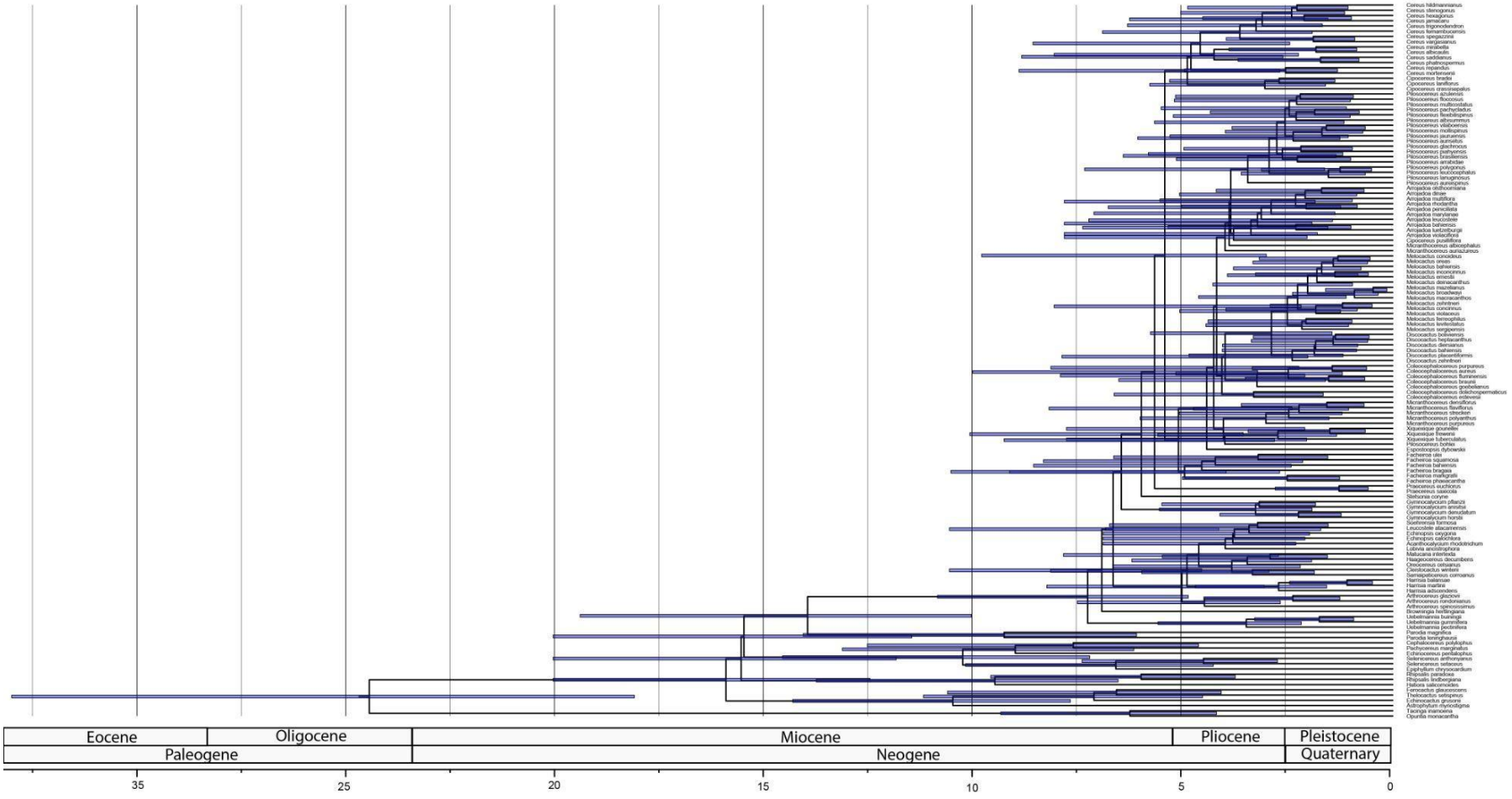


Figure S2. Set of shift in diversification rate of the tribe Cereeae estimated in BAMM, including the 95% credibility shift estimates. Colder colors in branches show slower rates, while warmer colors show faster rates. The best shift configuration is represented by the highest frequency estimates (f=0.72). Circles indicate a significant increase in diversification rates in each shift configuration and circle size corresponds to shift probability.

