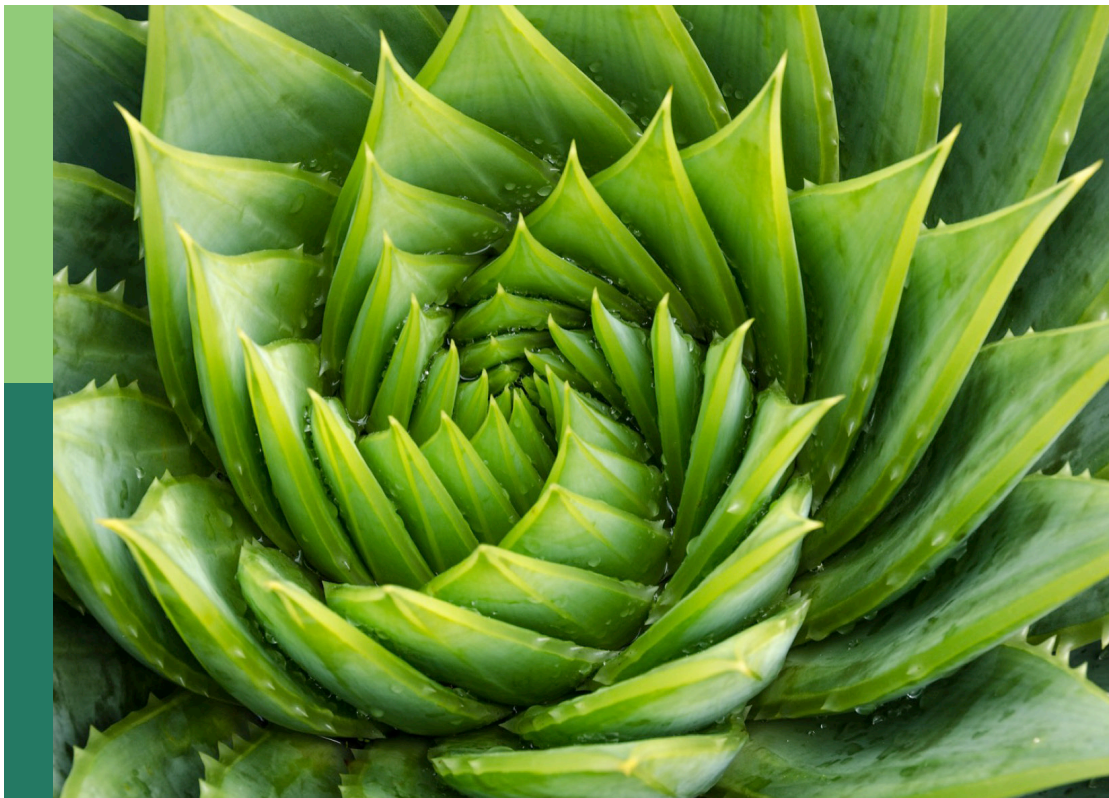
frontiers | Research Topics

Genomics, functional, evolutionary, and ecological perspectives on the biology of carnivorous plants

Edited by
Magdy S. Alabady, Jim Leebens-Mack, Kenji Fukushima, Jeremy D. Rentsch and Russell L. Malmberg

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-5780-8
DOI 10.3389/978-2-8325-5780-8

About Frontiers

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

Frontiers journal series

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

Dedication to quality

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

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Genomics, functional, evolutionary, and ecological perspectives on the biology of carnivorous plants

Topic editors

Magdy S. Alabady — University of Georgia, United States
Jim Leebens-Mack — University of Georgia, United States
Kenji Fukushima — National Institute of Genetics, Japan
Jeremy D. Rentsch — Francis Marion University, United States
Russell L. Malmberg — University of Georgia, United States

Citation

Alabady, M. S., Leebens-Mack, J., Fukushima, K., Rentsch, J. D., Malmberg, R. L., eds. (2024). *Genomics, functional, evolutionary, and ecological perspectives on the biology of carnivorous plants*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-5780-8

Table of contents

frontiers in Plant Science

ORIGINAL RESEARCH
published: 20 January 2021
doi: 10.3389/fpls.2020.625507

Transcriptomic and Proteomic Analyses of *Nepenthes ampullaria* and *Nepenthes rafflesiana* Reveal Parental Molecular Expression in the Pitchers of Their Hybrid, *Nepenthes* × *hookeriana*

Muhammad Mu'izzuddin Zulkapli, Nur Syatila Ab Ghani, Tiew Yik Ting, Wan Mohd Aizat and Hoe-Han Goh*

Institute of Systems Biology, Universiti Kebangsaan Malaysia, Bangi, Malaysia

OPEN ACCESS

Edited by:
Kenji Fukushima,
Julius Maximilian University of Würzburg, Germany

Reviewed by:
Andrej Pavlovič,
Palacký University, Olomouc, Czechia
Axel Mithöfer,
Max Planck Institute for Chemical Ecology, Germany

***Correspondence:**
Hoe-Han Goh
gohhh@ukm.edu.my

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

Received: *03 November 2020*
Accepted: *01 December 2020*
Published: *20 January 2021*

Citation:
Zulkapli MM, Ab Ghani NS, Ting TY, Aizat WM and Goh H-H (2021) Transcriptomic and Proteomic Analyses of Nepenthes ampullaria and Nepenthes rafflesiana Reveal Parental Molecular Expression in the Pitchers of Their Hybrid, Nepenthes × hookeriana. Front. Plant Sci. 11:625507. doi: 10.3389/fpls.2020.625507

Nepenthes is a genus comprising carnivorous tropical pitcher plants that have evolved trapping organs at the tip of their leaves for nutrient acquisition from insect trapping. Recent studies have applied proteomics approaches to identify proteins in the pitcher fluids for better understanding the carnivory mechanism, but protein identification is hindered by limited species-specific transcriptomes for *Nepenthes*. In this study, the proteomics informed by transcriptomics (PIT) approach was utilized to identify and compare proteins in the pitcher fluids of *Nepenthes ampullaria*, *Nepenthes rafflesiana*, and their hybrid *Nepenthes* × *hookeriana* through PacBio isoform sequencing (Iso-Seq) and liquid chromatography-mass spectrometry (LC-MS) proteomic profiling. We generated full-length transcriptomes from all three species of 80,791 consensus isoforms with an average length of 1,692 bp as a reference for protein identification. The comparative analysis found that transcripts and proteins identified in the hybrid *N.* × *hookeriana* were more resembling *N. rafflesiana*, both of which are insectivorous compared with omnivorous *N. ampullaria* that can derive nutrients from leaf litters. Previously reported hydrolytic proteins were detected, including proteases, glucanases, chitinases, phosphatases, nucleases, peroxidases, lipid transfer protein, thaumatin-like protein, pathogenesis-related protein, and disease resistance proteins. Many new proteins with diverse predicted functions were also identified, such as amylase, invertase, catalase, kinases, ligases, synthases, esterases, transferases, transporters, and transcription factors. Despite the discovery of a few unique enzymes in *N. ampullaria*, we found no strong evidence of adaptive evolution to produce endogenous enzymes for the breakdown of leaf litter. A more complete picture of digestive fluid protein composition in this study provides important insights on the molecular physiology of pitchers and carnivory mechanism of *Nepenthes* species with distinct dietary habits.

Keywords: carnivory, hybrid, proteomics, transcriptomics, *Nepenthes*, pitcher fluid

INTRODUCTION

Nepenthes comprises unique carnivorous tropical plants with pitcher organs at the ends of leaf tips for the capture, digestion, and absorption of insects to grow in nutrient-poor soil. There are more than 150 *Nepenthes* species geographically distributed in Southeast Asia, including Borneo, Philippines, and Sumatra (Murphy et al., 2020). The species diversification of this genus in their pitcher morphological features, ecology, and nutrient acquisition attracted many evolutionary studies of *Nepenthes* (Clarke and Moran, 2016).

Nepenthes species are mostly insectivorous but previous studies showed that *Nepenthes ampullaria*, which is predominantly found in the heath and swamp forests compared to open habitats like other species, possesses detritivore traits to trap leaf litter as a nutrient source (Moran et al., 2003). This genus is also well-known for natural and artificial hybridization. One of the common natural hybrids is *Nepenthes* × *hookeriana*, between *N. ampullaria* and *Nepenthes rafflesiana*. This hybridization was initially identified based on their common morphological characters (Clarke and Wong, 1997; Clarke, 2001) and later verified through protein and genetic marker analyses (Yulita and Mansur, 2012; Biteau et al., 2013), which also suggested a closer relationship between *N.* × *hookeriana* and *N. rafflesiana* than *N. ampullaria*. To date, there is no comprehensive report comparing the molecular expression in the pitchers and pitcher fluids of the hybrid and *N. rafflesiana* as carnivores, with *N. ampullaria* being an omnivore. This gap of knowledge was pointed out by Pavlovič (2012) on the adaptive radiation of *Nepenthes* nutrient sequestration strategies.

The pitchers with acidic fluids and secreted enzymes are important for trapping and digesting invertebrate prey (Ravee et al., 2018; Gilbert et al., 2020). Several digestive enzymes are commonly reported to be secreted into the pitcher fluids, which include aspartic proteases, nucleases, and pathogenesis-related (PR) proteins (Athauda et al., 2004; Hatano and Hamada, 2012; Buch et al., 2014; Rottloff et al., 2016; Fukushima et al., 2017). In comparison, knowledge of nutrient uptake and transportation is more limited. Furthermore, the regulatory mechanism of protein secretion and replenishment remains poorly understood (Wan Zakaria et al., 2016b, 2019; Goh et al., 2020). Protein identification in the pitcher fluids is limited by the unusual amino acid composition and the limited sequence information for *Nepenthes* (Lee et al., 2016). There are only 760 UniProtKB entries for the taxonomy *Nepenthes* as of August 2020, the majority of which are the maturase K and ribosomal protein sequences apart from those digestive enzymes mentioned above.

Species-specific transcriptome sequences are ideal for protein identification. Hence, we applied proteomics informed by transcriptomics (PIT) approach in this study to compare protein profiles of the hybrid *N.* × *hookeriana* with its parents *N. rafflesiana* and *N. ampullaria* for comparative protein profile analysis to elucidate the pitcher fluid protein composition of each species. Comparing the molecular profiles among the three species not only can explore the differences in fluid protein composition due to dietary habits but also validate their relationship. Furthermore, this study provides a reference list of endogenous proteins secreted upon pitcher opening for further studies on the regulation of secreted proteins into the pitcher fluids.

MATERIALS AND METHODS

Plant Materials

Nepenthes plants (*N. ampullaria*, *N. rafflesiana*, and *N.* × *hookeriana*) originated from the Endau Rompin National Park Malaysia were grown in a common garden at the experimental terrace (2°55′11.5″N 101°47′01.4″E) of Universiti Kebangsaan Malaysia under natural weather conditions. Developing pitchers were monitored daily and covered with mesh nets to prevent insect entry. Newly opened pitchers within 24 h of lid opening were harvested between June to August 2015 in the morning 9:00–10:00 am (Zulkapli et al., 2017). All of the fluids from individual pitchers were poured into separate Falcon tubes, while the whole pitcher tissues excluding the tendril were kept in separate plastic bags and frozen in liquid nitrogen before stored at −80°C.

PacBio Isoform Sequencing

Total RNA was extracted from pitcher tissues using the modified cetyltrimethylammonium bromide (CTAB) protocol (Abdul-Rahman et al., 2017). The quality and integrity of extracted total RNA were determined using Nanodrop ND-1000 (Thermo Fisher Scientific Inc., United States) and Agilent 2,100 bioanalyzer (Agilent Technologies, United States), respectively. Total RNA with RNA integrity number (RIN) >8 was submitted for library preparation and sequencing at Icahn Medical Institute (Mount Sinai, New York City, United States). One replicate per species with the highest RIN was chosen for sequencing. Full-length cDNAs were prepared using SMARTer PCR cDNA synthesis kit (Clontech) according to the manufacturer's protocols. Double-stranded cDNAs were subjected to size selection using BluePippin (Sage Science, MA, United States) at the MW range of 1–3 kb. The PCR amplification profile was 95°C for 2 min, 15 cycles × (98°C for 20 s, 65°C for 15 s, 72°C for 4 min), and a final extension at 72°C for 5 min. Due to low yield after selection, *N. ampullaria* sample was further amplified for nine cycles (98°C for 20 s, 65°C for 15 s, 72°C for 1 min 45 s) and was size-selected again before preparing the SMRTbell library with a minimum of 1 μg of dsDNA based on the manufacturer's SMRTbell template protocol (SMRTbell Template Preparation Kit 1.0). The SMRTbell libraries were purified by two sequential 0.45 × AMPure PB purifications (Pacific Biosciences) after exonuclease digestion of incomplete SMRTbell templates. Libraries were quantified by fluorimetry and assayed for quantity and size distribution by Bioanalyzer. A single SMRTbell library for individual species was sequenced using SMRT Cell v3 with P6-C4 chemistry on the PacBio RS II platform (Pacific Biosciences), each at a concentration of 110 pM (Zulkapli et al., 2017).

Subreads <300-bp and reads with quality <0.75 (corresponding to a predicted error rate of >25%) were filtered out. Sub-reads were filtered and subjected to circular consensus sequence (CCS) read analysis using PacBio SMRT Analysis Server v2.3.0 following the RS_IsoSeq protocol. In brief, cDNA primer and poly-A tails were removed and the read of inserts (ROIs) were classified into full-length and non-full-length. Iterative clustering for error correction (ICE) algorithm was also used and quiver polishing was performed to generate consensus isoform sequences at a high QV value of 0.99 and expected size of 1–2 kb. For the reference transcriptome, raw reads from all three species were combined for the CCS read analysis.

Raw sequences were deposited into the Sequence Read Archive (SRA) under BioProject PRJNA299862 with the following identifiers: SRX2692198 (*N. ampullaria*), SRX2692197 (*N. rafflesiana*), and SRX2692196 (*N.* × *hookeriana*) (Zulkapli et al., 2017). Consensus isoform sequences can be accessed from the TSA repository: GGLJ00000000.1 (*N. ampullaria*), GGLG00000000.1 (*N. rafflesiana*), and GGLF00000000.1 (*N.* × *hookeriana*).

Transcriptomics Analysis

Bioinformatics methods were applied to analyze the consensus isoform sequences aiming to compare the transcripts between the three *Nepenthes* species, including BLAST, TransDecoder, Trinotate, OrthoVenn, WEGO, and KAAS. Consensus isoform sequences of the hybrid *N.* × *hookeriana* were searched against the consensus isoform sequences of *N. ampullaria* and *N. rafflesiana* using local BLASTN v.2.6.0 with an *E*-value cut-off of $1e^{-5}$.

Trinotate (Bryant et al., 2017) was used for functional annotation of individual transcriptomes based on different methods that include homology search (BLAST+/UniProt), protein domain identification (HMMER/Pfam), prediction of signal peptide (SignalP), and transmembrane domain (TmHMM), as well as searches against eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups), Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases.

The GO annotations from Trinotate report were plotted using Web Gene Ontology Annotation (WEGO) (Tyanova et al., 2016), with further selection on the GO groups related to four unique physiology of carnivorous plants, namely trapping, digestion, absorption, and defense. The functional annotation in KEGG GENES database was obtained using KEGG Automatic Annotation Server (KAAS) that assigns KEGG Orthology (KO) to KEGG pathways given a set of protein sequences as input (Moriya et al., 2007). Venn diagram analysis was performed using Venny version 2.1.0 (Oliveros, 2007-2015).

Protein-coding sequences (CDS) predicted from the consensus isoform sequences using TransDecoder (Haas et al., 2013) were used as a reference dataset for protein identification and comparative analysis using OrthoVenn2 (Xu et al., 2019) with default parameters: *E*-value cutoff of $1e^{-5}$ for all-to-all similarity comparisons and the inflation value of 1.5 for the generation of orthologous clusters using the Markov Cluster Algorithm. The reference predicted protein sequences[1] were further annotated using eggNOG 5.0 (Huerta-Cepas et al., 2019) for functional categorization using eMapper v2.0[2] with default settings. Overrepresentation and underrepresentation analyses of KEGG pathway and eggNOG were performed using the hypergeometric test function in MS Excel.

Protein Extraction and Processing

Pitcher fluids were filtered through 25 mm acrodisc syringe filter with PVDF membrane of 0.2 μm pore size (Pall, United States) pre-conditioned using 1 mL UHP water (Mili-Q). Proteins were then concentrated by ultrafiltration through a Microsep Advance Centrifugal Devices with Omega membrane (Pall, United States) at a molecular weight cut-off of 10 kDa, rinsed with 1 mL of UHP water. Supernatants (>10 kDa) were collected and further concentrated to 100 μL through speed vacuum (Wan Zakaria et al., 2018). Aliquots of 20 μL were used for sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and the remaining 80 μL were pooled together from nine biological replicates for LC-MS/MS analysis.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis was performed using Bio-Rad electrophoresis apparatus (Bio-Rad, United States). Loading buffer (0.2 M Tris–HCl [pH 6.8], 10% SDS, 20% glycerol, 10 mM β-mercaptoethanol, 1% bromophenol blue, and water) was added to the protein sample and heated at 95°C for 10 min to break down the disulfide bonds. Two gel layers were prepared, namely the stacking gel with bis-acrylamide concentration of 12.5% (pH 8.8) and the separating gel with bis-acrylamide concentration of 4% (pH 6.8). The electrophoresis was performed at 75 V for 25 min followed by 125 V for 90 min. The MS-incompatible silver staining method was performed (Wan Zakaria et al., 2019). Gels were fixed in fixation solution with 30% ethanol and 10% acetic acid overnight before washed thoroughly in 30% ethanol for 20 min and soaked in distilled water for 20 min followed by sensitivity enhancing solution containing 8 mM sodium thiosulfate pentahydrate. The gels were washed thrice with distilled water and soaked in silver stain solution containing 11 mM silver nitrate and 0.15% formaldehyde. Then, the gels were washed thrice and soaked in the development solution with 0.5 M sodium carbonate and 0.2% formaldehyde. Once protein strips appeared, the reaction was stopped by washing the gels with stop solution containing 0.5 mM EDTA for 10 min. Gels were rinsed with distilled water, analyzed, and captured via Densitometer accompanied with Quantity One version 4.6.7 (Bio-Rad, United States). The protein band sizes were estimated using the BLUEstain protein ladder (11–245 kDa) (GoldBio).

Proteomics Analysis of Pitcher Fluids Using nanoLC-MS/MS

For gel-free liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis, solid phase extraction (SPE) was performed using the 1 cc Oasis HLB cartridges containing the

[1] https://doi.org/10.6084/m9.figshare.13270241.v1

[2] http://eggnog-mapper.embl.de/

Oasis HLB sorbent (OASIS, United States). SPE eluent was dried through speed vacuum and was rehydrated by 170 μL of 0.5% formic acid with 20 μL aliquot for SDS-PAGE and the remaining 150 μL was used for in-solution digestion. Trypsin digestion was conducted at a ratio of 1:100 (Wan Zakaria et al., 2018). Digested sample dried through speed vacuum and rehydrated by 35 μL of 0.5% formic acid with 3 μL aliquot for SDS-PAGE and the remaining was used for Zip-Tip protocol. Zip-Tip protocol was performed using Thermo Scientific Pierce C18 Tips (Thermo, United States) prior to MS analysis. Sample was speed vacuum and rehydrated with 45 μL of 0.1% formic acid prior to LC-MS/MS analysis.

All experiments were performed in a nanoflow LC system, Easy-nLC (Thermo) equipped with EASY-Spray Column Acclaim PepMap C18 and coupled (Thermo) to Orbitrap Fusion Tribrid mass spectrometer (Thermo, United States). Protein samples were loaded onto the pre-column and the peptides were analyzed using linear-gradient program with flow rate of 300 nL/min for 0.1% formic acid (solution A) and 0.1% formic acid in acetonitrile (solution B) and gradients were set as followed: (i) 5–40% B for 91 min, (ii) 95% B for 2 min, (iii) 95% B for 6 min, and (iv) 5% B for 2 min. Each pooled sample from nine biological replicates was injected three times in two independent analyses for data collection resulting in six spectra for each species.

MS/MS data were retrieved using mass spectrometer LTQ Orbitrap XL (Thermo Scientific). Full scan profile mass spectra (OTMS1) was obtained using the following parameters in top speed mode: scan range (m/z) = 201–1800 Da; cycle time = 3 s; resolution = 120,000; AGC target = $4.0e^5$; maximum injection time = 50 ms; precursor selection with charge state of 2–7; dynamic exclusion duration = 20 s; intensity = 5000. The parameters used for MSMS (ITMS2) analyses were as followed: rapid scan rate; CID NCE = 30%; HCD NCE = 28%; isolation window = 1.6 m/z; AGC target = $1.0e^2$; maximum injection time = 250 ms. Raw data, sequence files, and results were deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol et al., 2019) with data set identifier PXD007599.

