

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

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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 Cactaceae591 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 $3 \times$ tiling 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 $8\times$, 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)
<i>Cereus</i> 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)
<i>Cereus</i> -noncoding	560,025	15.26	53,768	23,603 (0.042)
<i>Cereus</i> -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*.

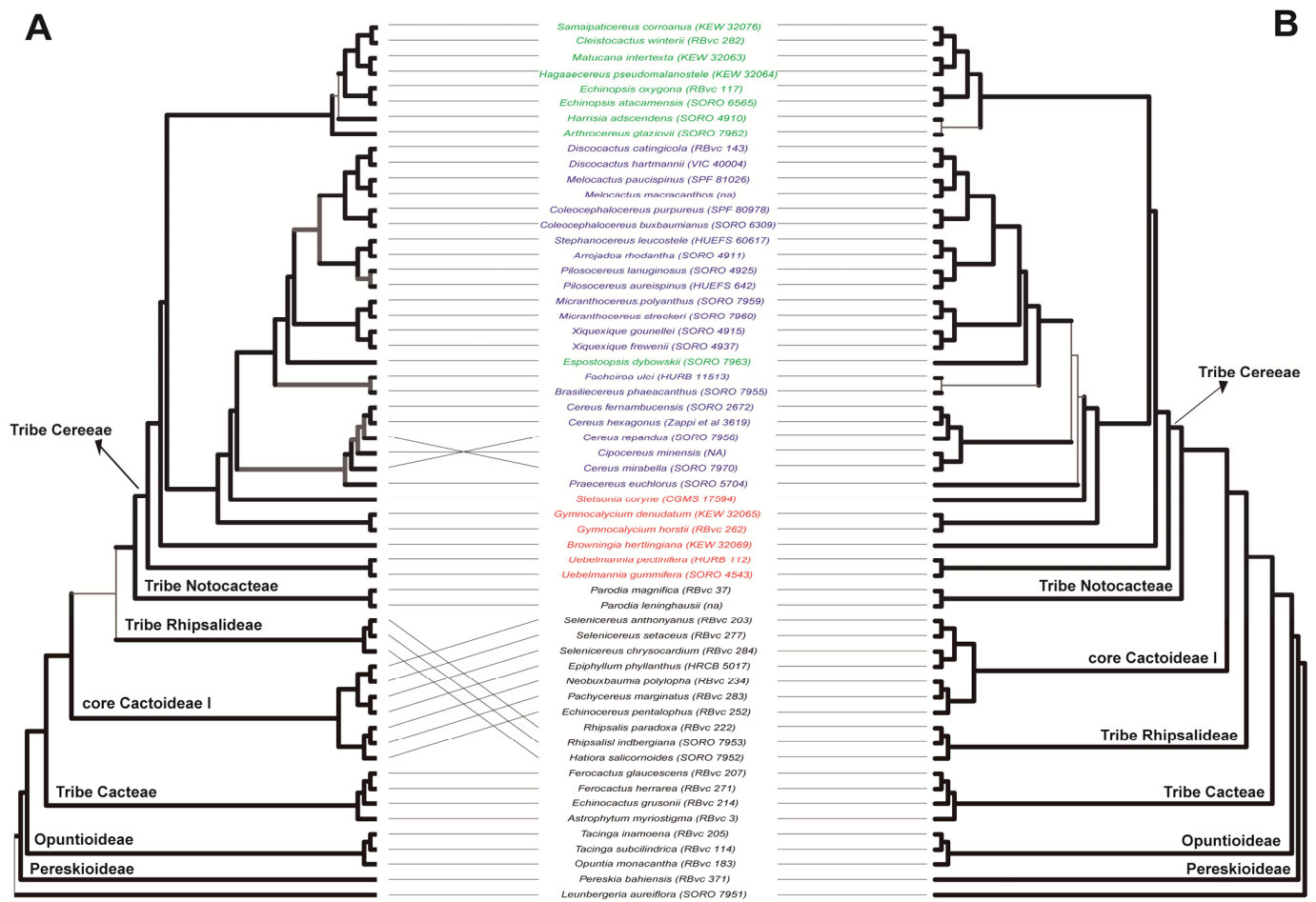


Figure 1. Comparison of coalescent-based species trees of Cactaceae estimated with Cactaceae-coding (A) and Cactaceae-all (B) sub-datasets. Cactaceae-coding and Cactaceae-all sub-datasets included 290 and 512 nuclear regions, respectively. The branches with high support (LPP > 0.95) are colored with 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).