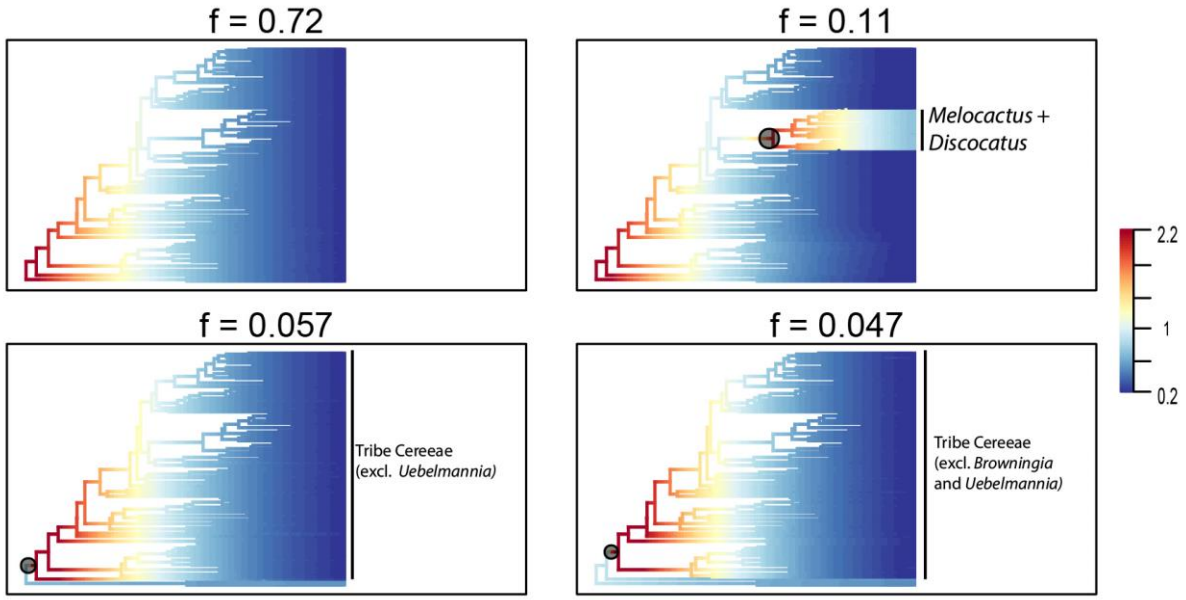
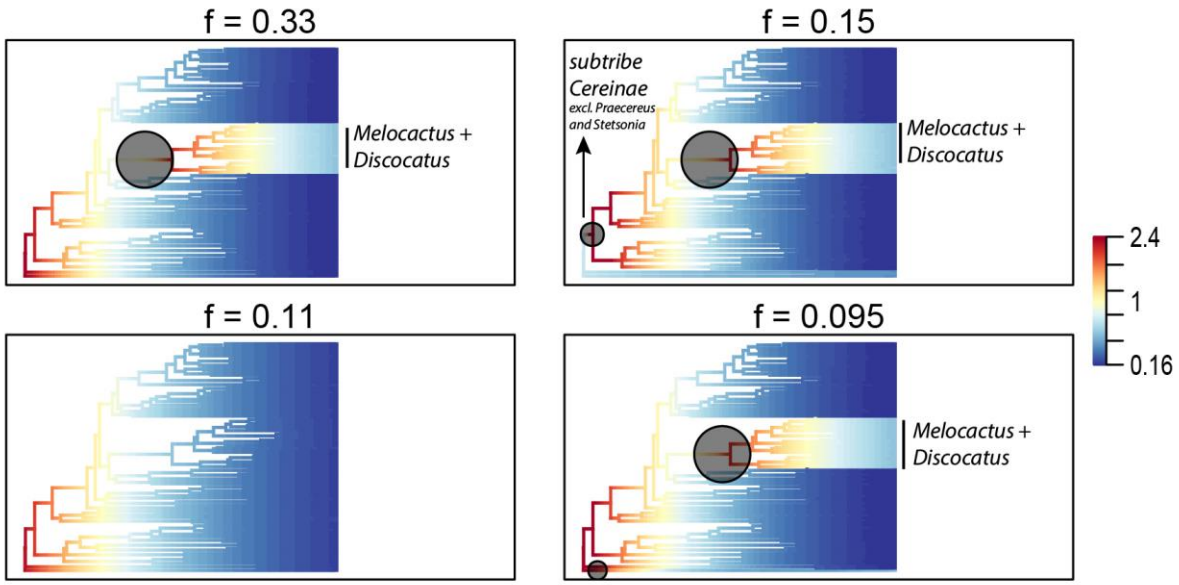