Protein Identification

Liquid chromatography tandem mass spectrometry raw files were processed for peptide identification using MaxQuant version 1.5.3.30 (Tyanova et al., 2016) through peptide to spectrum matching (PSM) pipeline using three digestion (Trypsin/P) settings: specific, semispecific, and unspecific. Carbamidomethylation and methionine oxidation were used as fixed and variable modifications. The MS/MS spectrum was searched through Andromeda searching engine (Cox et al., 2011) integrated with MaxQuant, against the reference data set of predicted proteins obtained from TransDecoder, in addition to 55 previously reported protein sequences (Lee et al., 2016; Rottloff et al., 2016) with 358 potential contaminants and reversed sequences. Initial mass tolerance was set to 4.5, 20 ppm on MS, and 0.5 Da on MS/MS. For peptide identification, the false discovery rate was set to 0.01, minimum peptide length was 7 amino acids and the maximum mis-cleavages allowed were 2. For peptide quantification, MaxLFQ algorithm (Cox et al., 2014) was used based on default parameters with minimum ratio count set to 1. For peptide matching, the retention time window was set to 30 s. Proteins with more than one peptide or one peptide with at least one MS/MS and intensity values greater than 0 were considered identified and present. Proteins obtained from MaxQuant identification were used for sequence comparisons using constraint-based alignment tool (COBALT) (Papadopoulos and Agarwala, 2007) and Clustal Omega (Sievers and Higgins, 2018).

RESULTS

Transcriptomics and proteomics analyses of *N. ampullaria, N. rafflesiana*, and their hybrid, *N.* × *hookeriana* were conducted with an overview of the methods illustrated in **Supplementary Figure 1**.

Transcriptome Profiling of *Nepenthes* through PacBio Isoform Sequencing

The transcriptome libraries of individual *Nepenthes* species were generated using PacBio isoform sequencing (Iso-Seq). A total of 26,130, 30,558, and 33,279 consensus isoforms were generated for *N. ampullaria*, *N. rafflesiana*, and *N.* × *hookeriana*, respectively, with an average length of 1,625, 1,680, and 1,722 bp (**Supplementary Table 1**). The three transcriptomes from individual *Nepenthes* species were combined to form a reference transcriptome containing a total of 80,791 consensus isoforms with an average length of 1,692 bp. Despite having the lowest number of read of insert (ROI), the hybrid expressed the highest number of consensus isoforms, indicating more varied transcripts in the hybrid. The local BLASTN search of the hybrid consensus isoforms against the parents found more hits with *N. rafflesiana* (93.2%) than *N. ampullaria* (89.2%) at similarity >60%, which suggests more similar transcriptome sequences between the hybrid and *N. rafflesiana* (**Supplementary Figure 2**).

Functional annotation for individual *Nepenthes* transcriptomes and the reference was performed through the Trinotate pipeline using transcript and predicted peptide sequences (**Table 1** and **Supplementary File 1**). Homology searches of reference transcriptome performed using BLASTX and BLASTP found 53,917 (66.7%) hits to UniProt, 34,524 (42.7%) hits to Pfam, 22,560 (27.9%) hits to eggNOG, 37,816 (46.8%) hits with KO, and 58,635 (72.5%) hits to Arabidopsis genes. In total, 64,455 (79.8%) consensus isoforms of the reference transcriptome were functionally annotated with at least one database. Furthermore, predictions of signal peptides (SignalP) and transmembrane helices (TmHMM) annotated 3,121 (3.9%) and 10,510 (13.0%) peptides, respectively.

The annotated transcriptomes of the three *Nepenthes* species were assigned with 495 GO terms according to three main GO categories: biological process, molecular function, and cellular component with 310, 100, and 85 terms, respectively, (**Supplementary File 2** and **Supplementary Table 2**). The distribution of GO annotation was visualized in WEGO analysis

TABLE 1 | Functional annotation of *Nepenthes* transcriptomes using Trinotate pipeline.

Number	*N. ampullaria*	*N. rafflesiana*	*N.* × *hookeriana*	References
Consensus isoform	26,130	30,558	33,279	80,791
Protein (CDS)	19,463	26,677	30,096	65,757
Functional annotation				
BLASTX	16,540	21,615	23,587	53,440
BLASTP	11,139	15,094	17,176	37,078
BLASTX/P	16,664	21,814	23,768	53,917
GO_BLAST	15,198	19,782	21,726	48,887
Pfam	10,455	14,192	16,103	34,524
GO_Pfam	7,836	10,408	12,107	25,561
eggNOG	6,984	9,045	9,896	22,560
SignalP	918	1,265	1,500	3,121
TmHMM	3,150	4,219	4,990	10,510
KO	11,785	15,170	16,553	37,816
AGI	18,258	23,491	25,696	58,535

BLASTX, hits of isoform sequences to UniProt database; BLASTP, hits of predicted protein sequences to UniProt database; BLASTX/P, combined hits from BLASTX and BLASTP; GO_BLAST, GO annotation from BLAST; Pfam, hits to Pfam database; GO_Pfam, GO annotation from Pfam; eggNOG, hits to orthologous genes in eggNOG database; SignalP, prediction of signal peptide; TmHMM, prediction of transmembrane protein; KO, KEGG Orthology from KAAS; AGI, BLASTN hits to Arabidopsis genes.

(**Figure 1**). The top terms for cellular component were cell, cell part, and organelle; for molecular function were the catalytic activity and binding; while for biological process were cellular process, metabolic process, and single-organism process, which showed significant differences between the three species. The analysis showed more significant differences of GO annotations between the two parent species (*N. ampullaria vs. N. rafflesiana*, 15.4%) than between the hybrid and *N. rafflesiana* (8.7%) compared to *N. ampullaria* (12.3%), indicating more similar functional genes between the hybrid and *N. rafflesiana* (**Supplementary Table 2**).

Meanwhile, KEGG annotation using KAAS against the KEGG GENES database found hits to 2,432, 2,846, and 2,663 KO, which mapped to 395, 398, and 396 KEGG pathways for *N. ampullaria, N. rafflesiana*, and *N.* × *hookeriana*, respectively, (**Supplementary File 2** and **Supplementary Table 3**). Majority (392) of the KEGG pathways were common in all three species. Two KEGG pathways, mucin type O-glycan biosynthesis and glycosaminoglycan biosynthesis-keratan sulfate with beta-galactoside alpha-2,3-sialyltransferase (K00780), were unique to *N. ampullaria*. A protein O-mannose beta-1,4-N-acetylglucosaminyltransferase for mannose type O-glycan biosynthesis (K18207) was unique to *N. rafflesiana*; while a 2,4-dihydroxy-1,4-benzoxazin-3-one-glucoside dioxygenase (K13229) for benzoxazinoid biosynthesis was unique in the hybrid. Meanwhile, a UDP-N-acetylglucosamine acyltransferase (K00677) in cationic antimicrobial peptide (CAMP) resistance was found only in the parents (*N. ampullaria* and *N. rafflesiana*). Three pathways shared between the hybrid and *N. rafflesiana* were absent in *N. ampullaria*: glycosphingolipid biosynthesis - lacto and neolacto series with a lactosylceramide 4-alpha-galactosyltransferase (K01988), caprolactam degradation with an alcohol dehydrogenase (NADP+) (K00002), and a crocetin glucosyltransferase (K21371) for the biosynthesis of various secondary metabolites—part 1.

Transdecoder analysis predicted a total of 19,463, 26,677, 30,096, and 65,757 protein-CDS from 14,523, 19,683, 22,192, and 48,663 consensus isoforms, respectively, for *N. ampullaria*, *N. rafflesiana, N.* × *hookeriana*, and reference (**Table 1** and **Supplementary Table 4**). A total of 53,235 (83.5%) predicted protein sequences of the reference can be functionally categorized by eggNOG 5.0 mapper (**Supplementary File 1**). Comparative analysis by OrthoVenn using predicted protein sequences identified 3,500 orthologous protein clusters shared among the three *Nepenthes* species in which 1,676 were single-copy gene clusters with ∼65% singletons without orthologs among the species (**Figure 2** and **Supplementary Table 4**). The higher number of clusters shared between the hybrid and *N. rafflesiana* (76.3%) than with *N. ampullaria* (63.9%) corroborates the results from BLASTN analysis that reflects a closer genetic relationship between the hybrid and *N. rafflesiana* based on our samples in this study.

Proteomic Profiling of *Nepenthes* Pitcher Fluids

Proteins were extracted from nine biological replicates of pitcher fluids for each *Nepenthes* species and examined through SDS-PAGE analysis with silver staining at each step of sample processing (**Supplementary Figure 3**). These nine replicates were pooled for each species and analyzed using the nanoLC-MS/MS. MS data were processed using MaxQuant for searching against predicted protein sequences of the reference transcriptome and protein sequences from previous studies (Lee et al., 2016; Rottloff et al., 2016) through peptide spectrum matches (PSM). Proteins were identified using specific, semispecific, and unspecific digestion settings for a comprehensive analysis due to the hydrolytic proteins in the pitcher fluids with a possibility of non-specific digestion (Lee et al., 2016). The analysis identified a total of 220 proteins from *Nepenthes* pitcher fluids: 125 in *N. rafflesiana*, 113 in *N. ampullaria*, and 94 in *N.* × *hookeriana* (**Supplementary File 3**). The least number of fluid proteins was identified in the hybrid despite having the highest number of transcriptomic predicted protein sequences (**Table 1**). According to the Venn diagram analysis (**Figure 3**), more proteins were shared between *N.* × *hookeriana* and *N. rafflesiana* (50 proteins) than *N. ampullaria* (36 proteins), which is consistent with the OrthoVenn analysis. The number of unique proteins in *N. ampullaria* (51, 45%), *N. rafflesiana* (49, 39%), and *N.* × *hookeriana* (33, 35%) were proportionally higher than that of OrthoVenn analysis (3.5–7.8%, **Figure 2**). A total of 25 proteins were shared among the three species as listed in **Table 2**.

Based on the eggNOG 5.0 functional categorization of all proteins found in the pitcher fluids, "Translation, ribosomal structure and biogenesis," "Transcription," "Posttranslational modification, protein turnover, chaperones," and "Function

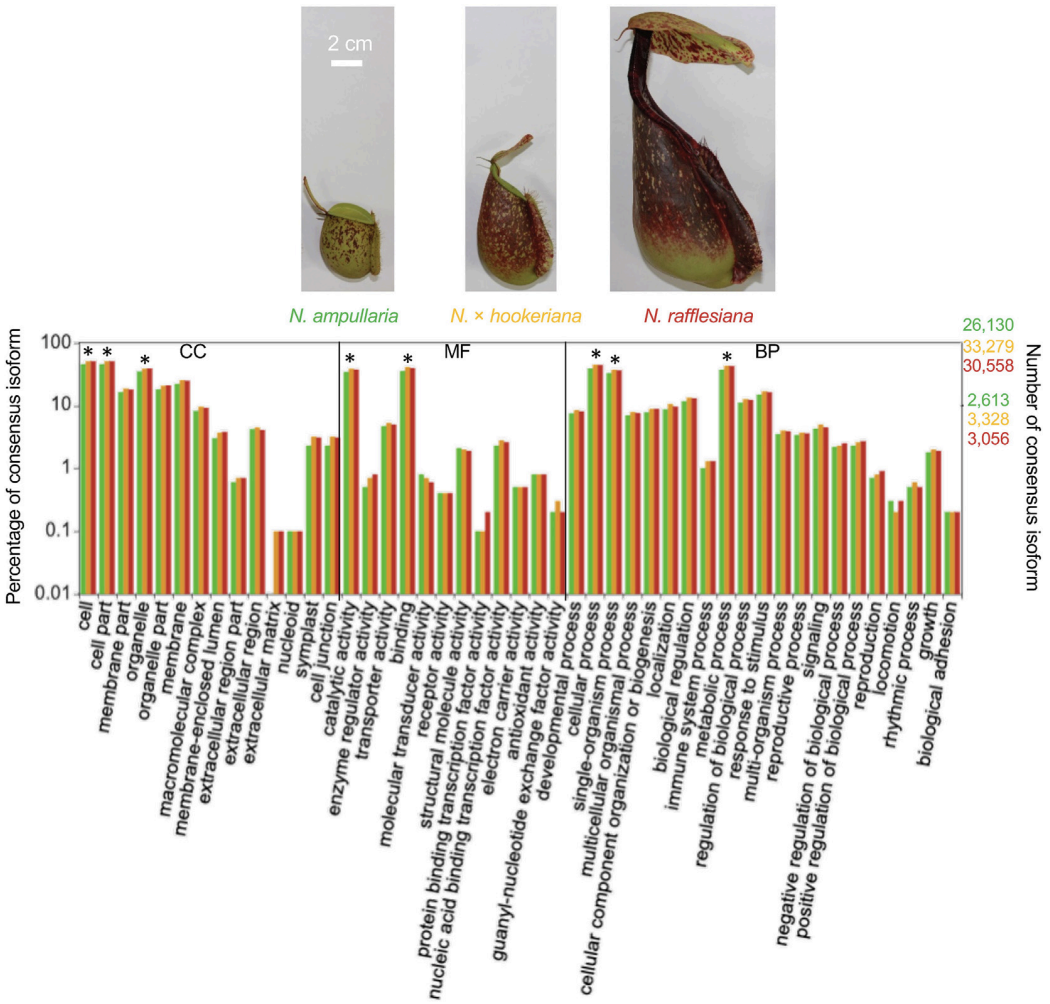


FIGURE 1 | The distribution of Gene Ontology (GO) terms of annotated consensus isoforms from all three *Nepenthes* species using WEGO based on cellular component (CC), molecular function (MF), and biological process (BP). Asterisks (*) represent significant ($P < 0.001$) differences in Pearson Chi-Square test. The bar chart is color-coded according to the font colors. Photos of the pitcher samples from the three species are displayed above with the same scale.

unknown" were found to be overrepresented ($P < 0.05$); "RNA processing and modification," "Amino acid transport and metabolism," and "Signal transduction mechanisms" were found to be underrepresented ($P < 0.05$) proportionally to the reference transcriptome (**Supplementary File 3**). There was no specific overrepresentation in individual species, except the hybrid with a disproportionally higher number of proteins with "Function unknown."

The distribution of GO annotation for shared and unique proteins in *Nepenthes* species was analyzed in WEGO (**Figure 4A**). The identified proteins from MaxQuant analysis were grouped into 18 biological processes, with eight biological processes common in all three species including cellular process, metabolic process, response to stimulus, biological regulation, cellular component organization or biogenesis, establishment of localization, and developmental process in the order of protein abundance. Two GO terms, catalytic activity (GO:0003824) and binding (GO:0005488), were annotated for more than 40% of proteins under the molecular function category, which indicates the predominant functions of proteins in the digestive pitcher fluids. Meanwhile, the comparison of GO annotation between the hybrid and parents found 17 common GO terms between the hybrid and *N. rafflesiana* in biological process compared to five with *N. ampullaria* (**Figure 4B**).

Pitcher Fluid Proteins Related to Carnivory Traits of *Nepenthes*

To explore carnivory mechanism of the three *Nepenthes* species, we focused on proteins with GO terms related to physiological properties of carnivorous plants, namely trapping, digestion, nutrient absorption, and defense. The proteins annotated with "catalytic activity" are mainly hydrolases, oxidoreductases, and transferases comprising nepenthesins, neprosins, purple acid phosphatases, lipid phosphate phosphatase, S-like ribonuclease (RNaseS), glucosidases, glucanases, and peroxidases commonly reported in *Nepenthes* pitcher fluids (**Table 3**).

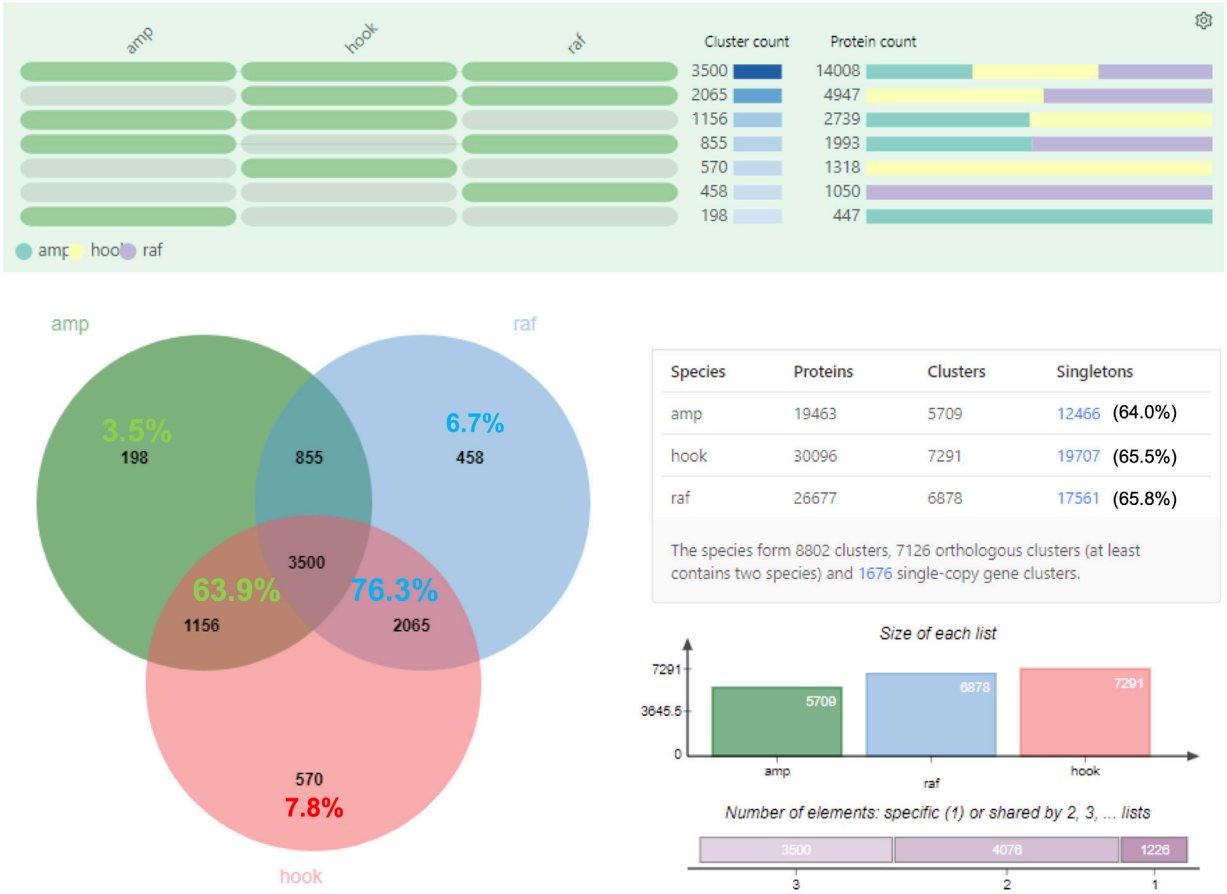


FIGURE 2 | OrthoVenn analysis on the distribution of orthologous protein clusters of predicted protein sequences. Top part shows the cluster and protein counts in different sections of the Venn diagram, which shows the number of protein clusters. Percentages in the shared regions of Venn diagram between the hybrid and parents are calculated based on the total clusters in the hybrid; the percentages in the non-shared sections are based on respective species. The table on the lower right summarizes the numbers of proteins, clusters, and singletons in each species. Singletons are protein sequences that do not form any cluster with other sequences and are shown with percentages out of total number of protein sequences in the respective species. The charts on the lower right show the size of protein clusters in each species and the cumulative numbers of shared elements based on the Venn diagram. amp: *N. ampullaria*; raf: *N. rafflesiana*; hook: *N.* × *hookeriana*.

All five reported nepenthesins were found in all species, except for Nep2 that was detected only in *N. rafflesiana* but the Nep2 transcripts were found in all species (**Table 3**). A recently reported Nep6 discovered in *N.* × *ventrata* (Wan Zakaria et al., 2019) was not detected despite that the sequence can be found in the transcriptomes of *N. rafflesiana* and hybrid (**Supplementary File 1**). The prolyl endoprotease neprosin Npr1 can be found in all three species while Npr2 was only found in *N. rafflesiana* and hybrid despite the presence of transcript in *N. ampullaria*. We identified a longer sequence (381 amino acids) of Npr2 (c68976/4/1377| m.37184) compared to the partial sequence (304 amino acids) reported by Lee et al. (2016) in *N. rafflesiana* with 81.9% sequence identity (**Supplementary Figure 4A**). An interesting protease identified in this study is the cysteine-type protease, vignain (c114505/1/1264| m.49694) uniquely found in *N. ampullaria*, which showed 76% sequence identity to partial sequence of peptidase C1 domain-containing protein (GenBank ID: GAV62544.1) present in *Cephalotus follicularis*, a carnivorous pitcher plants from a different plant order. Cathepsin propeptide inhibitor domain (I29) was detected in the sequence, which was found in the N-terminal of several peptidase C1, such as caspase that acts as a propeptide. The cysteine-type protease sequence was compared to a putative protease *Nv*CP1 from *N. ventricosa* reported by Stephenson and Hogan (Athauda et al., 2004) with 49.3% sequence identity (**Supplementary Figure 4B**). The transcripts of vignain were also found in *N. rafflesiana* and the hybrid despite not detected in their pitcher fluids.

This study also found two proteins potentially involved in lipid metabolism, namely non-specific lipid transfer protein GPI-anchored 1 (LTPG1) and lipid phosphate phosphatase 2 (LPP2). These lipid transfer proteins (LTPs) were found in all three *Nepenthes* species. The presence of LTPG1 was reported in *N. alata* and *N. mirabilis* but multiple sequence alignment found limited sequence similarity with 23.2 and 19.8%, respectively, (**Supplementary Figure 4C**). Two β-1,3-glucanases were detected in *N. rafflesiana* but absent in *N. ampullaria*, while a thaumatin-like protein (TLP) was only found in

FIGURE 3 | Venn diagram analysis of proteins identified from proteomics analysis. Selected proteins for each intersection are listed according to the font colors. Asterisks (*) indicate proteins reported previously in the pitcher fluids of other *Nepenthes* species. Refer to **Supplementary File 3** for details.