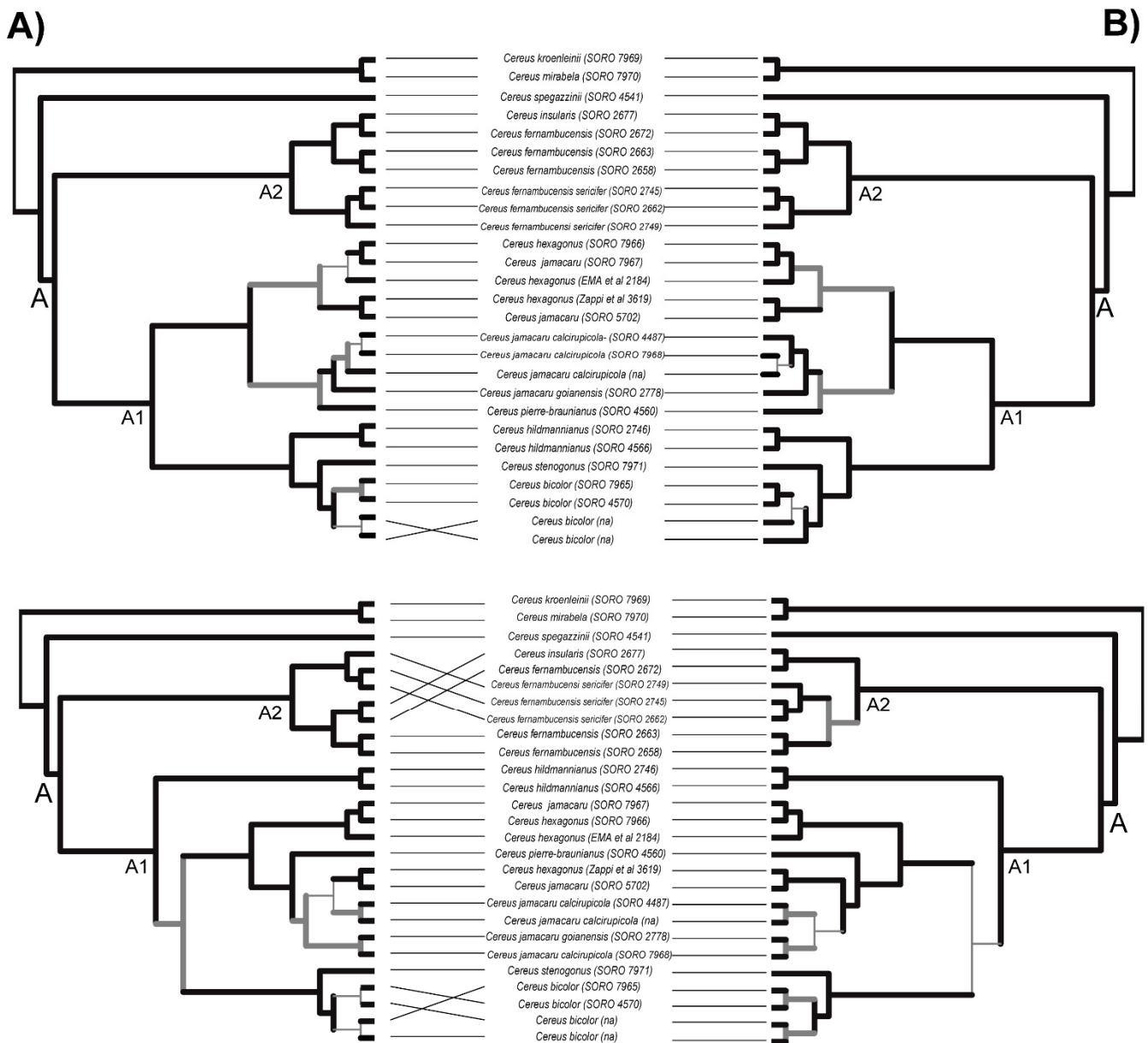


Figure 2. Comparison of phylogenies of Cereus clade A inferred with Cereus-noncoding (A) and Cereus-all (B) sub-datasets, using coalescent-based (upper) and ML concatenated (down) approaches. The branches with high support (LPP > 0.95 and BS > 95) are highlighted with thick branch lines, the branches 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 Egli [48] for the tribe Cereeae, our results recovered the subtribe Trichocereinae as polyphyletic and the subtribes Cereinae and Rebutinae as paraphyletic groups. For instance, *Espositoopsis 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

- 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](#)]
- 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](#)]
- 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](#)]
- Bárceñas, R.T.; Yesson, C.; Hawkins, J.A. Molecular systematics of the Cactaceae. *Cladistics* **2011**, *27*, 470–489. [[CrossRef](#)]
- 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](#)]
- 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](#)]
- 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](#)]
- 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](#)]
- Maddison, W.P. Gene trees in species trees. *Syst. Biol.* **1997**, *46*, 523–536. [[CrossRef](#)]
- 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](#)]
- 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](#)]
- 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](#)]
- 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](#)]
- 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](#)]
- 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](#)]
- 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](#)]
- 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](#)]
- 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](#)]
- 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](#)]
- 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](#)]
- 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 Papuanian 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, Cereaceae) 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.; Jermin, 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. Schlumberger, 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 Fakultät der Universität Zurich, Zurich, Switzerland, 2013.
 73. Taylor, N.; Zappi, D. Notes on plants called *Cereus hexagonus* (Cactaceae). *Bradleya* **2019**, *37*, 17–25. [[CrossRef](#)]