Figure S3. Set of shifts of diversification rate of subtribe Cereinae estimated in BAMM, including the 95% credibility shift estimates. Colder colors in branches show slower rates, while warmer colors show faster rates. The best shift configuration is represented by the higher frequency estimates (f=0.33). Circles indicate a significant increase in diversification rates in each shift configuration and circle size corresponds to shift probability.



Concluding remarks and future perspectives

The present thesis comprises three chapters that delineated the steps through the investigation of the ecological and evolutionary patterns involved in the diversification of arid habitats, using a South American cactus clade as a biological model. The first chapter selected nuclear regions, including coding and noncoding regions, useful to infer deep and shallow phylogenetic relationships within Cactaceae. Overall, the nuclear region selected to construct the lineage-specific probe set for Cactaceae (named Cactaceae591) has the potential to unravel contentious phylogenetic relationships in Cactaceae, especially those from the subfamily Cactoideae. Moreover, we highlighted the use of genomic datasets allied to coalescent-based phylogenetic inferences, given the extensive incomplete lineage sorting presented in this family.

In the second chapter, we used the Cactaceae591 panel and publicly available dataset to explore phylogenetic relationships within the tribe Cereeae, a South American cactus group. The use of genome-scale datasets and coalescent-based analysis were fundamental to untangling the contentious relationships within tribe Cereeae. This chapter showed for the first time a well-supported phylogenetic tree along all relationships of tribe Cereeae. As previously mentioned by Guerrero et al. (2018), this group has shown several non-monophyletic genera, which poses an urgent need for a taxonomic arrangement within the tribe Cereeae. We proposed a new taxonomic classification for the tribe Cereeae and expanded the classification of some genera in the subtribe Cereinae. This work highlights the use of genomic datasets as a mandatory step to delineate evolutionary study in Cactaceae, mainly within the tribe Cereeae. Moreover, the new taxonomic classification is fundamental to guide further phylogenomic studies as well as ancestral trait and range reconstructions in tribe Cereeae. Lastly, further studies with a comprehensive sampling may focus on early divergent lineages of tribe Cereeae and generic relationships within the subtribe Trichocereinae.

The last chapter investigated divergence time, ancestral range distribution, and biotic and abiotic factors involved in the diversification of taxa inhabiting arid and dry regions in South America, using the tribe Cereeae as a biological model. The major clades of Cereeae diversified during the late Miocene, a period of global cooling and expansion of arid environments around the globe. The most recent common ancestor of the tribe Cereeae was distributed in a widespread area reaching Central and Southern Andes, including Chacoan region, through Espinhaço Range. These biogeographical regions were segregated, probably due to the establishment of Cerrado domain and marine transgression. The colonization of the

Caatinga domain occurred through long-distance dispersal events, followed by in-situ diversification and colonization of dry habitats in adjacent biomes. Moreover, the biotic traits had shown a greater influence on diversification rates than abiotic traits. The most important biotic traits associated with the increase in diversification rates were globose growth form and hummingbird pollination. Here we have also used machine learning methods to screen for biotic and abiotic traits related to diversification rates within the tribe Cereeae. By employing machine learning techniques, this study aimed to identify biotic and abiotic traits associated with diversification rates within the tribe Cereeae, which may help uncover new traits for further investigation in diversification analysis. Finally, the inclusion of representatives of subtribe Aylosterinae and Rebutinae in further biogeographical analysis may help test the scenario of a widespread dry area during the late Miocene and consecutive segregation after the Cerrado establishment by the end of Miocene.