N. ampullaria with also two chitinases possibly involved in polysaccharide metabolism and/or defense response. There were more peroxidases detected in *N. rafflesiana* than *N. ampullaria* with roles in the regulation of reactive oxygen species (ROS).

In this study, we discovered several new proteins in *Nepenthes* pitcher fluids involved in secondary metabolism. Some of these proteins are cytochrome P450, isoflavone 2′-hydroxylase, isoflavone reductase homolog, and 12-oxophytodienoate reductase (jasmonic acid (JA) biosynthesis) potentially involved in secondary metabolism, anti-microbial properties, and stress response.

Putative functions for identified proteins in the newly opened pitchers of the three *Nepenthes* species are portrayed in the model of *Nepenthes* carnivory mechanism adapted from Lee et al. (2016); (**Figure 5**). This model supports that the digestive processes can readily occur upon pitcher opening through endogenous hydrolytic proteins even in the absence of prey. Four main types of metabolism identified include polysaccharide, protein, nucleic acid, and lipid digestion. The digestion of prey is likely to be initiated by glucanase and chitinase that digest the cell wall and outer parts of insects, providing nitrogen and phosphate to the plant.

TABLE 2 | List of 25 identified pitcher fluid proteins shared by *N. ampullaria*, *N. rafflesiana*, and *N.* × *hookeriana*.

Biological process	Protein	Function
Protein metabolism	Nepenthesin-1 (Nep1)	Aspartic protease
	Nepenthesin-3 (Nep3)	
	Nepenthesin-4 (Nep4)	
	Nepenthesin-5 (Nep5)	
	Neprosin-1 (Npr1)	Prolyl endoprotease
Lipid metabolism	Lipid phosphate phosphatase 2 (LPP2)	Lipid transfer
	Non-specific lipid transfer protein GPI-anchored 1 (LTPG1)	
Nucleic acid metabolism	Purple acid phosphatase (*Nr*PAP1)	Metallophosphatase
Polysaccharide metabolism	DOMON-like domain with heme-binding motif (*Nr*Dom1)	Glycoside hydrolase
Cell wall-related metabolism	Fasciclin-like arabinogalactan protein 7	Surface adhesion
	Fasciclin-like arabinogalactan protein 13	
	*NAC domain-containing protein 43 (NAC043)	Secondary cell wall biogenesis
Secondary metabolism	*12-oxophytodienoate reductase 3	Jasmonic acid biosynthesis
ROS regulation	Cation peroxidase 1 (*Nr*Prx1)	Peroxidase
	Cation peroxidase 1 (Prx1)	
	*Glutathione S-transferase (DHAR2)	ROS scavenging
Signal transduction	*Calmodulin-binding receptor-like cytoplasmic kinase 3 (CRCK3)	Serine family protein kinase
	*Serine/threonine protein phosphatase 2A 57 kDa regulatory subunit (B'KAPPA)	Phosphatase 2A regulatory subunit
Protein regulation	*F-box*/LRR-repeat protein At4g29420	Component of E3 ubiquitin ligase complex
Transporter	*Metal-nicotianamine transporter (YSL3)	Metal transporter
	*Probable sugar phosphate/phosphate translocator At3g11320	Carbohydrate transport
Gene regulation	*Polycomb protein Pcl-like	Transcription suppressor
	*RNA polymerase-associated protein C651.09c	Transcription elongation
Uncategorized	*Stress response protein NST1	
	*NADH:flavin oxidoreductase/NADH oxidase	

**Proteins not reported previously.*

DISCUSSION

A Reference Transcriptome for Protein Identification

We have generated full-length transcriptomes for species-specific protein profiling of three *Nepenthes* species during the early stage of pitcher opening to identify endogenous proteins that may contribute to the carnivory traits of *Nepenthes*. The PacBio sequencing of transcriptome from individual species provided a reference of 65,757 predicted protein sequences for protein identification. Previously, Biteau et al. (2013) examined just-opened pitchers of *N. ampullaria* and 12 other *Nepenthes* species using acetone precipitation with SDS-PAGE method followed by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) MS analysis. Meanwhile, Rottloff et al. (2016) investigated fluids from closed pitchers of several *Nepenthes* species including *N. rafflesiana* using one-dimensional SDS-PAGE followed by electrospray ionization-tandem MS (ESI-MS/MS) analysis. Lee et al. (2016) first reported proteins in *N.* × *ventrata* pitcher fluids using PIT approach through in-gel and gel-free proteomic analyses based on *N. rafflesiana* transcriptome from Illumina RNA sequencing. More recently, Wan Zakaria et al. (2019) investigated the protein replenishment in the pitcher fluids of *N.* × *ventrata* through species-specific RNA-seq analysis (Wan Zakaria et al., 2016a) and label-free quantitative proteomics (LC–MS/MS) (Wan Zakaria et al., 2018), which led to the discovery of a new nepenthesin-6 (Nep6) found to be replenished after its depletion upon pitcher opening. We also adopted the PIT approach to identify proteins in the pitcher fluids of newly opened pitchers. Due to limited protein content in *Nepenthes* pitcher fluids as previously reported (Buch et al., 2015; Lee et al., 2016), which used pooled samples of up to 1,000 pitcher fluids, we pooled nine samples to yield more concentrated proteins for nanoLC-MS/MS analysis. To our knowledge, this is the first study to analyze the transcriptome and proteome of *N.* × *hookeriana* in relation to its parent species, *N. ampullaria* and *N. rafflesiana,* to compare the pitcher fluid protein compositions related to different dietary habits.

Proteins Commonly Found in the Pitcher Fluids

A total of 220 proteins were found in pitcher fluids of the three *Nepenthes* species, including proteins known to be involved in the digestive mechanism of *Nepenthes*. Previously, several classes of proteins from *Nepenthes* pitcher fluid had been discovered, such as proteins involved in digestion, pitcher maturation, pathogenesis-related (PR), or defense (Stephenson and Hogan, 2006; Hatano and Hamada, 2008; Buch et al., 2014; Rottloff et al., 2016). Proteome analysis of *N. alata* found six proteins, including three novel PR proteins, namely TLP, β-1,3-glucanase, and β-D-xylosidase, that exhibit anti-microbial properties (Hatano and Hamada, 2012; Buch et al., 2014), which were also found in our study, except for β-D-xylosidase (**Table 3**). Lee et al. (2016) identified 36 proteins while

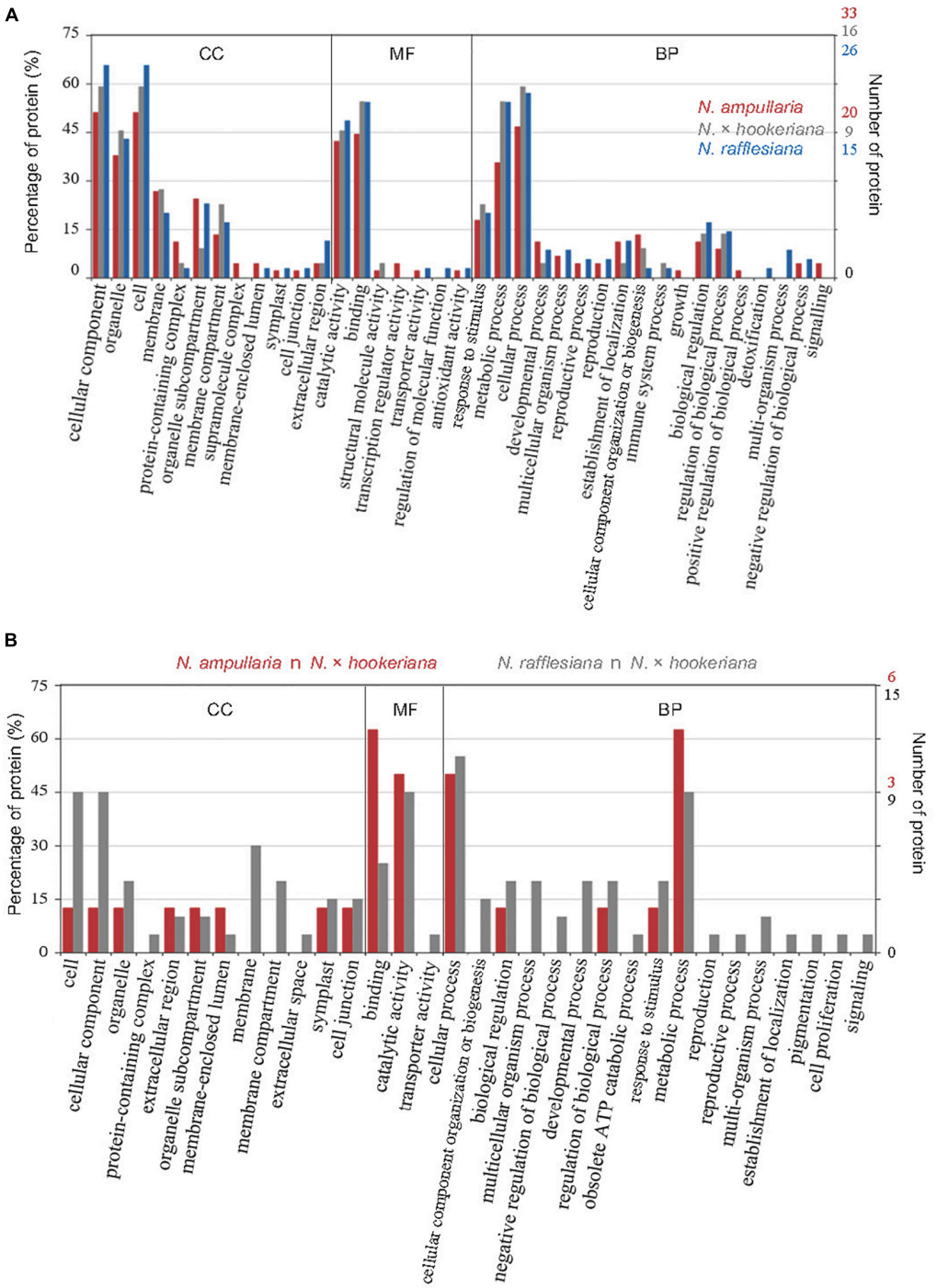


FIGURE 4 | GO distribution of identified proteins from proteomics analysis based on cellular component (CC), molecular function (MF), and biological process (BP). **(A)** Distribution of GO terms shared among all *Nepenthes* species. **(B)** Distribution of GO terms shared between parent and hybrid species, that is, *N. ampullaria* vs. *N.* × *hookeriana* (red), and *N. rafflesiana* vs. *N.* × *hookeriana* (gray).

Wan Zakaria et al. (2019) found 32 proteins from insect-fed and no insect-fed *N.* × *ventrata* pitcher fluids, respectively. Fukushima et al. (2017) focused on well-known secreted proteins (aspartic protease, class III peroxidase, GH18/class III chitinase, GH19/class IV chitinase, β-1,3-glucanase, TLP, purple acid phosphatase, PR-1-like protein, and RNase T2) from four different families of pitchers plants with three independent carnivorous origins, including *N. alata*. They showed convergent evolution of plant carnivory in the amino acid substitutions of some of the conserved digestive enzymes.

TABLE 3 | List of identified endogenous proteins important in *Nepenthes* carnivory traits from non-fed pitchers.

Function	Protein	Presence‡			Reported in other species
		A	**H**	**R**	
Protein metabolism	Nepenthesin-1, Nep1	√	√	√	*N.* × *ventrata* (Lee et al., 2016); *N. alata* (Hatano and Hamada, 2008); *N. distillatoria*, *N. gracilis* (Athauda et al., 2004); *N. sanguinea* (Rottloff et al., 2016); and *N. mirabilis* (Buch et al., 2015)
	Nepenthesin-2, Nep2	–*	–*	√	
	Nepenthesin-3, Nep3	√	√	√	
	Nepenthesin-4, Nep4	√	√	√	
	Nepenthesin-5, Nep5	√	√	√	
	Neprosin-1, Npr1	√	√	√	*N.* × *ventrata* (Lee et al., 2016; Wan Zakaria et al., 2019)
	Neprosin-2, Npr2	–*	√	√	
	Vignain	√	–*	–*	*N. ventricosa* (Stephenson and Hogan, 2006)
Nucleic acid metabolism	S-like ribonuclease, *Nr*RNaseS	–	√	√	*N.* × *ventrata* (Lee et al., 2016; Wan Zakaria et al., 2019); *N. bicalcarata* (Stephenson and Hogan, 2006); and *N. ventricosa* (Nishimura et al., 2014)
Lipid metabolism	Non-specific lipid transfer protein GPI-anchored 1, LTPG1	√	√	√	*N. alata* (Hatano and Hamada, 2008); *N. mirabilis* (Rottloff et al., 2016)
	Lipid phosphate phosphatase 2, LPP2	√	√	√	
Metallo-phosphatase	Purple acid phosphatase, PAP1	√	√	√	*N.* × *ventrata* (Lee et al., 2016; Wan Zakaria et al., 2019)
Polysaccharide	β-1,3-glucanase, *Nr*Glu1	–	√	√	*N.* × *ventrata* (Lee et al., 2016)
metabolism/	β-1,3-glucanase, *Na*BGlu1	–	–	√	*N. alata* (Hatano and Hamada, 2008; Rottloff et al., 2016)
Defense response	Chitinase class III, Chit3	√	√	–	*N.* × *ventrata* (Lee et al., 2016; Wan Zakaria et al., 2019); *N. alata* (Hatano and Hamada, 2008; Rottloff et al., 2011, 2016); *N. khasiana* (Eilenberg et al., 2006); *N. singalana*, *N. gracilis*, *N. mirabilis*, and *N. rafflesiana* (Rottloff et al., 2011)
	Chitinase class IV, Chit1	√	–	√	
	Thaumatin-like protein, TLP1	√	–*	–*	*N. alata* (Hatano and Hamada, 2008; Rottloff et al., 2016); *N. albomarginata, N. mirabilis, N. sanguinea* (Rottloff et al., 2016); *N.* × *ventrata* (Lee et al., 2016; Wan Zakaria et al., 2019)
ROS regulation	Cationic peroxidase 1, *Nr*Prx1	√	√	√	*N.* × *ventrata* (Lee et al., 2016; Wan Zakaria et al., 2019); *N. alata* (Hatano and Hamada, 2008); and *N. bicalcarata* (Rottloff et al., 2016)
	Cationic peroxidase 1, Prx1	–*	√	√	
	Cationic peroxidase 3, *Nr*Prx3	–	–	√	
Secondary metabolism	Cytochrome P450	√	√	–*	
	Isoflavone 2′-hydroxylase	√	–*	√	
	Isoflavone reductase homolog	–*	√	√	
	12-oxophytodienoate reductase 3, OPR3	√	√	√	

‡*A, N. ampullaria; H, Hybrid; R, N. rafflesiana;* √*, Presence; –, Absence.*
indicates the presence of transcript in the pitcher tissues but no protein was detected in the pitcher fluids. Refer* ***Supplementary File 3* *for details on the transcript ID.*

Overall, the number of proteins identified from previous studies were much lower than our findings of 94 to 125 identified proteins for individual species with the highest number from *N. rafflesiana* and 25 proteins shared in all three species. This might be due to the differences in datasets and analysis pipelines used for protein identification. Our findings from the functional annotation of *Nepenthes* transcriptomes (**Figure 2**) and identified proteins from proteomic analyses (**Figure 3**) revealed that *N.* × *hookeriana* is more similar to *N. rafflesiana* as compared to *N. ampullaria*, which is consistent to findings from the genetic analysis (Yulita and Mansur, 2012) that suggest a greater genetic similarity between the two species than *N. ampullaria*. Since our samples were obtained originally from natural habitat instead of controlled breeding, we cannot exclude the possibility of genetic similarity derived from hybrid backcrossing with *N. rafflesiana*. Nevertheless, a higher number of common proteins between *N.* × *hookeriana* and *N. rafflesiana* observed in protein clustering of transcriptomics and proteomics analyses suggested that a similar set of proteins were secreted during the early stage of pitcher opening in our samples, and both species have similar enzymes for prey digestion.

The 25 proteins found in all three *Nepenthes* species reflect their importance for early processes in newly opened *Nepenthes* pitchers. GO annotations for transcriptome and proteome discovered biological processes and molecular functions of the proteins identified in the pitcher fluids with the lowest significant

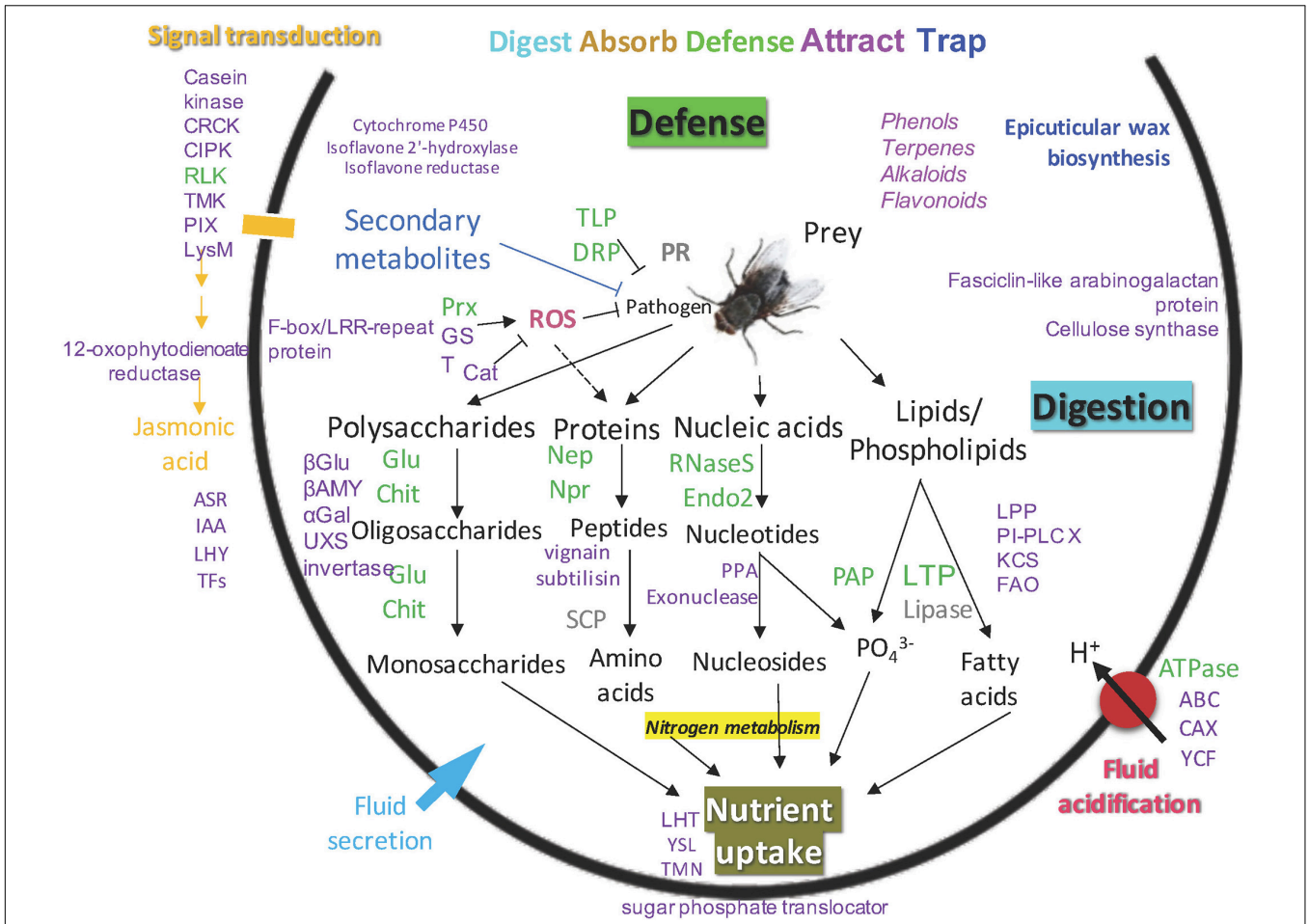


FIGURE 5 | An updated model of carnivory mechanism in *Nepenthes* pitchers, green: common proteins found in pitcher fluids, gray: common proteins not detected in this study, purple: new proteins found in this study, blue: new biological process identified in this study.

difference between *N. rafflesiana* and the hybrid. Some of the identified proteins in this study involved four main types of metabolisms, which are the metabolisms of proteins (10), lipids (7), nucleic acids (5), and polysaccharides (5). A high number of proteins involved in the catalytic activity were found with enzymatic roles for digestion, such as the hydrolase activities nepenthesins, purple acid phosphatase, and lipid phosphate phosphatase 2, which were common in all three species (**Table 2**).

Many of these proteins are found in previous studies (Athauda et al., 2004; Hatano and Hamada, 2012; Buch et al., 2014; Rottloff et al., 2016; Fukushima et al., 2017), which suggested that these endogenous proteins were secreted during pitcher opening in preparation for defense and prey digestion. These proteins include the conserved aspartic proteases, nepenthesins (Nep1-Nep5) that function to digest prey, mainly insects and plant debris by hydrolyzing peptides (Athauda et al., 2004; Lee et al., 2016). Most of these nepenthesins were found in other *Nepenthes* species such as *N.* × *ventrata*, *N. alata*, *N. distillatoria*, and *N. gracilis* (Athauda et al., 2004; Hatano and Hamada, 2008; Nishimura et al., 2014; Rottloff et al., 2016), except the Nep2, which was only found in *N. rafflesiana* with the same similarity to the *Nr*Nep2 sequence found by Lee et al. (2016) in *N.* × *ventrata* pitcher fluids. Nep1 contains carbohydrate moieties and glycosylation sites important for protein stability to prevent denaturation. However, it was not the case for Nep2, as observed in *N. gracilis* (Bariola and Green, 1997), which could explain the probable instability of Nep2 in the pitcher fluids, hence not found in *N. ampullaria* and *N.* × *hookeriana* despite the presence of transcripts (**Table 3**). Another recently reported nepenthesin in *N.* × *ventrata*, Nep6 (Wan Zakaria et al., 2019), was not detected in any of the species but the sequence was found in the reference transcriptome (c202852/1/1129| m.78569) attributed by *N. rafflesiana*. It is noteworthy that nine out of 10 proteases reported in this study were found in *N. rafflesiana* compared to six in the hybrid and *N. ampullaria*. Furthermore, only Nep4 and CLPX were identified by the "specific" trypsin digestion setting, while majority of other proteases were identified by "semispecific" and "unspecific" digestion settings, suggesting protein self-hydrolysis in the pitcher fluids during protein extraction (**Supplementary File 3**).

The presence of pathogenesis or defense-related proteins such as TLP, β-1,3-glucanase, and class III and class IV chitinases

were not consistent, which suggests differential protein secretion in the three *Nepenthes* species. Both chitinases were secreted in the pitcher fluids of *N. ampullaria* but only one in each of the other species (**Table 3**). These proteins were found to be prey-induced in *N.* × *ventrata* and *N. alata* (Lee et al., 2016; Rottloff et al., 2016) with proposed contribution to the anti-microbial environment in the pitcher fluids apart from digestion (Hatano and Hamada, 2012; Buch et al., 2013). TLP putatively functions to fight pathogens from the ingested prey, while the glycoside hydrolases (GHs), β-1,3-glucanases, and chitinases function in hydrolyzing polysaccharides, such as the cell walls of pathogens, insects, and leaves (Minic and Jouanin, 2006; Buch et al., 2013).

On the other hand, a nuclease, the S-like ribonuclease (RNaseS) was identified in *N.* × *hookeriana* and *N. rafflesiana*, with high sequence similarity to a similar protein in *N.* × *ventrata*, *N. bicalcarata,* and *N. ventricosa* (Stephenson and Hogan, 2006; Nishimura et al., 2014). In non-carnivorous plants, the protein is useful for self-defense against pathogen attacks from the prey (Bariola and Green, 1997; Sangaev et al., 2011). The expression of RNaseS in carnivorous plants showed tissue-specific constitutive expression in *Drosera* and *Cephalotus* and is prey-induced in *Dionaea* (Okabe et al., 2005; Nishimura et al., 2013). The presence of RNaseS in *N. rafflesiana* and the hybrid could indicate the conservation of ribonuclease activity for anticipated insect prey digestion.

New Proteins Found in the Pitcher Fluids

In this study, we discovered 21 new pitcher fluid proteins involved in protein regulation (**Supplementary File 3**). These previously unreported proteins mainly function in protein ubiquitination, such as BTB/POZ domain-containing protein (POB1), E3 ubiquitin-protein ligases, F-box/kelch-repeat protein, and F-box/LRR-repeat protein. This suggests that the turnover of secreted proteins is actively regulated in the pitcher fluids. However, no proteasomal protein was found. Therefore, the half-life of secreted proteins in the pitcher fluids poses an interesting biological question to be addressed in the future.

We also discovered 10 proteins related to signal transduction, which may play roles in regulating gene expression with the 28 detected transcription factors (**Figures 4**, **5** and **Supplementary File 3**). It is intriguing to find these proteins in the pitcher fluids, which are expected to be intracellular. Likewise, for the 15 proteins functioning in protein translation or synthesis, such as eukaryotic translation initiation factor 1A (EIF1A), 30S and 60S ribosomal proteins, arginyl-tRNA—protein transferase 1, and valine—tRNA ligase. Similarly, there were 10 proteins related to intracellular trafficking or cytoskeleton, such as actin, armadillo repeat-containing kinesin-like protein, and katanin. Some of these unexpected proteins are reported in the previous studies of *N.* × *ventrata* pitcher fluids (Lee et al., 2016; Wan Zakaria et al., 2019).

The carnivory mechanism of carnivorous plants has been proposed to evolve from the plant defense mechanism through JA signaling (Yilamujiang et al., 2016; Pavlovič and Mithöfer, 2019). It is therefore interesting to discover a 12-oxophytodienoate reductase (OPR3) involved in the biosynthesis of JA and lipid metabolism (Chini et al., 2018) in the pitcher fluids of all three species. It had been reported that JA may induce the proteolytic activity of nepenthesin in *Nepenthes* (Buch et al., 2015). Other phytohormone-related proteins include auxin response proteins (IAA9 and 27), abscisic stress-ripening protein 1 (ASR1), LHY, and WRKY transcription factors, which could play a role in stress response.

Apart from OPR3, we detected several other proteins involved in secondary metabolism, which include the cytochrome P450, isoflavone 2′-hydroxylase, and isoflavone reductase homolog. The cytochrome P450 functions to convert carlatone to carlartonic acid and is involved in flavonoid pathway (Abe et al., 2014); isoflavone 2′-hydroxylase functions in the biosynthesis of isoflavonoid-derived antimicrobial compounds (Akashi et al., 1998); and isoflavone reductase functions in the biosynthetic pathway of isoflavonoid phytoalexin (Cheng et al., 2015). Previous studies identified flavonoids and naphthoquinones in *N. khasiana* (Eilenberg et al., 2006), naphthoquinones (plumbagin and 7-methyl-juglone) in the opened pitcher fluids of *N. ventricosa* (Buch et al., 2013), while dihydronaphthoquinone glucosides rossoliside, plumbaside A, and plumbagin were reported in *N. insignis* (Rischer et al., 2002). These metabolites contain anti-microbial properties that prevent microbial competition for nutrient absorption. Efforts in the profiling of secondary metabolites from *Nepenthes* pitchers and their bioactivity are on-going (Rosli et al., 2017, 2018; Dávila-Lara et al., 2020). Meanwhile, genes involved in the biosynthesis of secondary metabolites such as phenylpropanoids, sesquiterpenoids, and triterpenoids in *N. ampullaria* were found to be influenced by endogenous protein depletion (Goh et al., 2020). Proteins involved in secondary metabolism were also reported to be important for response against environmental stress such as pathogen attack that led to the synthesis of secondary metabolites from different pathways (Chini et al., 2018; Goh et al., 2020). Further studies are needed to ascertain the roles of these proteins in secondary metabolism and stress response. Multi-omics integration will help elucidate the genes or enzymes involved in the biosynthesis pathways of secondary metabolites important for pitcher physiology.

Despite the discovery of many new proteins, most of them are expected to be functional intracellularly, such as OPR3 in the peroxisomes, transcription factors in the nucleus, and the membrane-localized transporters. Since our experimental design is based on species-specific transcriptomes using newly opened pitchers without prey, it is unlikely that these proteins are contaminants from the microbes or insects. However, we cannot exclude the possibility that these proteins could be attributed by microbial symbionts of the pitcher plants that could be present even in closed pitchers, although the fluids are unsuitable for microbial growth (Buch et al., 2013). The significance of these seemingly intracellular proteins in the pitcher fluids warrants further studies. It is noteworthy that the discovery of extracellular OPR3 corroborates the presence of jasmonyl-isoleucine (JA-Ile) in the digestive fluid (Yilamujiang et al., 2016). This suggests the possibility of the biosynthesis of phytohormones or secondary metabolites extracellularly.

On the other hand, there is no strong evidence in this study to suggest an adaptive evolution of *N. ampullaria* with novel enzymes for digesting leaf litter, which has been hypothesized to depend on infauna of the pitcher fluids (Moran et al., 2003; Moran and Clarke, 2010). This is consistent with the findings that pitcher fluids of *N. ampullaria* are heavily populated with aquatic organisms (Cresswell, 1998), perhaps due to the less acidic pitcher fluids compared to other *Nepenthes* species at a trade-off of hydrolytic enzymes functioning at suboptimal pH (Saganová et al., 2018). Nonetheless, some of the unique endogenous proteins discovered in *N. ampullaria* could potentially contribute to nutrient sequestration, for example, a cysteine-type peptidase vignain, an alpha-galactosidase, a beta-glucosidase, a cellulose synthase A (CESA), and a catalase (**Figure 3**). Apart from these unique enzymes, the finding that both prey-induced chitinases (Chit1 and Chit3) were found in the newly opened pitchers suggests differential secretion of proteins in *N. ampullaria* could contribute to its success in being an omnivore to derive nutrients from both insects and leaf litter. However, this remains speculative without functional validation through genetic transformation or transfection, which unfortunately is still unavailable.

CONCLUSION

The comparison of protein content in pitcher fluids of three *Nepenthes* species through transcriptomic and proteomic analyses revealed distinct profiles of secreted proteins, especially hydrolytic enzymes and defense-related proteins. Despite no evidence of novel enzymes for leaf litter digestion in *N. ampullaria*, this study provides information on the molecular compositions of individual *Nepenthes* species with differential secretion of endogenous proteins apart from the distinct morphological traits between the parent species and hybrid that reflect inter-species diversity. Furthermore, many interesting biological questions that are raised on the functions of new proteins discovered in this study manifest wonders on the molecular physiology of secreted proteins in the pitcher fluids to be elucidated in future studies.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MZ, WA, and H-HG designed the experiments. MZ performed the experiments and analyzed the data. H-HG performed funding project administration and supervision. NA and TT assisted in data organization and discussion. MZ, NA, TT, WA, and H-HG wrote and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the Universiti Kebangsaan Malaysia (DIP-2014-008). The research group is currently supported by the Malaysian Ministry of High Education (FRGS/1/2019/STG05/UKM/02/10) and UKM research university grant (DIP-2020-005).

ACKNOWLEDGMENTS

We thank Prof. Dr. Jumaat Haji Adam for contributing pitcher samples with access to the *Nepenthes* experimental terrace. We express our gratitude to the two reviewers and editor for their constructive comments in improving this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.625507/full#supplementary-material

Supplementary Figure 1 | Overview of the transcriptomics and proteomics studies of *Nepenthes* pitcher fluids.

Supplementary Figure 2 | BLASTN analysis of consensus isoform sequences. **(A)** BLASTN searches of individual transcriptomes against the reference transcriptome. BLASTN analysis of hybrid transcriptome against the parent transcriptomes showing **(B)** the percentage of hybrid sequences that found hits and **(C)** the percentage of parent sequences that found hits with the hybrid in reciprocal searches.

Supplementary Figure 3 | SDS-PAGE and silver staining analysis of pitcher fluid protein extraction and processing. **(A–C)** SDS-PAGE gel results for 20 μL aliquots after filtration (left) and after concentration (right). **(A)** *N. ampullaria* [amp]. **(B)** *N. rafflesiana* [raff]. **(C)** *N.* × *hookeriana* [hook]. Individual lanes represent the nine biological replicates. **(D)** SDS-PAGE gel results of pooled samples after solid phase extraction (left) and peptide digestion (right).

Supplementary Figure 4 | Multiple sequence alignment analysis of selected sequences found in proteomic analysis. **(A)** Pairwise sequence alignment of c68976/4/1377| m.37184 transcript (top) and truncated Neprosin-2 transcript (bottom) from *N. rafflesiana* as reported by Lee et al. (2016). **(B)** Pairwise sequence alignment of c114505/1/1264 m.49694 (top) and *Nv*CP1 (bottom), with a sequence identity of 49.34%. **(c)** Sequence alignment of c171049/1/1434 m.73671 and LTP sequences from *N. alata* and *N. mirabilis*.

Supplementary Table 1 | Statistics of transcriptome obtained from PacBio sequencing using the RS_IsoSeq protocol. The analysis had been performed for independently individual species. Reference dataset is generated from the combined analysis of all three transcriptomes.

Supplementary Table 2 | WEGO analysis of the gene ontology (GO) annotation.

Supplementary Table 3 | Summary of KO analysis.

Supplementary Table 4 | OrthoVenn cluster analysis of predicted protein sequences.

Supplementary File 1 | Functional annotation of transcriptomes.

Supplementary File 2 | GO and KO analysis.

Supplementary File 3 | Proteomics analysis.

REFERENCES

Abdul-Rahman, A., Suleman, N. I., Zakaria, W. A., Goh, H. H., Noor, N. M., and Aizat, W. M. (2017). RNA extractions of mangosteen (*Garcinia mangostana* L.) pericarps for sequencing. *Sains Malaysiana* 46, 1231–1240. doi: 10.17576/jsm-2017-4608-08

Abe, S., Sado, A., Tanaka, K., Kisugi, T., Asami, K., Ota, S., et al. (2014). Carlactone is converted to carlactonoic acid by MAX1 in *Arabidopsis* and its methyl ester can directly interact with AtD14 in vitro. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18084–18089. doi: 10.1073/pnas.1410801111

Akashi, T., Aoki, T., and Ayabe, S.-I. (1998). CYP81E1, a cytochrome P450 cDNA of licorice (*Glycyrrhiza echinata*L.), encodes isoflavone 2'-hydroxylase. *Biochem. Biophys. Res. Commun.* 251, 67–70. doi: 10.1006/bbrc.1998.9414

Athauda, S. B., Matsumoto, K., Rajapakshe, S., Kuribayashi, M., Kojima, M., Kubomura-Yoshida, N., et al. (2004). Enzymic and structural characterization of nepenthesin, a unique member of a novel subfamily of aspartic proteinases. *Biochem. J.* 381, 295–306. doi: 10.1042/bj20031575

Bariola, P. A., and Green, P. J. (1997). "Plant ribonucleases," in *Ribonucleases*, eds G. D'Alessio and J. F. Riordan (Amsterdam: Elsevier), 163–190. doi: 10.1016/b978-012588945-2/50006-6

Biteau, F., Nisse, E., Miguel, S., Hannewald, P., Bazile, V., Gaume, L., et al. (2013). A simple SDS-PAGE protein pattern from pitcher secretions as a new tool to distinguish *Nepenthes* species (Nepenthaceae). *Am. J. Bot.* 100, 2478–2484. doi: 10.3732/ajb.1300145

Bryant, D. M., Johnson, K., Ditommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., et al. (2017). A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* 18, 762–776. doi: 10.1016/j.celrep.2016.12.063

Buch, F., Kaman, W. E., Bikker, F. J., Yilamujiang, A., and Mithöfer, A. (2015). Nepenthesin protease activity indicates digestive fluid dynamics in carnivorous *Nepenthes* plants. *PLoS One* 10:e0118853. doi: 10.1371/journal.pone.0118853

Buch, F., Pauchet, Y., Rott, M., and Mithöfer, A. (2014). Characterization and heterologous expression of a PR-1 protein from traps of the carnivorous plant *Nepenthes mirabilis*. *Phytochemistry* 100, 43–50. doi: 10.1016/j.phytochem.2014.01.014

Buch, F., Rott, M., Rottloff, S., Paetz, C., Hilke, I., Raessler, M., et al. (2013). Secreted pitfall-trap fluid of carnivorous *Nepenthes* plants is unsuitable for microbial growth. *Ann. Bot.* 111, 375–383. doi: 10.1093/aob/mcs287

Cheng, Q., Li, N., Dong, L., Zhang, D., Fan, S., Jiang, L., et al. (2015). Overexpression of soybean isoflavone reductase (GmIFR) enhances resistance to *Phytophthora sojae* in soybean. *Front. Plant Sci.* 6:1024. doi: 10.3389/fpls.2015.01024

Chini, A., Monte, I., Zamarreño, A. M., Hamberg, M., Lassueur, S., Reymond, P., et al. (2018). An OPR3-independent pathway uses 4, 5-didehydrojasmonate for jasmonate synthesis. *Nat. Chem. Biol.* 14, 171–178. doi: 10.1038/nchembio.2540

Clarke, C. (2001). *Nepenthes of Sumatra and Peninsular Malaysia*. Kota Kinabalu: Natural History Publications (Borneo).

Clarke, C., and Moran, J. A. (2016). Climate, soils and vicariance – their roles in shaping the diversity and distribution of *Nepenthes* in Southeast Asia. *Plant Soil* 403, 37–51. doi: 10.1007/s11104-015-2696-x

Clarke, C., and Wong, K. M. (1997). *Nepenthes of Borneo*. Kota Kinabalu: Natural History Publications (Borneo).

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* 13, 2513–2526. doi: 10.1074/mcp.m113.031591

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–1805. doi: 10.1021/pr101065j

Cresswell, J. E. (1998). Morphological correlates of necromass accumulation in the traps of an eastern tropical pitcher plant, *Nepenthes ampullaria* Jack, and observations on the pitcher infauna and its reconstitution following experimental removal. *Oecologia* 113, 383–390. doi: 10.1007/s004420050390

Dávila-Lara, A., Rodríguez-López, C. E., O'connor, S. E., and Mithöfer, A. (2020). Metabolomics analysis reveals tissue-specific metabolite compositions in leaf blade and traps of carnivorous *Nepenthes* plants. *Int. J. Mol. Sci.* 21:4376. doi: 10.3390/ijms21124376

Eilenberg, H., Pnini-Cohen, S., Schuster, S., Movtchan, A., and Zilberstein, A. (2006). Isolation and characterization of chitinase genes from pitchers of the carnivorous plant *Nepenthes khasiana*. *J. Exp. Bot.* 57, 2775–2784. doi: 10.1093/jxb/erl048

Fukushima, K., Fang, X., Alvarez-Ponce, D., Cai, H., Carretero-Paulet, L., Chen, C., et al. (2017). Genome of the pitcher plant *Cephalotus* reveals genetic changes associated with carnivory. *Nat. Ecol. Evol.* 1:59.

Gilbert, K. J., Bittleston, L. S., Tong, W., and Pierce, N. E. (2020). Tropical pitcher plants (*Nepenthes*) act as ecological filters by altering properties of their fluid microenvironments. *Sci. Rep.* 10:4431.

Goh, H.-H., Baharin, A., Salleh, F. I. M., Ravee, R., Zakaria, W. N. A. W., and Noor, N. M. (2020). Transcriptome-wide shift from photosynthesis and energy metabolism upon endogenous fluid protein depletion in young *Nepenthes ampullaria* pitchers. *Sci. Rep.* 10:6575.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084

Hatano, N., and Hamada, T. (2008). Proteome analysis of pitcher fluid of the carnivorous plant *Nepenthes alata*. *J. Proteome Res.* 7, 809–816. doi: 10.1021/pr700566d

Hatano, N., and Hamada, T. (2012). Proteomic analysis of secreted protein induced by a component of prey in pitcher fluid of the carnivorous plant *Nepenthes alata*. *J. Proteomics* 75, 4844–4852. doi: 10.1016/j.jprot.2012.05.048

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314.

Lee, L., Zhang, Y., Ozar, B., Sensen, C. W., and Schriemer, D. C. (2016). Carnivorous nutrition in pitcher plants (*Nepenthes* spp.) via an unusual complement of endogenous enzymes. *J. Proteome Res.* 15, 3108–3117. doi: 10.1021/acs.jproteome.6b00224

Minic, Z., and Jouanin, L. (2006). Plant glycoside hydrolases involved in cell wall polysaccharide degradation. *Plant Physiol. Biochem.* 44, 435–449. doi: 10.1016/j.plaphy.2006.08.001

Moran, J. A., and Clarke, C. M. (2010). The carnivorous syndrome in *Nepenthes* pitcher plants: current state of knowledge and potential future directions. *Plant Signal. Behav.* 5, 644–648. doi: 10.4161/psb.5.6.11238

Moran, J. A., Clarke, C. M., and Hawkins, B. J. (2003). From carnivore to detritivore? Isotopic evidence for leaf litter utilization by the tropical pitcher plant *Nepenthes ampullaria*. *Int. J. Plant Sci.* 164, 635–639. doi: 10.1086/375422

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185.

Murphy, B., Forest, F., Barraclough, T., Rosindell, J., Bellot, S., Cowan, R., et al. (2020). A phylogenomic analysis of *Nepenthes* (Nepenthaceae). *Mol. Phylogenet. Evol.* 144:106668. doi: 10.1016/j.ympev.2019.106668

Nishimura, E., Jumyo, S., Arai, N., Kanna, K., Kume, M., Nishikawa, J.-I., et al. (2014). Structural and functional characteristics of S-like ribonucleases from carnivorous plants. *Planta* 240, 147–159. doi: 10.1007/s00425-014-2072-8

Nishimura, E., Kawahara, M., Kodaira, R., Kume, M., Arai, N., Nishikawa, J. I., et al. (2013). S-like ribonuclease gene expression in carnivorous plants. *Planta* 238, 955–967. doi: 10.1007/s00425-013-1945-6

Okabe, T., Iwakiri, Y., Mori, H., Ogawa, T., and Ohyama, T. (2005). An S-like ribonuclease gene is used to generate a trap-leaf enzyme in the carnivorous plant *Drosera adelae*. *FEBS Lett.* 579, 5729–5733. doi: 10.1016/j.febslet.2005.09.043

Oliveros, J. C. (2007-2015). *Venny. An Interactive Tool for Comparing Lists with Venn's Diagrams*. Available online at: https://bioinfogp.cnb.csic.es/tools/venny/index.html (accessed September 20, 2020).