APPENDIX - Published papers during PhD

Perez, M. F., Bonatelli, I. A., **Romeiro-Brito, M.**, Franco, F. F., Taylor, N. P., Zappi, D. C., & Moraes, E. M. (2022). Coalescent-based species delimitation meets deep learning: Insights from a highly fragmented cactus system. Molecular Ecology Resources, 22(3), 1016-1028.

Received: 3 January 2021 | Revised: 16 September 2021 | Accepted: 12 October 2021

DOI: 10.1111/1755-0998.13534

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

Coalescent-based species delimitation meets deep learning: Insights from a highly fragmented cactus system

Manolo F. Perez[1,2] | Isabel A. S. Bonatelli[1,3] | Monique Romeiro-Brito[1] | Fernando F. Franco[1] | Nigel P. Taylor[4] | Daniela C. Zappi[5] | Evandro M. Moraes[1]

[1]Departamento de Biologia, Universidade Federal de São Carlos, Sorocaba, Brazil

[2]Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, Brazil

[3]Departamento de Ecologia e Biologia Evolutiva, Universidade Federal de São Paulo, Diadema, Brazil

[4]University of Gibraltar, The Alameda, Gibraltar

[5]Programa de Pós Graduação em Botânica, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, Brazil

Correspondence
Evandro M. Moraes, Departamento de Biologia, Universidade Federal de São Carlos, Sorocaba, SP, Brazil.
Email: emarsola@ufscar.br

Funding information
Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Number: 03940/2019-0 and 305301/2018-7; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: 001; Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Number: 2012/22857-8, 2012/22943-1 and 2015/06160-5

Abstract

Delimiting species boundaries is a major goal in evolutionary biology. An increasing volume of literature has focused on the challenges of investigating cryptic diversity within complex evolutionary scenarios of speciation, including gene flow and demographic fluctuations. New methods based on model selection, such as approximate Bayesian computation, approximate likelihoods, and machine learning are promising tools arising in this field. Here, we introduce a framework for species delimitation using the multispecies coalescent model coupled with a deep learning algorithm based on convolutional neural networks (CNNs). We compared this strategy with a similar ABC approach. We applied both methods to test species boundary hypotheses based on current and previous taxonomic delimitations as well as genetic data (sequences from 41 loci) in *Pilosocereus aurisetus*, a cactus species complex with a sky-island distribution and taxonomic uncertainty. To validate our method, we also applied the same strategy on data from widely accepted species from the genus *Drosophila*. The results show that our CNN approach has a high capacity to distinguish among the simulated species delimitation scenarios, with higher accuracy than ABC. For the cactus data set, a splitter hypothesis without gene flow showed the highest probability in both CNN and ABC approaches, a result agreeing with previous taxonomic classifications and in line with the sky-island distribution and low dispersal of *P. aurisetus*. Our results highlight the cryptic diversity within the *P. aurisetus* complex and show that CNNs are a promising approach for distinguishing complex evolutionary histories, even outperforming the accuracy of other model-based approaches such as ABC.

KEYWORDS
approximate Bayesian computation, convolutional neural networks, deep learning, fragmented systems, recent diversification, species delimitation

1 | INTRODUCTION

Recognizing species boundaries has long been a major challenge for biologists. The main difficulty is to some degree related to the numerous existing species concepts. The use of specific definitions can lead to alternative strategies for identifying species boundaries in empirical data sets (Carstens et al., 2013; de Queiroz, 2007). However, different species concepts can be considered elements of diverse properties that are associated with the dynamics of the speciation continuum. After the proposal of the unified species concept (Queiroz, 2007), the roles of species concept theory and species delimitation methodologies became apparent. The view of species as independent segments

Amaral, D. T., Minhós-Yano, I., Oliveira, J. V. M., **Romeiro-Brito, M.**, Bonatelli, I. A. S., Taylor, N. P., ... & Franco, F. F. (2021). Tracking the xeric biomes of South America: The spatiotemporal diversification of Mandacaru cactus. Journal of Biogeography, 48(12), 3085-3103.