Papadopoulos, J. S., and Agarwala, R. (2007). COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23, 1073–1079. doi: 10.1093/bioinformatics/btm076

Pavlovič, A. (2012). Adaptive radiation with regard to nutrient sequestration strategies in the carnivorous plants of the genus *Nepenthes*. *Plant Signal. Behav.* 7, 295–297. doi: 10.4161/psb.18842

Pavlovič, A., and Mithöfer, A. (2019). Jasmonate signalling in carnivorous plants: copycat of plant defence mechanisms. *J. Exp. Bot.* 70, 3379–3389. doi: 10.1093/jxb/erz188

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* 47, D442–D450.

Ravee, R., Salleh, F. M., and Goh, H. H. (2018). Discovery of digestive enzymes in carnivorous plants with focus on proteases. *PeerJ* 2018:e4914. doi: 10.7717/peerj.4914

Rischer, H., Hamm, A., and Bringmann, G. (2002). *Nepenthes insignis* uses a C2-portion of the carbon skeleton of L-alanine acquired via its carnivorous organs, to build up the allelochemical plumbagin. *Phytochemistry* 59, 603–609. doi: 10.1016/s0031-9422(02)00003-1

Rosli, M. A., Azizan, K. A., Baharum, S. N., and Goh, H. H. (2017). Mass spectrometry data of metabolomics analysis of *Nepenthes* pitchers. *Data Brief* 14, 295–297. doi: 10.1016/j.dib.2017.07.068

Rosli, M. A. F., Azizan, K. A., and Goh, H. H. (2018). Antioxidant activity of pitcher extracts from three *Nepenthes* species. *Sains Malaysiana* 47, 3069–3075. doi: 10.17576/jsm-2018-4712-17

Rottloff, S., Miguel, S., Biteau, F., Nisse, E., Hammann, P., Kuhn, L., et al. (2016). Proteome analysis of digestive fluids in *Nepenthes* pitchers. *Ann. Bot.* 117, 479–495.

Rottloff, S., Stieber, R., Maischak, H., Turini, F. G., Heubl, G., and Mithöfer, A. (2011). Functional characterization of a class III acid endochitinase from the traps of the carnivorous pitcher plant genus, *Nepenthes*. *J. Exp. Bot.* 62, 4639–4647. doi: 10.1093/jxb/err173

Saganová, M., Bokor, B., Stolárik, T., and Pavlovič, A. (2018). Regulation of enzyme activities in carnivorous pitcher plants of the genus *Nepenthes*. *Planta* 248, 451–464. doi: 10.1007/s00425-018-2917-7

Sangaev, S., Kochetov, A., Ibragimova, S., Levenko, B., and Shumny, V. (2011). Physiological role of extracellular ribonucleases of higher plants. *Russ. J. Genet. Appl. Res.* 1, 44–50. doi: 10.1134/s2079059711010060

Sievers, F., and Higgins, D. G. (2018). Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* 27, 135–145. doi: 10.1002/pro.3290

Stephenson, P., and Hogan, J. (2006). Cloning and characterization of a ribonuclease, a cysteine proteinase, and an aspartic proteinase from pitchers of the carnivorous plant *Nepenthes ventricosa* Blanco. *Int. J. Plant Sci.* 167, 239–248. doi: 10.1086/499284

Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* 11, 2301–2319. doi: 10.1038/nprot.2016.136

Wan Zakaria, W. N. A., Aizat, W. M., Goh, H. H., and Mohd Noor, N. (2019). Protein replenishment in pitcher fluids of *Nepenthes*× *ventrata* revealed by quantitative proteomics (SWATH-MS) informed by transcriptomics. *J. Plant Res.* 132, 681–694. doi: 10.1007/s10265-019-01130-w

Wan Zakaria, W. N. A., Aizat, W. M., Goh, H. H., and Noor, N. M. (2018). Proteomic analysis of pitcher fluid from *Nepenthes* × *ventrata*. *Data Brief* 17, 517–519. doi: 10.1016/j.dib.2018.01.037

Wan Zakaria, W. N. A., Loke, K. K., Goh, H. H., and Mohd Noor, N. (2016a). RNA-seq analysis for plant carnivory gene discovery in *Nepenthes* × *ventrata*. *Genomics Data* 7, 18–19. doi: 10.1016/j.gdata.2015.11.007

Wan Zakaria, W. N. A., Loke, K. K., Zulkapli, M. M., Mohd Salleh, F. I., Goh, H. H., and Mohd Noor, N. (2016b). RNA-seq analysis of *Nepenthes ampullaria*. *Front. Plant Sci.* 6:1229. doi: 10.3389/fpls.2015.01229

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., et al. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 47, W52–W58.

Yilamujiang, A., Reichelt, M., and Mithöfer, A. (2016). Slow food: insect prey and chitin induce phytohormone accumulation and gene expression in carnivorous *Nepenthes* plants. *Ann. Bot.* 118, 369–375. doi: 10.1093/aob/mcw110

Yulita, K. S., and Mansur, M. (2012). The occurrence of hybrid in *Nepenthes hookeriana* Lindl. from Central Kalimantan can be detected by RAPD and ISSR markers. *HAYATI J. Biosci.* 19, 18–24. doi: 10.4308/hjb.19.1.18

Zulkapli, M. M., Rosli, M. A. F., Salleh, F. I. M., Mohd Noor, N., Aizat, W. M., and Goh, H. H. (2017). Iso-Seq analysis of *Nepenthes ampullaria*, *Nepenthes rafflesiana* and *Nepenthes*× *hookeriana* for hybridisation study in pitcher plants. *Genomics Data* 12, 130–131. doi: 10.1016/j.gdata.2017.05.003

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 21 April 2021
doi: 10.3389/fpls.2021.655004

UPLC-TOF-MS/MS-Based Metabolomics Analysis Reveals Species-Specific Metabolite Compositions in Pitchers of *Nepenthes ampullaria*, *Nepenthes rafflesiana*, and Their Hybrid *Nepenthes* × *hookeriana*

*Muhammad Aqil Fitri Rosli, Ahmed Mediani, Kamalrul Azlan Azizan, Syarul Nataqain Baharum and Hoe-Han Goh**

Institute of Systems Biology, Universiti Kebangsaan Malaysia, Bangi, Malaysia

OPEN ACCESS

Edited by:
Jeremy D. Rentsch, Francis Marion University, United States

Reviewed by:
Axel Mithöfer, Max Planck Institute for Chemical Ecology, Germany
Weng Ngai Lam, National University of Singapore, Singapore

***Correspondence:**
Hoe-Han Goh
gohhh@ukm.edu.my

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

Received: *18 January 2021*
Accepted: *15 March 2021*
Published: *21 April 2021*

Citation:
Rosli MAF, Mediani A, Azizan KA, Baharum SN and Goh H-H (2021) UPLC-TOF-MS/MS-Based Metabolomics Analysis Reveals Species-Specific Metabolite Compositions in Pitchers of Nepenthes ampullaria, Nepenthes rafflesiana, and Their Hybrid Nepenthes × hookeriana. Front. Plant Sci. 12:655004. doi: 10.3389/fpls.2021.655004

Hybridization is key to the evolution and diversity of plants in nature. Nepenthaceae comprises a family of diverse tropical carnivorous pitcher plant species with extensive hybridization. However, there is no study to date on the metabolite expression of hybrids in this family. We performed a non-targeted metabolomics analysis of the pitchers of two *Nepenthes* species with different dietary habits, namely, the semi-detritivorous *N. ampullaria* and carnivorous *N. rafflesiana* with their hybrid (*N.* × *hookeriana*) for a comparative study. The whole-pitcher samples were extracted in methanol:chloroform:water (3:1:1) via sonication-assisted extraction and analyzed using ultra-performance liquid chromatography time-of-flight mass spectrometry (UPLC-TOF-MS) followed by data analysis to profile chemical compositions. A total of 1,441 metabolite features were profiled from the three species in which 43.3% of features in the hybrid samples were not found in either of its parents. The partial least squares discriminant analysis (PLS-DA) found 324 metabolite features with variable in projection (VIP) values greater than one in which 55 features were statistically significant. This showed that the hybrid is closer to *N. rafflesiana*, which is consistent to the previous study on gene and protein expressions. A total of 105 metabolites were putatively identified with manual searches using public metabolite databases. Phenols were detected to be the most abundant secondary metabolites due to a high flavonoid content, especially in *N. rafflesiana*. The most abundant feature 476.3s:449.102 was found to be the most significant VIP for distinguishing between the three species as a chemical marker. This is the first study comparing metabolites in the carnivory organs of different *Nepenthes* species with comprehensive profiling and putative identification.

The differential metabolite compositions in the pitchers of different species might have ecological implications with the hybrid showing intermediate phenotype between the parents as well as manifesting unique metabolites. However, there is no clear evidence of metabolites related to the differences in dietary habits between the hybrid and the two parent species.

Keywords: carnivorous plant, flavonoid, hybrid, LC-MS, metabolomics, *Nepenthes*, pitcher

INTRODUCTION

Nepenthes L. is the sole genus of paleotropical carnivorous pitcher plants from the monotypic Nepenthaceae family of Caryophyllales order that is highly distributed around the warm and humid equatorial countries, such as Indonesia, Philippines, and Malaysia (Adam, 1997). *Nepenthes* pitcher plants thrive on poor soils and acquire limiting nutrients mainly from insect prey via the unique fluid-filled pitcher organ developed at the leaf tip connected by a tendril, which allows prey trapping, digestion, and absorption (Clarke and Robinson, 2018). Fascinating pitcher shapes and sizes have been observed in this diverse family of carnivorous plants with over 150 extant species (Murphy et al., 2020), which are linked to the various ecological adaptations related to nutrient acquisition or dietary habits. For example, *Nepenthes ampullaria* Jack have evolved a semi-detritivorous habit with a vestigial pitcher lid bent away from the pitcher opening surrounded by a relatively flat peristome and growing on the forest floor to ease the trapping of fallen leaf litter, thus less dependent on catching prey (Moran et al., 2003; Pavlovič et al., 2011). In comparison, a typical carnivorous *N. rafflesiana* Jack produces larger pitchers with attractive coloration and a functional lid covering the pitcher opening, some of which secrete fragrance and sweet nectars to lure insect prey (Di Giusto et al., 2010). Additionally, the viscoelastic fluid (Bazile et al., 2015) and slippery epicuticular wax crystals of the inner pitcher wall with the anisotropic effect of lunate cells (Wang et al., 2018) also contribute to the success of prey capture and retention.

Hybridization is the formation of organisms by cross-fertilization between individuals of different species, which is key to the evolution and diversification of 30–70% flowering plants (Soltis and Soltis, 2009). Hybrids generally express phenotypes intermediate to those of parents, which sometimes result in heterosis or hybrid vigor such that the hybrids survive better than both parents, but the opposite is also possible (Orians, 2000; Goulet et al., 2017). Studies comparing the metabolite expression of hybrids with that of their parent species are limited, but existing evidence suggests that hybrids may express more diverse combinations of secondary metabolites or produce unique phytochemicals novel to both parental species (Cheng et al., 2011; Lätti et al., 2011). This leads to an interesting question on the expression of metabolites in the hybrids of *Nepenthes* species.

Hybridization in the species-rich Nepenthaceae is extensive, either naturally or artificially driven by high horticultural demand as ornamental plants. According to Harry James Veitch in 1906 (Veitch, 1979), the first *Nepenthes* artificial hybridization was initiated by John Dominy who produced *N.* × *dominicana* from *N. rafflesiana* with an unnamed species in 1862, followed by *N.* × *hybrida* from *N. khasiana* with an unnamed Bornean species in 1866. Despite the long history, studies on *Nepenthes* hybrids and their chemical compositions are generally limited. Studies of *Nepenthes* have largely focused on the proteins of pitcher fluids (Ravee et al., 2018). Recent studies have reported the effects of endogenous protein depletion on protein secretion in pitcher fluids of *N.* × *ventrata* (Wan Zakaria et al., 2019) and gene expression in pitchers of *N. ampullaria* with transcriptome analysis (Goh et al., 2020). The same proteomics informed by the transcriptomics approach was applied to explore the expression of genes and proteins in the pitchers and fluids, respectively, in *N. ampullaria*, *N. rafflesiana*, and their hybrid *N.* × *hookeriana* (Zulkapli et al., 2017; Zulkapli et al., 2021).

To date, information is only available on certain targeted metabolites that have been successfully isolated and characterized from some *Nepenthes* species. Advanced metabolite identification using high-performance liquid chromatography (HPLC), nuclear magnetic resonance (NMR), and gas chromatography–mass spectrometry (GC-MS) was performed on a few *Nepenthes* species in several studies, but the isolated and characterized compounds were targeted (Eilenberg et al., 2010; Kováčik et al., 2012). There is only one recent untargeted metabolite profiling study comparing the feeding effect on leaf blades and pitchers of *N.* × *ventrata* based on cheminformatics fingerprinting without compound identification (Dávila-Lara et al., 2020). The authors found little impact of feeding on the metabolite composition of pitchers with only small changes in the polar metabolites.

The current study aims to unveil the effects of hybridization in pitcher plants based on comparative metabolic profiling through an untargeted approach for identifying metabolites. Liquid chromatography–mass spectrometry (LC-MS) is used to classify and compare chemical profiles between *N.* × *hookeriana*, a natural hybrid with its parents, *N. ampullaria* and *N. rafflesiana*. The parentage of *N.* × *hookeriana* was previously discovered based on its morphological resemblance to *N. ampullaria* and *N. rafflesiana* (Danser, 1928), which was confirmed by molecular markers, including amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), and inter-simple sequence repeats (ISSR) (Kiew et al., 2003; Yulita and Mansur, 2012). Differences in the morphology and diet between *N. ampullaria* and *N. rafflesiana* may lead to differences in the metabolite compositions. Therefore, this study provides a comprehensive metabolite profiling of these three species to investigate how the parental molecular phenotypes are manifested in *N.* × *hookeriana*.

MATERIALS AND METHODS

Pitcher Sampling

Pitcher tissues of closely related lowland *Nepenthes* species, *N. ampullaria* Jack and *N. rafflesiana* Jack (Bunawan et al., 2017) with their hybrid *N.* × *hookeriana* Lindl., all of which were originally transplanted from Endau-Rompin National Park, were sampled from an experimental terrace at Universiti Kebangsaan Malaysia (2°55′12.7″N, 101°46′59.7″E) (**Supplementary Figure 1**). The pitchers were monitored and measured throughout their development (**Supplementary Figure 2**) to capture the timing of pitcher opening. Pitchers after 7 days of lid opening were chosen as they are reported to be fully functional traps (Bauer et al., 2009). The sampling was performed in the morning between 09:00 and 11:00 h during October to December 2015 with at least four biological replicates from each species (Rosli et al., 2018). Whole pitchers were harvested without the tendril, emptied, and rinsed with deionized water before rapidly frozen in liquid nitrogen and stored at −80°C.

Phytochemical Extraction

The whole-pitcher samples excluding the tendril were ground until fine homogenous powders using pestle and mortar prechilled with liquid nitrogen, collected in a Falcon tube, and freeze-dried for 48 h. Dried powder samples (100 mg) were extracted using 200 μL of methanol:chloroform:water (3:1:1) via the sonication-assisted method (Shyur et al., 2013; Rosli et al., 2017). Samples were vortexed, sonicated in a water bath sonicator at room temperature for 15 min with maximum frequency, vortexed again, and then centrifuged at 10,000 *g* for 10 min. The extracts were filtered through a 0.22-μm PTFE membrane syringe filter before stored at −80°C.

Liquid Chromatography-Mass Spectrometry (LC-MS) Analysis

The chromatographic separation was carried out using a Thermo Scientific C18 column (Acclaim™ Polar Advantage II, 3 × 150 mm, 3 μm particle size) on UltiMate 3000 UHPLC (Dionex) system according to previous reports (Goh et al., 2016; Rosli et al., 2017). The gradient elution was performed for 22 min total run time, using (A) water containing 0.1% formic acid, and (B) 100% acetonitrile as the mobile phase with 0.4 mL/min flow rate at 40°C. The gradient was started at 5% solvent B for 3 min (0–3 min), then increased to 80% solvent B for 7 min (3–10 min) and maintained at 80% solvent B for 5 min (10–15 min). Finally, the gradient was returned to 5% solvent B in 7 min (15–22 min). The injection volume for each sample was 1 μL, and the samples were prepared by dissolving 10 μL of sample extracts in LC-grade methanol (990 μL) spiked with internal standards not present in the studied sample, namely, ribitol, riboflavin, and vanillin at a concentration of 50 ppm each with known basic characteristics, such as retention time (rt) and exact mass. Internal standards are important in metabolic profiling analysis serving as references for relative quantitative analysis and validation on the performance of chromatographic and MS systems (Stobiecki and Kachlicki, 2013). At least four independent biological replicates of pitcher extracts from each species were analyzed.

High-resolution mass spectrometry analysis was carried out using the Bruker Daltonics MicrOTOF-Q III with the following settings: capillary voltage at 4,500 V, nebulizer pressure at 1.2 bar, drying gas flow at 8 L/min with the source temperature at 200°C, and m/z range from 50 to 1,000 Da. The mode of electrospray ionization (ESI) was compared preliminary before the metabolite profiling LC-MS analysis. Most of the metabolites identified through negative mode were present in the total ion chromatogram of positive mode (data not shown). Thus, the positive ESI mode was selected following a higher number of metabolites profiled in preliminary LC-MS analysis and other studies (Goh et al., 2016; Rosli et al., 2017; Dávila-Lara et al., 2020). Independent from the MS analysis of individual samples, the MS/MS analysis was performed for each species using the pooled replicates of all extracts at equal amounts using automated fragmentation settings (Auto-MS/MS) over a mass-to-charge precursor ion ranged between 500 and 1,000 Da (Vargas et al., 2016).

Mass Spectrometry Data Processing

Mass spectrometry data processing was done according to Krug et al. (2008) with some modifications. The mass spectrometry raw data was extracted from Bruker DataAnalysis (version 4.1) and aligned using Bruker Compass ProfileAnalysis (version 2.1). The advanced bucketing setting was selected with the following parameters: $\Delta m/z$ = 20 mDa, Δrt = 10 s, threshold of signal-to-noise ratio = 5, and smoothing width = 4. The correlation coefficient was set at 0.7. The exact mass and predicted molecular formula obtained from SmartFormula, a feature in ProfileAnalysis (version 2.1), was selected based on the highest score with low molecular mass error tolerance range (Δppm = 5) and manually matched to accessible public databases, including METLIN[1] developed for QTOF instruments (Guijas et al., 2018), MassBank[2] the longest standing community database (Horai et al., 2010), and MetFrag[3] (Ruttkies et al., 2016) for putative metabolite identification.

Putative metabolites were retrieved from the databases for having molecular weights within the molecular mass error tolerance range of Δppm < 20 to the query m/z values via positive mode adduct. Full-scan LC-MS data were acquired for statistical analysis to shortlist molecular ions with significant differences between samples, followed by independent precursor-ion (PI) scans to acquire MS/MS data of the PIs. These MS/MS data of precursor m/z and retention time were used to derive the structural information with confidence level 1 (CL1), while metabolite features without MS/MS data were putatively identified at confidence level 2 (CL2) according to the Metabolomics Standards Initiative (Blaženović et al., 2018). Erroneous metabolite identifications were removed through manual inspection based on chemotaxonomy information with

[1]https://metlin.scripps.edu/

[2]http://www.massbank.jp/Search

[3]https://msbi.ipb-halle.de/MetFrag/

reference to the KNApSAcK species–metabolite relationship database (Afendi et al., 2012).

Statistical Analysis

XCMS version 3.7.1[4] was used for a preliminary inspection of feature detection using the mzXML files and default parameters (ID#6675) optimized for "UPLC/Bruker Q-TOF pos" to generate the statistics of identified features and mirror plots of pairwise comparisons between the samples (Huan et al., 2017). For further statistical analysis, the data matrix generated from ProfileAnalysis (version 2.1) comprising peak intensity of each retention time (rt) and m/z value was normalized to the intensity of the most consistent internal standard (vanillin) using MetaboAnalyst 3.0[5], a free online tool for statistical and pathway analyses (Xia and Wishart, 2016). The normalized peak intensity data from MetaboAnalyst was exported into SIMCA-P+ (version 14.1, Umetrics, Umeå, Sweden) for multivariate analysis (MVA). The data was preprocessed with Pareto (Par) scaling which was applied to the normalized data to reduce the effects of background noise interference on sample clustering. MVA based on non-supervised principal component analysis (PCA) and supervised partial least squares discriminant analysis (PLS-DA) were applied to cluster samples according to the profiled metabolite features and to identify distinct metabolites that differentiate between the three species. One-way analysis of variance (ANOVA) with a significance of $P < 0.05$ was conducted on the normalized peak intensity data to differentiate the statistically significant metabolites among the three species. The variable importance in the projection (VIP) ranks the overall contribution of each variable to the PLS-DA model, and those variables with VIP > 1.0 were considered relevant for group discrimination (Xie et al., 2008). The list of important VIPs was based on PLS-DA loading plots with VIP scores > 1. Venn diagrams were generated using a web tool[6].

RESULTS

Morphological Changes During Pitcher Development

For the sampling of pitchers, daily observations of morphological changes were performed *in situ* (**Supplementary Figure 1**) throughout pitcher development (**Figure 1**). The three species showed different pitcher longevity with *N. ampullaria* pitchers being the most long-lived (>6 months) compared to the short-lived (≤2 months) *N. rafflesiana* pitchers with the hybrid pitchers showing intermediate longevity (2–3 months). It took around 1 month (22–33 days) from pitcher initiation at the tendril tip to pitcher opening for all three species. Pitcher development continues after opening and maturity within 7 days with a stable morphology, namely, the positioning of the lid, peristome structure, coloration, and pitcher size (**Figure 1** and **Supplementary Figure 2**). Therefore, 7-day-old mature pitchers after lid opening were chosen to standardize sampling for metabolomics analysis.

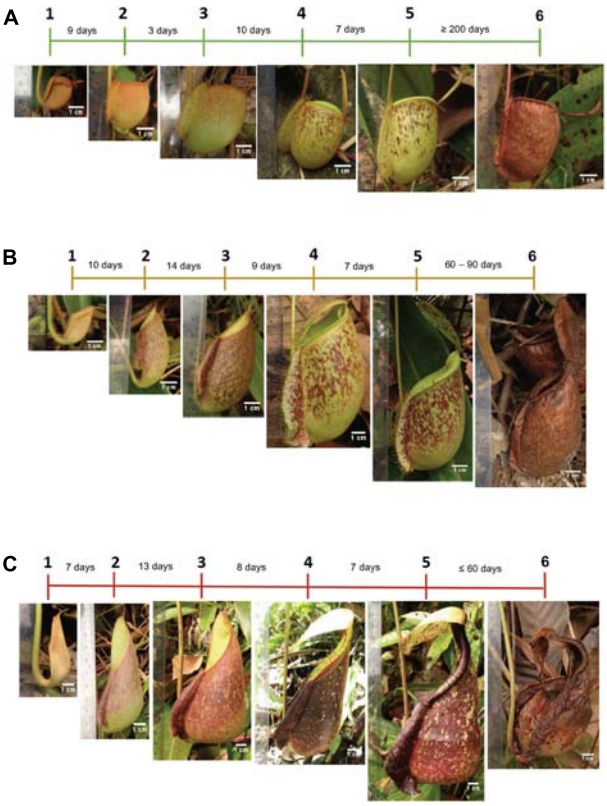


FIGURE 1 | The timeline of pitcher morphological changes of **(A)** *N. ampullaria*, **(B)** *N.* × *hookeriana*, and **(C)** *N. rafflesiana*. Different phases and durations of pitcher development are indicated in the timeline.

Metabolite Profiling

The chromatographic separation of the three *Nepenthes* samples and internal standards can be visualized on a base peak chromatogram (BPC), which shows the peak intensity based on the m/z value and retention time (rt) with extracted ion chromatogram (EIC). All three internal standards were successfully identified in the chromatogram, including ribitol at rt 1.9 min, riboflavin at rt 8.4 min, and vanillin at rt 9.5 min (**Figure 2**). Based on a comparison of all the chromatograms, all three internal standards were present and separated at similar rt in each sample. Samples from the three species showed metabolic separation patterns with different numbers of metabolite features. There were metabolite features uniquely present in one species or all species with different intensities. For instance, a metabolite feature with the highest intensity after internal standards at rt 8 min exhibited the highest intensity in *N. rafflesiana*, moderate in the hybrid, *N.* × *hookeriana*, and the lowest in *N. ampullaria*.

A statistical analysis was performed using XCMS using the raw data from all the samples for pairwise comparisons (**Supplementary Figure 3**) and feature detection (**Supplementary Figure 4**). Comparison between the two

[4]https://xcmsonline.scripps.edu/

[5]http://www.metaboanalyst.ca

[6]http://bioinformatics.psb.ugent.be/webtools/Venn/

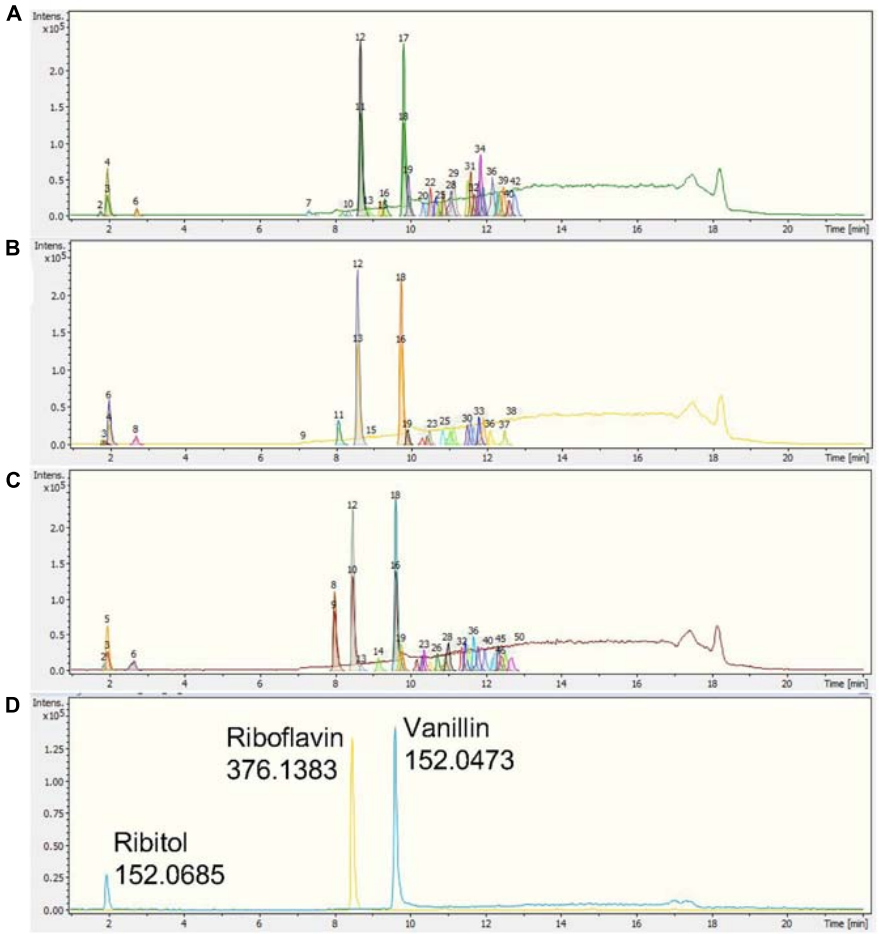


FIGURE 2 | Representative base peak chromatograms (BPCs) with extracted ion chromatograms (EICs) of **(A)** *N. ampullaria*, **(B)** *N.* × *hookeriana*, **(C)** *N. rafflesiana*, and **(D)** a blank with internal standards, namely, ribitol (1.9 min), riboflavin (8.4 min), and vanillin (9.5 min).

parent species showed the biggest difference with 22 significant differentially accumulated features (DAFs), followed by the hybrid and *N. ampullaria* (9 DAFs), and lastly between the hybrid and *N. rafflesiana* (2 DAFs) (**Supplementary Figure 3**).

After filtering and bucketing using Bruker Compass ProfileAnalysis, we successfully profiled a total of 1,441 metabolite features in pitcher samples from all three species (**Supplementary File 1**). The distribution of pitcher metabolite features for each species is shown in a Venn diagram (**Figure 3A**). The pitcher extracts with the highest number of profiled metabolite features were found in *N. ampullaria* with 841 features, followed by the hybrid species, *N.* × *hookeriana* (632) and *N. rafflesiana* (577). Only 221 metabolite features were present in all three species with many features unique to individual species. It is noteworthy that there were 274 (43.3%) features found in *N.* × *hookeriana* which were not found in any of the parent species. Similar numbers of metabolite features were shared by the hybrid with *N. ampullaria* (292, 46.2%) and *N. rafflesiana* (287, 45.4%). However, there were many more unique metabolite features in *N. ampullaria* (549, 65.3%) that were absent in the hybrid, compared to *N. rafflesiana* which shared 49.7% of its total metabolite features with the hybrid. The mass-to-charge ratio (m/z) of metabolite features were rather evenly distributed compared to the retention time (rt) (**Figure 3B**). Most of the unique features found in *N. ampullaria* were detected during early rt (25 s), while the majority of the shared features were found in the middle of rt (400–800 s).

Multivariate Analysis

An unsupervised principal component analysis (PCA) analysis shows the projections of each sample to other samples in a multidimensional space with each point representing an individual sample. The sample dispersions are related to the differences in metabolite compositions such that samples with higher similarities are closer together while samples with bigger differences are further apart. The parent clusters of *N. ampullaria* and *N. rafflesiana* were separated from each other, but the hybrid cluster was dispersed and closer to *N. rafflesiana* (**Figure 4**). This separation only accounted for 21.5% of variation on the first principal component, and the cumulative values of R^2 and Q^2 were lower than 0.5 despite having no outlier. This indicates high variations between biological replicates, which could be

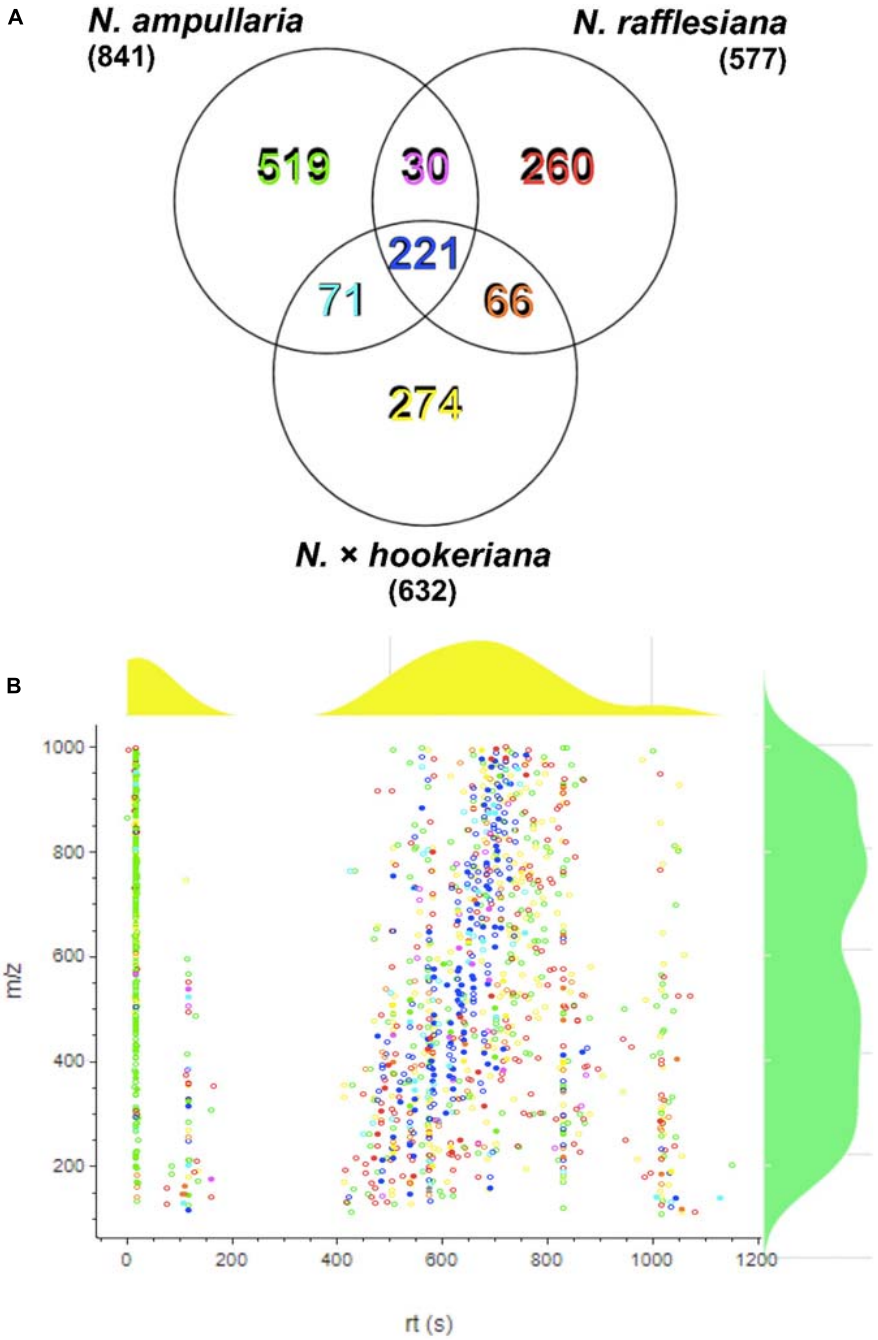


FIGURE 3 | Metabolite profiling. **(A)** Venn analysis of metabolite features identified from LCMS analysis of all three species. Numbers in parentheses show the total number for each species. **(B)** Scatter plot of all metabolite features with overlaying relative density plots for retention time (rt) along the X-axis and mass over charge (m/z) along the Y-axis. Symbol colors correspond to the font colors in the Venn diagram. Filled symbols represent features with putative identification.

due to environmental factors since sampling was performed on-site in the field.

A supervised partial least squares-discriminant analysis (PLS-DA) was then performed for sample clustering with species information (**Figure 5A**). *N. ampullaria* samples were closely clustered and separated from *N. rafflesiana* along the first component that accounted for 19.1% variations. *N.* × *hookeriana* samples were more dispersed and closer to *N. rafflesiana* on the first component with separation mainly on the second component that explained 10.8% variations. These two components appeared to be a good fit model with a predictive value of confidence prediction (Q2) at 35% lower than the value of R2Y at 88%, which suggests no overfitting as supported by a permutation test (**Supplementary Figure 5**).

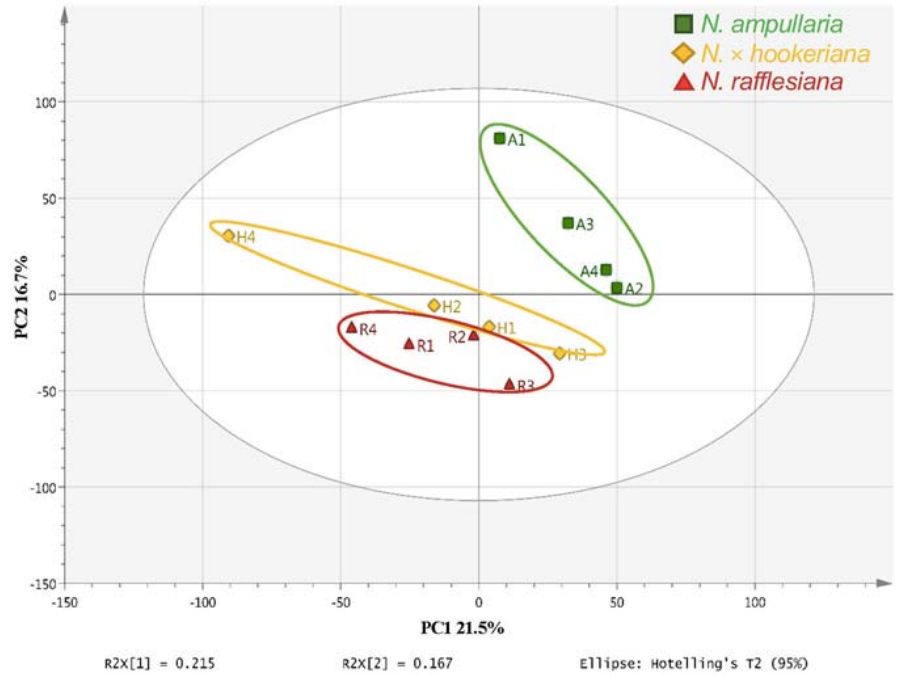


FIGURE 4 | PCA score plot of pitcher extracts from different *Nepenthes* species.

The metabolite features responsible for the discrimination between the three species can be observed in the PLS-DA loading plot (**Figure 5B**) with a higher contribution by the metabolites that are projected further from the center. A total of 324 (22.5%) metabolite features with variables important in projection (VIP) values of more than 1 were identified in the PLS-DA loading plot (**Figure 5B**), in which 102 VIPs were shared among the three species with *N. ampullaria* having the highest number of VIPs (**Figure 5C**). Apart from VIPs, univariate statistical analysis can identify statistically significant metabolites by comparing intensity differences of metabolite features in each sample, whereby only statistically significant features are considered important and maybe biological or chemical markers. Only 55 (17%) out of 324 VIPs were statistically significant ($P < 0.05$) based on the one-way variance analysis (ANOVA) in which 31 were unique to *N. ampullaria*. PCA performed based on these 55 significant VIPs can separate the three species into three distinct clusters and explained 63.2% of variation on the first principal component (PC1) and 14.2% for PC2 (**Supplementary Figure 6**). This supports that these metabolite features contribute significantly to the distinction of each species and account for their differences in metabolism.

Some of these significant metabolite features were plotted in a heat map with normalized intensity values (**Figure 6**). A total of 27 out of the 36 metabolite features showed higher abundance in *N. ampullaria* compared to six metabolites in *N. rafflesiana* and only three in the hybrid. This indicates that *N. ampullaria* displayed a more distinct metabolite distribution compared to *N. rafflesiana* and *N.* × *hookeriana*, which is consistent with the results from the multivariate analysis.

Metabolite Identification

Due to the labor-intensive metabolite identification, we mainly focused on VIP and metabolite features with high abundance for putative identification of metabolites by manually matching the accurate masses, fragmentation patterns, and MS/MS spectral data with reference to public metabolite databases, including METLIN, MassBank, and MetFrag (**Supplementary File 1**). For every annotated metabolite, we manually checked against the KNApSAcK database to make sure the metabolite is produced in plants. A total of 95 (29.3%) of 324 VIPs were among the 105 (7.3%) putatively identified metabolites from 1,441 metabolite features based on our manual approach.

DISCUSSION

Untargeted Metabolite Profiling of *Nepenthes* Pitcher Extracts

This untargeted metabolite profiling study is the first comparative metabolomics analysis of *Nepenthes* pitchers from different species. For non-model organisms, LC-MS analysis is the preferred method for metabolic profiling with good selection and sensitivity to analyze various types of metabolites with large mass range differences and different physicochemical properties, such as volatility, heat loss, and polarity (Stobiecki and Kachlicki, 2013). The chromatographic separation can reduce the sample complexity and increase the MS detection sensitivity with less background noise. Furthermore, the electrospray ionization (ESI) mode of soft ionization can produce large amounts of ions through the exchange of charges in solution and often form intact

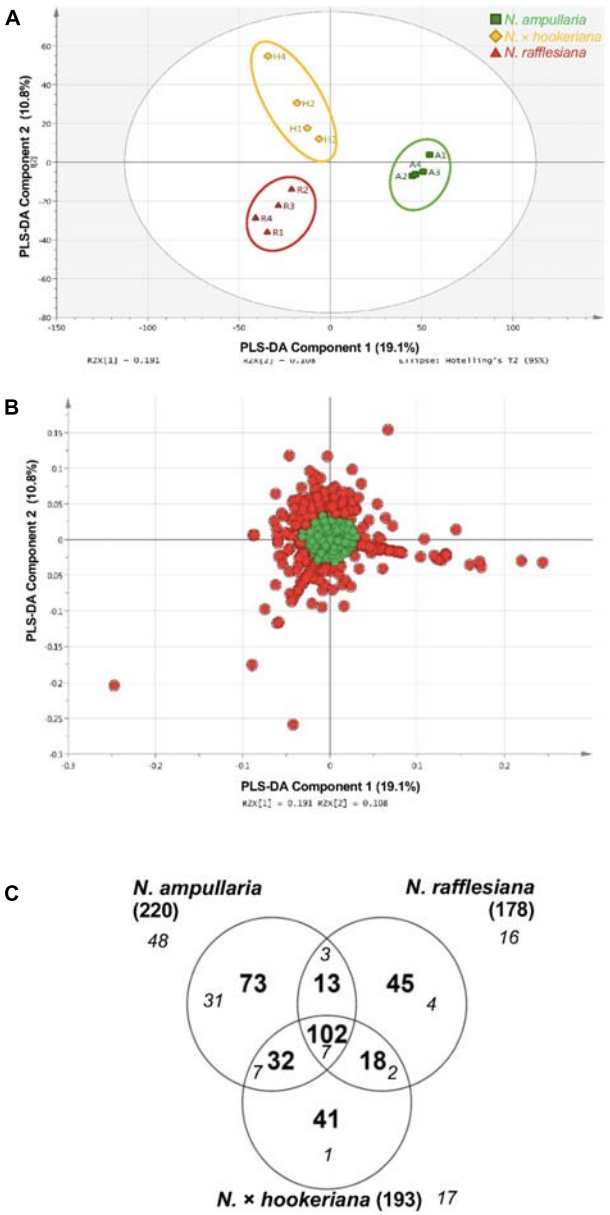


FIGURE 5 | PLS-DA analysis of pitcher extracts from different *Nepenthes* species. **(A)** PLS-DA score plot. **(B)** PLS-DA loading plot with red dots showing the VIP features. **(C)** Venn analysis of VIP features in three *Nepenthes* species. Numbers in parentheses show the total number for each species. Italic numbers in smaller font size represent the numbers of significant ($P < 0.05$) VIP.

molecular ions for initial identification of metabolites (Zhou et al., 2012). Time-of-flight (ToF) technology applied in this study further improves the mass accuracy to facilitate the identification and relative quantification of metabolites.