Received: 11 February 2021 | Revised: 26 August 2021 | Accepted: 27 August 2021

DOI: 10.1111/jbi.14265

RESEARCH ARTICLE

Journal of Biogeography WILEY

Tracking the xeric biomes of South America: The spatiotemporal diversification of Mandacaru cactus

Danilo Trabuco Amaral[1,2] | Isis Minhós-Yano[1] | João Victor M. Oliveira[1] | Monique Romeiro-Brito[1] | Isabel Aparecida Silva Bonatelli[1,3] | Nigel Paul Taylor[4] | Daniela Cristina Zappi[5] | Evandro Marsola Moraes[1] | Deren A. R. Eaton[6] | Fernando Faria Franco[1]

[1]Departamento de Biologia, Centro de Ciências Humanas e Biológicas, Universidade Federal de São Carlos (UFSCar), Sorocaba, Brazil

[2]Programa de Pós Graduação em Biologia Comparada, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo (USP), Ribeirão Preto, Brazil

[3]Departamento de Ecologia e Biologia Evolutiva, Universidade Federal de São Paulo (UNIFESP), Diadema, Brazil

[4]University of Gibraltar, Gibraltar Botany Garden campus, Gibraltar, Gibraltar

[5]Programa de Pós Graduação em Botânica, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, Brazil

[6]Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, New York, USA

Correspondence
Fernando Faria Franco, Rodovia João Leme dos Santos, Km 110, SP 264, ZIP 18052-780, Sorocaba, Brazil.
Email: franco@ufscar.br

Funding information
Conselho Nacional de Desenvolvimento Científico e Tecnológico; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior; Fundacao de Amparo a Pesquisa do Estado de Sao Paulo

Handling Editor: Lars Chatrou

Abstract

Aim: The interconnectedness and biotic interchange among Neotropical biomes are thought to play an important role in driving adaptation and diversification. However, how these processes are in synteny to trait evolution in species of open and xeric areas is poorly studied. Here, we investigate the spatial and temporal dimensions of evolution and candidate traits associated with biome shifts in xeric vegetation, focusing on the family Cactaceae.

Location: Xeric and open areas of South America.

Taxon: Genus *Cereus* Mill. (Cactaceae, Cereeae).

Methods: We applied biogeographical reconstructions on a time-calibrated phylogeny inferred from multilocus data (ddRAD-Seq) using Bayesian analyses on BEAST2, species distribution modelling in Maxent, the reconstruction of biome affinities and niche shift analyses based on abiotic traits (climate and soil) using Mk-model in BioGeoBEARS, and phenotypic trait-based analysis in Mesquite.

Results: The Cerrado domain is the ancestral area of *Cereus*, with most diversification events occurring in a time of intense orogenesis, climatic changes, and marine regressions within the last 5 Mya. Events of biome transition from the seasonally dry tropical forest (SDTF) were also associated with trait and niche shifts.

Main conclusions: The diversification of the xerophyte genus *Cereus* is associated with the climatic and geomorphological instabilities of the Pliocene and Pleistocene epochs. The Cerrado domain states an important region of dispersal for the genus. Some geographical range movements involved biome shifts associated with niche evolution while others were restricted to a simple biogeographical transition without niche change. Particular clades that experienced biome shifts displayed some phenotypic state changes, suggesting a role of biotic traits for environment transition. The results observed in *Cereus* may be a biogeographical pattern that should be tested with other cactus species, such as *Pilosocereus* spp., or species of xeric habitats, such as Annonaceae and Vochysiaceae.

KEYWORDS
biogeography, biome transition, niche shifts, Plio-Pleistocene, RAD-Seq, SDTF, trait shift

Amaral, D. T., Bonatelli, I. A., **Romeiro-Brito, M.**, Moraes, E. M., & Franco, F. F. (2022). Spatial patterns of evolutionary diversity in Cactaceae show low ecological representation within protected areas. Biological Conservation, 273, 109677.

Biological Conservation 273 (2022) 109677

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Spatial patterns of evolutionary diversity in Cactaceae show low ecological representation within protected areas

Danilo Trabuco Amaral [a,b,d,*], Isabel A.S. Bonatelli [c], Monique Romeiro-Brito [a], Evandro Marsola Moraes [a], Fernando Faria Franco [a]

[a] Departamento de Biologia, Centro de Ciências Humanas e Biológicas, Universidade Federal de São Carlos (UFSCar), Sorocaba, Brazil
[b] Programa de Pós Graduação em Biologia Comparada, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo (USP), Ribeirão Preto, Brazil
[c] Departamento de Ecologia e Biologia Evolutiva, Universidade Federal de São Paulo (UNIFESP), Diadema, São Paulo, Brazil
[d] Universidade Federal do ABC (UFABC), Centro de Ciências Naturais e Humanas (CCNH), Santo André, São Paulo, Brazil

ARTICLE INFO

Keywords:
Conservation
Drylands
Phylogenetic endemism
Protected areas
Refugia
Spatial patterns

ABSTRACT

Mapping biodiversity patterns across taxa and environments is crucial to address the evolutionary and ecological dimensions of species distribution, suggesting areas of particular importance for conservation purposes. Within Cactaceae, spatial diversity patterns are poorly explored, as are the abiotic factors that may predict these patterns. We gathered geographic and genetic data from 921 cactus species by exploring both the occurrence and genetic databases, which are tightly associated with drylands, to evaluate diversity patterns, such as phylogenetic diversity and endemism, paleo-, neo-, and superendemism, and the environmental predictor variables of such patterns in a global analysis. Hotspot areas of cacti diversity are scattered along the Neotropical and Nearctic regions, mainly in the desertic portion of Mesoamerica, Caribbean Island, and the dry diagonal of South America. The geomorphological features of these regions may create a complexity of areas that work as locally buffered zones over time, which triggers local events of diversification and speciation. Desert and dryland/dry forest areas comprise paleo- and superendemism and may act as both museums and cradles of species, displaying great importance for conservation. Past climates, topography, soil features, and solar irradiance seem to be the main predictors of distinct endemism types. The hotspot areas that encompass a major part of the endemism cells are outside or poorly covered by formal protection units. The current legally protected areas are not able to conserve the evolutionary diversity of cacti. Given the rapid anthropogenic disturbance, efforts must be reinforced to monitor biodiversity and the environment and to define/plan current and new protected areas.

1. Introduction

A meaningful observation regarding biodiversity is that organisms have uneven distribution across the globe, which can reveal how speciation, extinction, and dispersal events may have impacted species distribution (Lomolino et al., 2009). Naturalists have long been interested in explaining why some regions are biologically richer than others as a way to minimize the Wallacean shortfall of biodiversity knowledge. Mapping biodiversity patterns across taxa and environmental conditions is crucial to address the evolutionary and ecological dimensions of species distribution, such as endemism patterns that emerge in regions with significant concentrations of organisms with little representation elsewhere. Areas such as these are particularly meaningful for conservation purposes (Willig et al., 2003; Graham and Fine, 2008; Swenson et al., 2012; Rosauer and Jetz, 2015).

Endemism has multiple spatial and temporal dimensions that can be related to different taxonomic levels, from families to subspecies (Morrone, 2008). Species-rich areas tend to have high endemism, which is usually correlated with contemporary and historical climate regimes and topography (Sandel et al., 2011; Daru et al., 2015; Barratt et al., 2017; Fenker et al., 2020). The metrics used to estimate endemism (see appendix A in Supplementary material) are influenced by the defined spatial scale, which can range from large (e.g., continent) to small areas (mountain tops). On a temporal scale, endemism can be described as a result of recent speciation with no dispersion out of the ancestral area (neoendemism) or as the persistence of lineages extinct elsewhere

* Corresponding author at: Rodovia João Leme dos Santos, Km 110, SP 264, 18052-780 Sorocaba, Brazil.
E-mail address: danilo.trabuco@ufabc.edu.br (D.T. Amaral).

https://doi.org/10.1016/j.biocon.2022.109677
Received 16 December 2021; Received in revised form 27 June 2022; Accepted 29 July 2022

Franco, F. F., Amaral, D. T., Bonatelli, I. A., **Romeiro-Brito, M.,** Telhe, M. C., & Moraes, E. M. (2022). Evolutionary genetics of cacti: research biases, advances and prospects. Genes, 13(3), 452.