To compare the three *Nepenthes* species with four biological replicates of pitcher extracts from each species, their metabolite profile data were processed and normalized based on vanillin because it exhibited the most consistent peak intensities in all samples with the lowest coefficient of variation (CV) compared to ribitol and riboflavin (**Supplementary File 1**). Overall, multivariate analysis (MVA) showed that the metabolite composition of the hybrid is more like *N. rafflesiana* than *N. ampullaria*. This corroborates the recent report that the hybrid showed more similar transcriptome and secretome to *N. rafflesiana* than *N. ampullaria* (Zulkapli et al., 2021). Previously, it was proposed that the variation of colors and motifs were determined by male *Nepenthes* based on a genetic study (Yulita and Mansur, 2012). *N.* × *hookeriana* "spotted" as a hybrid between male *N. rafflesiana* and female *N. ampullaria* "green" was more like *N. rafflesiana* while *N.* × *hookeriana* "green" was hypothesized to be a hybrid between male *N. ampullaria* "green" and female *N. rafflesiana*. Hence, this explains the molecular compositions of samples in this study with *N.* × *hookeriana* "spotted" (**Figure 1**).

A recent untargeted metabolite profiling study of *N.* × *ventrata* pitchers was based on cheminformatics fingerprinting to infer metabolite classification without compound identification, which hindered functional interpretation (Dávila-Lara et al., 2020). At present, metabolite identification for untargeted metabolomics analysis largely depends on mass-based search using the m/z values of molecular ions of interest against databases (Xiao et al., 2012; Blaženović et al., 2018). In this study, manual searches against different public databases putatively identified a total of 105 metabolites with 95 VIPs in which only four metabolites were identified with MS/MS data. These putative metabolites are only suggestive because of limited information on *Nepenthes* metabolites in the public databases and the low hit of auto MS/MS spectra for most of the compounds. To our knowledge, this is the most comprehensive study of metabolite profiling ever reported in *Nepenthes* studies.

Due to the complexity and diversity of the chemical composition of metabolites, metabolite identification is still a big challenge in metabolomics analysis (Aretz and Meierhofer, 2016). Commonly, there are less than 20% identified compounds in untargeted analysis, such as the Metabolomics Workbench (Sud et al., 2016) or the European metabolomics repository MetaboLights (Haug et al., 2013) even for the model organisms. A comprehensive spectral database for LC-MS is almost impossible with various instruments, methods, and acquisition of metabolic data, such as collision, fragmentation, and various settings resulting in varied retention time (rt) and molecular fragments (Aretz and Meierhofer, 2016).

Functional Implications of Putatively Identified Metabolites in *Nepenthes* Pitchers

The most significant VIP metabolite feature (476.3s:449.102) was the most abundant metabolite in the pitchers with its peak clearly detectable at the eighth minute with different intensities in each species showing the highest abundance in *N. rafflesiana*, intermediate in the hybrid, and the lowest in *N. ampullaria* (**Figure 2**). VIP refers to metabolites that contribute to the discrimination between samples, and metabolites with high VIP

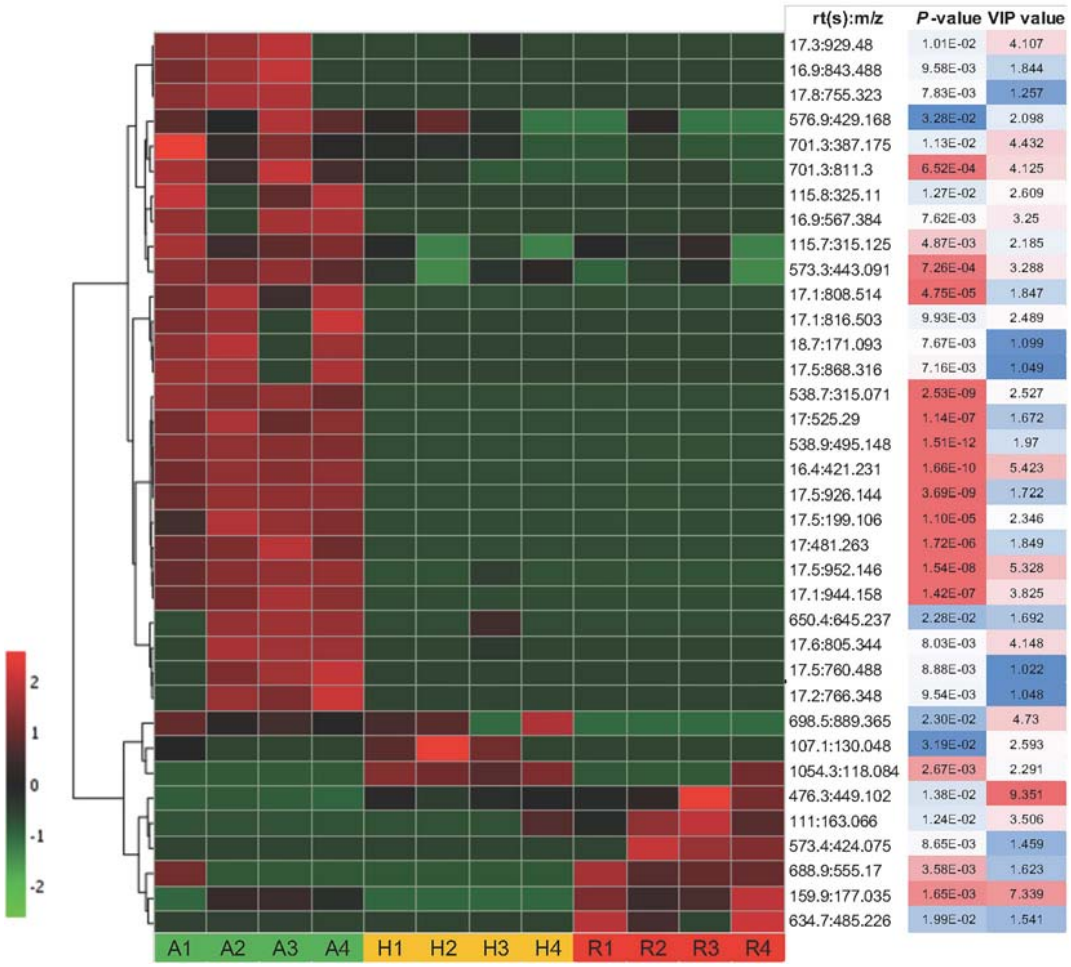


FIGURE 6 | Heat map of statistically significant features in all three species. The color temperature represents the relative peak intensity with hotter red color showing higher intensity. Putative identification of metabolite features can refer to **Supplementary File 1**.

values can serve as biological or chemical markers (Beatriz et al., 2014). A MS/MS spectrum match search using the precursor m/z with MS/MS peaks (**Supplementary Figure 7**) found 476.3s:449.102 matches to kaempferol 3-glucoside (astragalin) and cyanidin 3-O-glucoside (chrysenthemin), both with a METLIN score of 100 and Δppm = 11. However, further fragment similarity search only found hit to astragalin at a fragment peak of m/z 287.054 with Δppm = 9 (**Supplementary Figure 7D**). Nonetheless, we cannot exclude the possibility that 476.3s:449.102 could be other isomers, kaempferol O-glucosides, or luteolin O-glucosides without further validation and structural elucidation using reference compounds and NMR. Astragalin is a flavonol involved in the biosynthesis of flavones and flavonols. Astragalin is a type of yellow pigment that could protect plants from UV radiation (Nakabayashi, 1955). Therefore, this compound may contribute to the pitcher color and confer protection from UV radiation (Lavola et al., 2003). Astragalin is associated with various pharmacological and biological activities, including anti-inflammatory, antioxidant activity, anti-atopic dermatitis activity, TNF-a, IL-1b, and IL-6 production inhibiting activity, and inhibiting histamine release in human blood cells (Riaz et al., 2018).

This finding agrees with Kováčik et al. (2012) that found phenols constituted the most abundant metabolites in *Nepenthes*. The results of several other phytochemical studies also emphasized phenolic compounds in *Nepenthes* roots and leaves, particularly naphthoquinones (Cannon et al., 1980; Aung et al., 2002; Van Thanh et al., 2015). Flavonoids were found to be the dominant phenols in the pitcher extracts, apart from quinones, quinic acids, monolignols, lignans, and phenylpropanoids (**Supplementary File 1**). These flavonoids have antioxidant properties that protect plants from environmental stresses, especially oxidative stress and ultraviolet (UV) light (Pollastri and Tattini, 2011). Flavonoids involved in UV light protection include flavones and flavonols that also act as antioxidants in protecting plants from oxidative stress caused by temperatures higher than suboptimal temperatures (Herrmann, 1976; Jaakola and Hohtola, 2010). Flavan is a product of a double reduction of flavanone. Most of the flavans are soluble in lipids, making these compounds often found on the fruit peel and the cutin layer of

the leaf surface (Swanson, 2003). According to Mierziak et al. (2014), the strongest antifungal activity is shown by unmodified flavones and flavanones. Therefore, these compounds are commonly used in plants as phytoalexins to fight fungi and insects (Swanson, 2003). Chalcones are usually intermediates in flavonoid biosynthesis and generally have low compositions in most plants but display a variety of biological activities as traditional medicine (Rozmer and Perjési, 2016).

Anthocyanins are responsible for color in plants. Cyanidin and delphinidin are parts of anthocyanins that contribute to a variety of colors, including orange, red, and purple, depending on the types, levels, and structural modifications of anthocyanins, the presence of co-pigments such as flavones or flavonol, and pH variations (Yu and McGonigle, 2005). Anthocyanidins were present in pitchers of all species but in relatively low abundance (**Supplementary File 1**). These compounds may be involved in the pitcher pigmentation and coloration, such as spots on the pitchers.

The presence of flavonols and flavones, which were highly abundant in *N. rafflesiana*, may also contribute to the red coloration of pitchers, although the distribution of anthocyanidins in the species is low (**Supplementary File 1**). A high temperature can inhibit anthocyanin biosynthesis caused by chemical degradation from the accumulation of polyphenol oxidase and peroxide enzymes (Vaknin et al., 2005; Mori et al., 2007). This partly explains the effect of latitude on plant flavonoid biosynthesis (Jaakola and Hohtola, 2010). For example, a tropical herb plant *Polygonum minus* growing in the lowland and exposed to higher temperatures contained higher amounts of flavones and flavonols but lower anthocyanidin content compared to those from highland and exposed to lower temperatures (Goh et al., 2016). However, whether or not *N. rafflesiana* plants growing at lowland and highland exhibit different compositions of anthocyanidins remains to be investigated.

In previous reports, phenolic compound naphthoquinones such as plumbagin with numerous bioactivities are known to be highly abundant in *Nepenthes* (Likhitwitayawuid et al., 1998; Gwee et al., 2014). Plumbagin has been suggested with a possible role in attracting their insect prey with distinct volatile preferences toward the sticky trap of another insectivorous plant, *Drosera auriculata* (El-Sayed et al., 2016; Widhalm and Rhodes, 2016). According to Shin et al. (2007), some naphthoquinones with antifungal activities can be found in the roots of *N. rafflesiana*, *N. thorelii*, *N. gracilis*, and *N. insignis*. Studies have also shown that naphthoquinones can be found in the leaves of *N. khasiana* and *N. gracilis* (Aung et al., 2002; Eilenberg et al., 2010; Raj et al., 2011). Naphthoquinones might have several functions in *Nepenthes*, including photosynthesis, pigmentation and coloration, and protection against UV light, desiccation, and insects (Widhalm and Rhodes, 2016). Raj et al. (2011) speculated toxic or anesthetic effects of naphthoquinones on insect prey in *N. khasiana* pitchers. Furthermore, derivatives of naphthoquinones, droserone, and 5-O-methyl droserone could protect and maintain the sterility of the pitcher fluids against microbes from the trapped prey and together with plumbagin may act as molecular triggers in prey capture and digestion (Raj et al., 2011).

In this study, there were a few naphthoquinones putatively identified (**Supplementary File 1**), which have never been reported. There might be several factors that contribute to the detection of naphthoquinones, including the differences in the extraction using different solvents or analytical methods such as GC-MS (Raj et al., 2011). Besides that, Eilenberg et al. (2010) showed that naphthoquinones in *N. khasiana* pitcher fluids only present after chitin induction, which suggests the induced production of naphthoquinones.

As mentioned above, this study is limited by the uncertain metabolite identification for a detailed discussion of metabolites in *Nepenthes* pitchers. Furthermore, this also hindered further discussion on whether the different metabolite compositions in the pitchers were accountable by the different dietary habits of the parent and hybrid species. Nevertheless, this study is meaningful in finding potential chemical markers and bioactive compounds. Various reports described the curative effects of extracts from *Nepenthes* on diseases, for example, jaundice, hepatitis, gastric ulcers, ureteral stones, diarrhea, diabetes, cough, fever, hypertension, urinary system infections (Thao et al., 2016), malaria (Likhitwitayawuid et al., 1998), and Staphylococcus infection (Wiart et al., 2004), and recently on different kinds of oral cancer cells (Tang et al., 2019). Thus, further works on the validation of metabolites from *Nepenthes* as well as the analysis of their putative pharmaceutical uses are promising to explore new compounds and therapeutics. Subsequently, compounds with bioactivities of interest can be targeted for isolation, structural elucidation, and functional characterization.

CONCLUSION

This study provides a comprehensive profiling and putative identification of metabolites from *Nepenthes* pitchers through the untargeted LC-MS approach with a total of 1,441 metabolite features and 105 putatively identified metabolites. MVA performed on profiled metabolites showed that the chemical compositions of hybrid pitchers were more similar to that of *N. rafflesiana* than *N. ampullaria*. Astragalin, which is a pharmacologically useful flavonol, was putatively identified as the most significant compound in distinguishing between the three *Nepenthes* species with potential as a chemical marker. However, the influence of dietary habit on the differences in the metabolite compositions between the three taxa remains unclear.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

MR and H-HG conceived the study and experiments. MR performed the experiments and analyzed data. MR, AM, KA, SB, and H-HG discussed the data and wrote the manuscript. H-HG

acquired the funding for this work. All authors read and agreed to the present version of the manuscript.

FUNDING

This work was funded by FRGS/2/2014/SG05/UKM/02/4 from the Malaysian Ministry of High Education and UKM Research University Grant DIP-2014-008 awarded to H-HG. The research group is currently supported by DIP-2020-005 and FRGS/1/2019/STG05/UKM/02/10.

ACKNOWLEDGMENTS

We thank Prof. Dr. Jumaat Haji Adam for providing access to the experimental terrace and contributing the pitcher samples. We are also grateful to both reviewers for their constructive comments in improving this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.655004/full#supplementary-material

Supplementary Figure 1 | *Nepenthes* pitcher plants in this study. **(A–C)** *Nepenthes ampullaria.* **(D,E)** *Nepenthes* × *hookeriana.* **(F–H,J,K)** *Nepenthes rafflesiana* **(A,D,F)** Whole pitcher plants. **(B)** A cluster of *N. ampullaria* pitchers on the ground. **(C,E,H)** A single pitcher. **(G)** An aerial pitcher of *N. rafflesiana*. **(I)** Experimental terrace for pitcher sampling of all three species. **(J)** Characteristics of a typical *Nepenthes* pitcher. **(K)** External and internal structures of a *N. rafflesiana* pitcher.

Supplementary Figure 2 | (A) Measurement of pitcher length. **(B)** Pitcher length at each growth phase. Error bars showing standard errors of six biological replicates.

Supplementary Figure 3 | Mirror cloud plots of differentially accumulated features (DAFs) in pairwise comparisons of **(A)** *N. ampullaria* and *N. rafflesiana,* **(B)** *N. ampullaria* and *N.* × *hookeriana*, and **(C)** *N. rafflesiana* and *N.* × *hookeriana*. The numbers of significant ($P < 0.001$, rt > 1) DAFs are shown in parentheses and listed on the side with details in ascending *P*-value. DAFs are shown for lower (top) or higher (bottom) accumulation in the first species of comparison. Circle size depicts the average fold change. Y-axis indicates the m/z value of the feature; the further away from the origin, the higher the m/z. The superimposed, raw base-peak chromatograms (BPC) of all runs are shown in the background, colored according to different samples. Insets show peak intensity boxplots of selected DAFs with m/z values.

Supplementary Figure 4 | Venn analysis of identified features in XCMS pairwise comparisons of all three species. A: *N. ampullaria*; H: *N.* × *hookeriana*; R: *N. rafflesiana.*

Supplementary Figure 5 | Permutation test of PLS-DA validation model.

Supplementary Figure 6 | PCA score plot analysis based on significant features. Each symbol represents different species of *Nepenthes*, green square: *N. ampullaria*, yellow diamond: *N.* × *hookeriana*, and red triangle: *N. rafflesiana.*

Supplementary Figure 7 | Putative identification of metabolite feature 476.3s:449.102 based on tandem MS data. **(A)** MS/MS spectra of 476.3s:449.102 with the list of peaks (m/z and intensity). **(B)** Chemical structure of astragalin and chrysenthemin. Comparison for putative metabolite identification based on **(C)** METLIN MS/MS spectrum and **(D)** Fragment Similarity Search. The blue square in **(A)** shows the precursor ion. For **(C)**, the asterisk indicates the highest peak of precursor ion, while the black and cyan lines represent the MS/MS spectrum of the putative metabolite and the search fragments, respectively.

Supplementary File 1 | Normalized peak intensity values of metabolite features with putative metabolite identification.

REFERENCES

Adam, J. H. (1997). Prey spectra of bomean *Nepenthes* species (Nepenthaceae) in relation to their habitat. *Pertanika J. Trop. Agric. Sci.* 20, 121–134.

Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., et al. (2012). KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant Cell Physiol.* 53:e1. doi: 10.1093/pcp/pcr165

Aretz, I., and Meierhofer, D. (2016). Advantages and pitfalls of mass spectrometry based metabolome profiling in systems biology. *Int. J. Mol. Sci.* 17:632. doi: 10.3390/ijms17050632

Aung, H., Chia, L., Goh, N., Chia, T., Ahmed, A., Pare, P., et al. (2002). Phenolic constituents from the leaves of the carnivorous plant *Nepenthes gracilis*. *Fitoterapia* 73, 445–447. doi: 10.1016/s0367-326x(02)00113-2

Bauer, U., Willmes, C., and Federle, W. (2009). Effect of pitcher age on trapping efficiency and natural prey capture in carnivorous *Nepenthes rafflesiana* plants. *Ann. Bot.* 103, 1219–1226. doi: 10.1093/aob/mcp065

Bazile, V., Moguédec, G. L., Marshall, D. J., and Gaume, L. (2015). Fluid physico-chemical properties influence capture and diet in *Nepenthes* pitcher plants. *Ann. Bot.* 115, 705–716. doi: 10.1093/aob/mcu266

Beatriz, G.-P., Lennart, E., and Johan, T. (2014). Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS). *J. Chemom.* 28, 623–632. doi: 10.1002/cem.2627

Blaženović, I., Kind, T., Ji, J., and Fiehn, O. (2018). Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. *Metabolites* 8:31. doi: 10.3390/metabo8020031

Bunawan, H., Yen, C. C., Yaakop, S., and Noor, N. M. (2017). Phylogenetic inferences of *Nepenthes* species in peninsular Malaysia revealed by chloroplast (trnL intron) and nuclear (ITS) DNA sequences. *BMC Res. Notes* 10:67. doi: 10.1186/s13104-017-2379-1

Cannon, J., Lojanapiwatna, V., Raston, C., Sinchai, W., and White, A. (1980). The quinones of *Nepenthes rafflesiana*. the crystal structure of 2, 5-dihydroxy-3, 8-dimethoxy-7-methylnaphtho-1, 4-quinone (nepenthone-E) and a synthesis of 2, 5-dihydroxy-3-methoxy-7-methylnaphtho-1, 4-quinone (nepenthone-C). *Aust. J. Chem.* 33, 1073–1093. doi: 10.1071/ch9801073

Cheng, D., Vrieling, K., and Klinkhamer, P. G. L. (2011). The effect of hybridization on secondary metabolites and herbivore resistance: implications for the evolution of chemical diversity in plants. *Phytochem. Rev.* 10, 107–117. doi: 10.1007/s11101-010-9194-9

Clarke, S. J., and Robinson, A. S. (2018). "Species of carnivorous plants," in *Carnivorous Plants: Physiology, Ecology, and Evolution*, eds A. Ellison and L. Adamec (Oxford: Oxford University Press), 411–413.

Danser, B. H. (1928). "The Nepenthaceae of the Netherlands Indies," in *Bulletin du Jardin Botanique de Buitenzorg, Série III*, Vol. 9. Research Triangle Park, NC: Natural History Publications, 249–438.