genes

MDPI

Review

Evolutionary Genetics of Cacti: Research Biases, Advances and Prospects

Fernando Faria Franco [1,†], Danilo Trabuco Amaral [1,2,†], Isabel A. S. Bonatelli [3], Monique Romeiro-Brito [1], Milena Cardoso Telhe [1] and Evandro Marsola Moraes [1,*]

1 Departamento de Biologia, Centro de Ciências Humanas e Biológicas, Universidade Federal de São Carlos (UFSCar), Sorocaba 18052-780, Brazil; franco@ufscar.br (F.F.F.); danilo.trabuco@gmail.com (D.T.A.); romeiro.monique@gmail.com (M.R.-B.); milena.telhe@gmail.com (M.C.T.)
2 Programa de Pós-graduação em Biologia Comparada, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo (USP), Ribeirão Preto 14040-901, Brazil
3 Departamento de Ecologia e Biologia Evolutiva, Universidade Federal de São Paulo (UNIFESP), Diadema, São Paulo 04021-001, Brazil; belbonatelli@gmail.com
* Correspondence: emarsola@ufscar.br
† These authors contributed equally to this work.

Abstract: Here, we present a review of the studies of evolutionary genetics (phylogenetics, population genetics, and phylogeography) using genetic data as well as genome scale assemblies in Cactaceae (Caryophyllales, Angiosperms), a major lineage of succulent plants with astonishing diversity on the American continent. To this end, we performed a literature survey (1992–2021) to obtain detailed information regarding key aspects of studies investigating cactus evolution. Specifically, we summarize the advances in the following aspects: molecular markers, species delimitation, phylogenetics, hybridization, biogeography, and genome assemblies. In brief, we observed substantial growth in the studies conducted with molecular markers in the past two decades. However, we found biases in taxonomic/geographic sampling and the use of traditional markers and statistical approaches. We discuss some methodological and social challenges for engaging the cactus community in genomic research. We also stressed the importance of integrative approaches, coalescent methods, and international collaboration to advance the understanding of cactus evolution.

Keywords: Cactaceae; literature survey; evolutionary genetics; genomics

Citation: Franco, F.F.; Amaral, D.T.; Bonatelli, I.A.S.; Romeiro-Brito, M.; Telhe, M.C.; Moraes, E.M. Evolutionary Genetics of Cacti: Research Biases, Advances and Prospects. *Genes* **2022**, *13*, 452. https://doi.org/10.3390/genes13030452

Academic Editor: Zengjian Jeffrey Chen

Received: 25 January 2022
Accepted: 25 February 2022
Published: 1 March 2022

1. Introduction

Cactaceae (Caryophyllales, Angiosperms) is the major lineage of succulent plants, originating during the Eocene-Oligocene transition [1–3], and it is recognized by its remarkable diversity [4,5]. It is one of the most conspicuous examples of species radiation in the Americas, with intense diversification in the last 10 million years [1]. Cactus species have been traditionally studied by morphologists, and several taxonomic reorganizations have been proposed in the last century (e.g., [4,6–8]). As an emblematic and ecologically relevant group, many evolutionary studies using genetic-based markers have been published in recent decades (e.g., [9–15]). However, in several examples, classical molecular markers often lack sufficient information to resolve phylogenetic relationships (e.g., [16–19]) and genetic variation at intraspecific level (e.g., [20–22]). Considering the efforts undertaken so far in sampling traditional molecular markers for the family Cactaceae, and that the increasing accessibility to new sequencing technologies [23,24] has fostered genomic sampling in Cactaceae, we believe that a turning point has been reached and that this is the appropriate context for a review of current knowledge achieved using those molecular markers.

Evolutionary studies using molecular markers have been routinely performed to address questions on the natural history of living organisms, such as life histories, population genetics, gene flow, taxonomy, and evolutionary relationships [25]. Moreover, target taxa

Romeiro-Brito, M., Khan, G., Perez, M. F., Zappi, D. C., Taylor, N. P., Olsthoorn, G., ... & Moraes, E. M. (2023). Revisiting phylogeny, systematics, and biogeography of a Pleistocene radiation. American Journal of Botany, 110(3), e16134.

Received: 14 August 2022 | Accepted: 5 January 2023

DOI: 10.1002/ajb2.16134

RESEARCH ARTICLE

Botany

Revisiting phylogeny, systematics, and biogeography of a Pleistocene radiation

Monique Romeiro-Brito[1] | **Gulzar Khan**[2] | **Manolo F. Perez**[3] | **Daniela C. Zappi**[4] | **Nigel P. Taylor**[5] | **Gerardus Olsthoorn**[6] | **Fernando F. Franco**[1] | **Evandro M. Moraes**[1]

[1]Departamento de Biologia, Universidade Federal de São Carlos (UFSCar), Sorocaba, SP 18052-780, Brazil

[2]Institute for Biology and Environmental Sciences, Carl von Ossietzky-University Oldenburg, Carl von Ossietzky-Str. 9-11, 26111 Oldenburg, Germany

[3]Departamento de Genética e Evolução, Universidade Federal de São Carlos (UFSCar), São Carlos, SP 13565-905, Brazil

[4]Programa de Pós-Graduação em Botânica, Instituto de Ciências Biológicas, Universidade de Brasília (UNB), PO Box 04457, Brasília, DF 70910-970, Brazil

[5]University of Gibraltar, Gibraltar Botanic Gardens Campus, The Alameda, PO Box 843, GX11 1AA, Gibraltar

[6]Citizen scientist, PO Box 169, Holambra, São Paulo 13825-000, Brazil

Correspondence

Evandro M. Moraes, Departamento de Biologia, Universidade Federal de São Carlos (UFSCar), Sorocaba, SP 18052-780, Brazil.
Email: emarsola@ufscar.br

Abstract

Premise: *Pilosocereus* (Cactaceae) is an important dry forest element in all subregions and transitional zones of the neotropics, with the highest diversity in eastern Brazil. The genus is subdivided into informal taxonomic groups; however, most of these are not supported by recent molecular phylogenetic inferences. This lack of confidence is probably due to the use of an insufficient number of loci and the complexity of cactus diversification. Here, we explored the species relationships in *Pilosocereus* in more detail, integrating multilocus phylogenetic approaches with the assessment of the ancestral range and the effect of geography on diversification shifts.

Methods: We used 28 nuclear, plastid, and mitochondrial loci from 54 plant samples of 31 *Pilosocereus* species for phylogenetic analyses. We used concatenated and coalescent phylogenetic trees and Bayesian models to estimate the most likely ancestral range and diversification shifts.

Results: All *Pilosocereus* species were clustered in the same branch, except *P. bohlei*. The phylogenetic relationships were more associated with the geographic distribution than taxonomic affinities among taxa. The genus began diversifying during the Plio-Pleistocene transition in the Caatinga domain and experienced an increased diversification rate during the Calabrian age.

Conclusions: We recovered a well-supported multispecies coalescent phylogeny. Our results refine the pattern of rapid diversification of *Pilosocereus* species across neotropical drylands during the Pleistocene and highlight the need for taxonomic rearrangements in the genus. We recovered a pulse of diversification during the Pleistocene that was likely driven by multiple dispersal and vicariance events within and among the Caatinga, Cerrado, and Atlantic Forest domains.

KEYWORDS

Cactaceae, Cereeae, dry forest, multilocus phylogeny, phylogenetic discordance, recent radiation, South America, species tree

Recent and rapid diversification usually leads to unresolved phylogenies containing polytomies (nodes with three or more branches) due to shorter internodes located deeper in the phylogeny (Whitfield and Lockhart, 2007). The most common approach to obtain a well-supported phylogeny in a system with rapid diversification is to increase the number of loci with sufficient polymorphism (Brito and Edwards, 2009). However, phylogenetic inferences at the species level have traditionally been performed with a limited set of plastid loci, generally rendering low resolution, especially in recently radiated taxa. In addition to the difficulties imposed by using limited molecular information, the usual procedure of concatenating all loci into a supergene (supermatrix approach) imposes additional concerns on phylogenetic inferences. For instance, this approach forces the inference of a single topology with the same coalescence time for all