Dávila-Lara, A., Rodríguez-López, C. E., O'Connor, S. E., and Mithöfer, A. (2020). Metabolomics analysis reveals tissue-specific metabolite compositions in leaf blade and traps of carnivorous *Nepenthes* plants. *Int. J. Mol. Sci.* 21:4376. doi: 10.3390/ijms21124376

Di Giusto, B., Bessière, J. M., Guéroult, M., Lim, L. B., Marshall, D. J., Hossaert-McKey, M., et al. (2010). Flower-scent mimicry masks a deadly trap in the carnivorous plant *Nepenthes rafflesiana*. *J. Ecol.* 98, 845–856. doi: 10.1111/j.1365-2745.2010.01665.x

Eilenberg, H., Pnini-Cohen, S., Rahamim, Y., Sionov, E., Segal, E., Carmeli, S., et al. (2010). Induced production of antifungal naphthoquinones in the pitchers

of the carnivorous plant *Nepenthes khasiana*. *J. Exp. Bot.* 61, 911–922. doi: 10.1093/jxb/erp359

El-Sayed, A. M., Byers, J. A., and Suckling, D. M. (2016). Pollinator-prey conflicts in carnivorous plants: when flower and trap properties mean life or death. *Sci. Rep.* 6:21065.

Goh, H. H., Baharin, A., Mohd Salleh, F. I., Ravee, R., Wan Zakaria, W. N. A., and Mohd Noor, N. (2020). Transcriptome-wide shift from photosynthesis and energy metabolism upon endogenous fluid protein depletion in young *Nepenthes ampullaria* pitchers. *Sci. Rep.* 10:6575.

Goh, H. H., Khairudin, K., Sukiran, N. A., Normah, M. N., and Baharum, S. N. (2016). Metabolite profiling reveals temperature effects on the VOCs and flavonoids of different plant populations. *Plant Biol.* 18, 130–139. doi: 10.1111/plb.12403

Goulet, B. E., Roda, F., and Hopkins, R. (2017). Hybridization in plants: old ideas, new techniques. *Plant Physiol.* 173, 65–78. doi: 10.1104/pp.16.01340

Guijas, C., Montenegro-Burke, J. R., Domingo-Almenara, X., Palermo, A., Warth, B., Hermann, G., et al. (2018). METLIN: a technology platform for identifying knowns and unknowns. *Anal. Chem.* 90, 3156–3164. doi: 10.1021/acs.analchem.7b04424

Gwee, P. S., Khoo, K. S., Ong, H. C., and Sit, N. W. (2014). Bioactivity-guided isolation and structural characterization of the antifungal compound, plumbagin, from *Nepenthes gracilis*. *Pharm. Biol.* 52, 1526–1531. doi: 10.3109/13880209.2014.902083

Haug, K., Salek, R. M., Conesa, P., Hastings, J., De Matos, P., Rijnbeek, M., et al. (2013). MetaboLights - An open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* 41, D781–D786.

Herrmann, K. (1976). Flavonols and flavones in food plants: a review. *Int. J. Food Sci.* 11, 433–448. doi: 10.1111/j.1365-2621.1976.tb00743.x

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., et al. (2010). MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* 45, 703–714. doi: 10.1002/jms.1777

Huan, T., Forsberg, E. M., Rinehart, D., Johnson, C. H., Ivanisevic, J., Benton, H. P., et al. (2017). Systems biology guided by XCMS Online metabolomics. *Nat. Methods* 14, 461–462. doi: 10.1038/nmeth.4260

Jaakola, L., and Hohtola, A. (2010). Effect of latitude on flavonoid biosynthesis in plants. *Plant Cell Environ.* 33, 1239–1247.

Kiew, R., Teo, L., and Gan, Y. (2003). Assessment of the hybrid status of some malesian plants using amplified fragment length polymorphism. *Telopea* 10, 225–233. doi: 10.7751/telopea20035617

Kováčik, J., Klejdus, B., and Repčáková, K. (2012). Phenolic metabolites in carnivorous plants: inter-specific comparison and physiological studies. *Plant Physiol. Biochem.* 52, 21–27. doi: 10.1016/j.plaphy.2011.11.007

Krug, D., Zurek, G., Schneider, B., Garcia, R., and Müller, R. (2008). Efficient mining of myxobacterial metabolite profiles enabled by liquid chromatography-electrospray ionisation-time-of-flight mass spectrometry and compound-based principal component analysis. *Anal. Chim. Acta* 624, 97–106. doi: 10.1016/j.aca.2008.06.036

Lätti, A. K., Riihinen, K. R., and Jaakola, L. (2011). Phenolic compounds in berries and flowers of a natural hybrid between bilberry and lingonberry (*Vaccinium* × *intermedium* Ruthe). *Phytochemistry* 72, 810–815. doi: 10.1016/j.phytochem.2011.02.015

Lavola, A., Aphalo, P. J., Lahti, M., and Julkunen-Tiitto, R. (2003). Nutrient availability and the effect of increasing UV-B radiation on secondary plant compounds in Scots pine. *Environ. Exp. Bot.* 49, 49–60. doi: 10.1016/s0098-8472(02)00057-6

Likhitwitayawuid, K., Kaewamatawong, R., Ruangrungsi, N., and Krungkrai, J. (1998). Antimalarial naphthoquinones from *Nepenthes thorelii*. *Planta Med.* 64, 237–241. doi: 10.1055/s-2006-957417

Mierziak, J., Kostyn, K., and Kulma, A. (2014). Flavonoids as important molecules of plant interactions with the environment. *Molecules* 19, 16240–16265. doi: 10.3390/molecules191016240

Moran, J. A., Clarke, C. M., and Hawkins, B. J. (2003). From carnivore to detritivore? Isotopic evidence for leaf litter utilization by the tropical pitcher plant *Nepenthes ampullaria*. *Int. J. Plant Sci.* 164, 635–639. doi: 10.1086/375422

Mori, K., Goto-Yamamoto, N., Kitayama, M., and Hashizume, K. (2007). Loss of anthocyanins in red-wine grape under high temperature. *J. Exp. Bot.* 58, 1935–1945. doi: 10.1093/jxb/erm055

Murphy, B., Forest, F., Barraclough, T., Rosindell, J., Bellot, S., Cowan, R., et al. (2020). A phylogenomic analysis of *Nepenthes* (Nepenthaceae). *Mol. Phylogenet. Evol.* 144:106668. doi: 10.1016/j.ympev.2019.106668

Nakabayashi, T. (1955). Isolation of astragalin and isoquercitrin from bracken, *Pteridium aquilinum*. *J. Agric. Chem. Soc. Jpn.* 19, 104–109. doi: 10.1271/bbb1924.19.104

Orians, C. M. (2000). The effects of hybridization in plants on secondary chemistry: implications for the ecology and evolution of plant - Herbivore interactions. *Am. J. Bot.* 87, 1749–1756. doi: 10.2307/2656824

Pavlovič, A., Slováková, L., and Šantrůček, J. (2011). Nutritional benefit from leaf litter utilization in the pitcher plant *Nepenthes ampullaria*. *Plant Cell Environ.* 34, 1865–1873. doi: 10.1111/j.1365-3040.2011.02382.x

Pollastri, S., and Tattini, M. (2011). Flavonols: old compounds for old roles. *Ann. Bot.* 108, 1225–1233. doi: 10.1093/aob/mcr234

Raj, G., Kurup, R., Hussain, A. A., and Baby, S. (2011). Distribution of naphthoquinones, plumbagin, droserone, and 5-O-methyl droserone in chitin-induced and uninduced *Nepenthes khasiana*: molecular events in prey capture. *J. Exp. Bot.* 62, 5429–5436. doi: 10.1093/jxb/err219

Ravee, R., Salleh, F. M., and Goh, H. H. (2018). Discovery of digestive enzymes in carnivorous plants with focus on proteases. *Peer J.* 2018:e4914. doi: 10.7717/peerj.4914

Riaz, A., Rasul, A., Hussain, G., Zahoor, M. K., Jabeen, F., Subhani, Z., et al. (2018). Astragalin: a bioactive phytochemical with potential therapeutic activities. *Adv. Pharmacol. Sci.* 2018:9794625.

Rosli, M. A. F., Azizan, K. A., Baharum, S. N., and Goh, H.-H. (2017). Mass spectrometry data of metabolomics analysis of *Nepenthes* pitchers. *Data Brief.* 14, 295–297. doi: 10.1016/j.dib.2017.07.068

Rosli, M. A. F., Azizan, K. A., and Goh, H.-H. (2018). Antioxidant activity of pitcher extracts from three *Nepenthes* species. *Sains Malays.* 47, 3069–3075. doi: 10.17576/jsm-2018-4712-17

Rozmer, Z., and Perjési, P. (2016). Naturally occurring chalcones and their biological activities. *Phytochem. Rev.* 15, 87–120. doi: 10.1007/s11101-014-9387-8

Ruttkies, C., Schymanski, E. L., Wolf, S., Hollender, J., and Neumann, S. (2016). MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J. Cheminform.* 8:3.

Shin, K.-S., Lee, S.-K., and Cha, B.-J. (2007). Antifungal activity of plumbagin purified from leaves of *Nepenthes ventricosa* x maxima against phytopathogenic fungi. *Plant Pathol. J.* 23, 113–115. doi: 10.5423/ppj.2007.23.2.113

Shyur, L. F., Liu, C. P., and Chien, S. C. (2013). "Metabolomics in herbal medicine research," in *The Handbook of Plant Metabolomics*, eds W. Weckwerth and G. Kahl (Hoboken, NJ: Wiley), 155–174. doi: 10.1002/9783527669882.ch8

Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039

Stobiecki, M., and Kachlicki, P. (2013). "Liquid chromatographic-mass spectrometric analysis of flavonoids," in *The Handbook of Plant Metabolomics*, eds W. Weckwerth and G. Kahl (Hoboken, NJ: Wiley), 197–213. doi: 10.1002/9783527669882.ch10

Sud, M., Fahy, E., Cotter, D., Azam, K., Vadivelu, I., Burant, C., et al. (2016). Metabolomics workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* 44, D463–D470.

Swanson, B. G. (2003). "Tannins and polyphenols," in *Encyclopedia of Food Sciences and Nutrition*, 2nd Edn, ed. B. Caballero (Oxford: Academic Press), 5729–5733. doi: 10.1016/b0-12-227055-x/01178-0

Tang, J. Y., Peng, S. Y., Cheng, Y. B., Wang, C. L., Farooqi, A. A., Yu, T. J., et al. (2019). Ethyl acetate extract of *Nepenthes adrianii* x *clipeata* induces antiproliferation, apoptosis, and DNA damage against oral cancer cells through oxidative stress. *Environ. Toxicol.* 34, 891–901. doi: 10.1002/tox.22748

Thao, N. P., Luyen, B. T. T., Koo, J. E., Kim, S., Koh, Y. S., Van Thanh, N., et al. (2016). In vitro anti-inflammatory components isolated from the carnivorous plant *Nepenthes mirabilis* (Lour.) Rafarin. *Pharm. Biol.* 54, 588–594. doi: 10.3109/13880209.2015.1067234

Vaknin, H., Bar-Akiva, A., Ovadia, R., Nissim-Levi, A., Forer, I., Weiss, D., et al. (2005). Active anthocyanin degradation in *Brunfelsia calycina* (yesterday-today-tomorrow) flowers. *Planta* 222, 19–26. doi: 10.1007/s00425-005-1509-5

Van Thanh, N., Thao, N. P., Huong, P. T. T., Lee, S. H., Jang, H. D., Cuong, N. X., et al. (2015). Naphthoquinone and flavonoid constituents from the carnivorous plant *Nepenthes mirabilis* and their anti-osteoporotic and antioxidant activities. *Phytochem. Lett.* 11, 254–259. doi: 10.1016/j.phytol.2015.01.009

Vargas, L. H. G., Neto, J. C. R., de Aquino Ribeiro, J. A., Ricci-Silva, M. E., Souza, M. T., Rodrigues, C. M., et al. (2016). Metabolomics analysis of oil palm (*Elaeis guineensis*) leaf: evaluation of sample preparation steps using UHPLC-MS/MS. *Metabolomics* 12:153.

Veitch, H. J. (1979). An abridged history of *Nepenthes*. *Carniv. Plant Newsl.* 8, 20–23.

Wan Zakaria, W. N. A., Aizat, W. M., Goh, H. H., and Mohd Noor, N. (2019). Protein replenishment in pitcher fluids of *Nepenthes* × *ventrata* revealed by quantitative proteomics (SWATH-MS) informed by transcriptomics. *J. Plant Res.* 132, 681–694. doi: 10.1007/s10265-019-01130-w

Wang, L., Tao, D., Dong, S., Li, S., and Tian, Y. (2018). Contributions of lunate cells and wax crystals to the surface anisotropy of *Nepenthes* slippery zone. *R. Soc. Open Sci.* 5:180766. doi: 10.1098/rsos.180766

Wiart, C., Mogana, S., Khalifah, S., Mahan, M., Ismail, S., Buckle, M., et al. (2004). Antimicrobial screening of plants used for traditional medicine in the state of Perak, Peninsular Malaysia. *Fitoterapia* 75, 68–73. doi: 10.1016/j.fitote.2003.07.013

Widhalm, J. R., and Rhodes, D. (2016). Biosynthesis and molecular actions of specialized 1,4-naphthoquinone natural products produced by horticultural plants. *Hortic. Res.* 3:16046.

Xia, J., and Wishart, D. S. (2016). Using metaboanalyst 3.0 for comprehensive metabolomics data analysis. *Curr. Protoc. Bioinform.* 2016, 14.10.11–14.10.91.

Xiao, J. F., Zhou, B., and Ressom, H. W. (2012). Metabolite identification and quantitation in LC-MS/MS-based metabolomics. *TRAC Trend Anal. Chem.* 32, 1–14. doi: 10.1016/j.trac.2011.08.009

Xie, G. X., Ni, Y., Su, M. M., Zhang, Y. Y., Zhao, A. H., Gao, X. F., et al. (2008). Application of ultra-performance LC-TOF MS metabolite profiling techniques to the analysis of medicinal Panax herbs. *Metabolomics* 4, 248–260. doi: 10.1007/s11306-008-0115-5

Yu, O., and McGonigle, B. (2005). Metabolic engineering of isoflavone biosynthesis. *Adv. Agron.* 86, 147–190. doi: 10.1016/s0065-2113(05)86003-1

Yulita, K. S., and Mansur, M. (2012). The occurrence of hybrid in *Nepenthes hookeriana* Lindl. from Central Kalimantan can be detected by RAPD and ISSR markers. *Hayati J. Biosci.* 19:18. doi: 10.4308/hjb.19.1.18

Zhou, B., Xiao, J. F., Tuli, L., and Ressom, H. W. (2012). LC-MS-based metabolomics. *Mol. Biosyst.* 8, 470–481.

Zulkapli, M. M., Ab Ghani, N. S., Ting, T. Y., Aizat, W. M., and Goh, H.-H. (2021). Transcriptomic and proteomic analyses of *Nepenthes ampullaria* and *Nepenthes rafflesiana* reveal parental molecular expression in the pitchers of their hybrid, *Nepenthes* × *hookeriana*. *Front. Plant Sci.* 11:625507. doi: 10.3389/fpls.2020.625507

Zulkapli, M. M., Rosli, M. A. F., Salleh, F. I. M., Mohd Noor, N., Aizat, W. M., and Goh, H. H. (2017). Iso-Seq analysis of *Nepenthes ampullaria*, *Nepenthes rafflesiana* and *Nepenthes* × *hookeriana* for hybridisation study in pitcher plants. *Genom. Data* 12, 130–131. doi: 10.1016/j.gdata.2017.05.003

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fpls.2021.643137

Nepenthes × *ventrata* Transcriptome Profiling Reveals a Similarity Between the Evolutionary Origins of Carnivorous Traps and Floral Organs

Anna V. Shchennikova, Alexey V. Beletsky, Mikhail A. Filyushin, Maria A. Slugina, Eugeny V. Gruzdev, Andrey V. Mardanov, Elena Z. Kochieva and Nikolay V. Ravin*

Institute of Bioengineering, Research Center of Biotechnology, The Russian Academy of Sciences, Moscow, Russia

The emergence of the carnivory syndrome and traps in plants is one of the most intriguing questions in evolutionary biology. In the present study, we addressed it by comparative transcriptomics analysis of leaves and leaf-derived pitcher traps from a predatory plant *Nepenthes ventricosa* × *Nepenthes alata*. Pitchers were collected at three stages of development and a total of 12 transcriptomes were sequenced and assembled *de novo*. In comparison with leaves, pitchers at all developmental stages were found to be highly enriched with upregulated genes involved in stress response, specification of shoot apical meristem, biosynthesis of sucrose, wax/cutin, anthocyanins, and alkaloids, genes encoding digestive enzymes (proteases and oligosaccharide hydrolases), and flowering-related MADS-box genes. At the same time, photosynthesis-related genes in pitchers were transcriptionally downregulated. As the MADS-box genes are thought to be associated with the origin of flower organs from leaves, we suggest that *Nepenthes* species could have employed a similar pathway involving highly conserved MADS-domain transcription factors to develop a novel structure, pitcher-like trap, for capture and digestion of animal prey during the evolutionary transition to carnivory. The data obtained should clarify the molecular mechanisms of trap initiation and development and may contribute to solving the problem of its emergence in plants.

Keywords: plant carnivory, *Nepenthes*, pitcher, transcriptome analysis, MADS-box family transcription factors

OPEN ACCESS

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

Reviewed by:
Sara V. Good, University of Winnipeg, Canada
Alejandra Vázquez-Lobo, Universidad Autónoma del Estado de Morelos, Mexico

****Correspondence:***
Anna V. Shchennikova
shchennikova@yandex.ru

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

Received: *17 December 2020*
Accepted: *03 May 2021*
Published: *28 May 2021*

Citation:
Shchennikova AV, Beletsky AV, Filyushin MA, Slugina MA, Gruzdev EV, Mardanov AV, Kochieva EZ and Ravin NV (2021) Nepenthes × ventrata Transcriptome Profiling Reveals a Similarity Between the Evolutionary Origins of Carnivorous Traps and Floral Organs. Front. Plant Sci. 12:643137. doi: 10.3389/fpls.2021.643137

INTRODUCTION

Predator plants attract arthropods that serve both as pollinators for sexual reproduction and as prey captured and digested in traps (active or passive); traps represent modified leaves developed to support plant survival in the nutrient-deficient environment (Jürgens et al., 2012; Pavlovič and Saganová, 2015; Bartlett, 2017). It is considered that carnivory evolved independently nine times; in particular, passive pitfall traps have at least six autonomous origins in different orders of flowering plants (Givnish, 2015). To date, the list of green predators contains more than 650 species representing 19 genera, 12 families, and five orders of the flowering plants, both monocots and eudicots (Pavlovič and Saganová, 2015). The most diverse lineage of carnivorous plants, Caryophyllales, includes one of the most famous genera of pitcher plants, *Nepenthes*, which

comprises 160–180 species as well as numerous natural and cultivated hybrids, inhabiting humid and sunny lowlands or tropical mountains deficient in nitrogenous nutrients soils (Murphy et al., 2020). Because of the large number of *Nepenthes* species residing in diverse habitats, the genus has become a model for studying plant adaptive radiation and speciation (Thorogood et al., 2018).

Nepenthes species develop typical photosynthetic leaves whose tips form a tendril as an extension of the midrib, to which a pitcher-like pitfall trap is attached (Owen and Lennon, 1999). The trap has a leaf-like lid (operculum), which initially seals the growing pitcher until it is ripe and ready to capture the prey; it also prevents fluid dilution during rain in mature traps (Bauer et al., 2008). Once open, the pitcher attracts the prey with rim (peristome)-located nectaries forming a slippery zone covered with a thick wax layer to catch and prevent the escape of the prey which then falls into a lower digestive zone containing multicellular glands that secrete an acidic fluid saturated with hydrolytic enzymes (chitinases and proteases) and antifungal factors (Gaume and Forterre, 2007; Scholz et al., 2010; Pavlovič and Mithöfer, 2019). The main pitcher characteristics attracting the prey are color patterns (UV fluorescence/visible wavebands), sweet fragrance of the secreted extrafloral nectar, and high CO_2 levels (Kurup et al., 2013; Baby et al., 2017).

Most studies on *Nepenthes* focus on the morphophysiological and functional characteristics of the pitcher, including biochemical composition of the secreted fluid in the digestive zone and transcriptomic and proteomic changes in response to prey and during early pitcher opening (Gorb et al., 2004; Hatano and Hamada, 2008, 2012; Rottloff et al., 2016; Wan Zakaria et al., 2016, 2019; Baby et al., 2017; Zulkapli et al., 2017; Goh et al., 2020); however, the data regarding the evolutionary transition from the non-carnivorous to carnivorous status in Nepenthes species are scarce. Comparative transcriptomics of carnivorous and non-carnivorous leaves in *Cephalotus follicularis* identified genetic signatures associated with prey attraction, capture, digestion, and nutrient absorption underlying the developmental switch between carnivorous and non-carnivorous leaf state (Fukushima et al., 2017). Recent transcriptomic analysis of genes controlling pitcher development in *Nepenthes khasiana* has revealed that ASYMMETRIC LEAVES 1 (AS1) and ERECTA (ER) contribute to the formation of the tendril from the midrib, whereas REVOLUTA (REV) may be positively associated with the initiation and maturation of the pitcher (Dkhar and Pareek, 2019). The most intriguing question is identification of molecular pathways that directed the emergence of the carnivory syndrome, as there are examples of both convergent (across unrelated lineages) and divergent (within genera) evolution of the trap (Thorogood et al., 2018). The pitcher-like trap has diverse morphology adapted to trapping and digestion of various preys. This is especially characteristic for the species of the *Nepenthes* genus, which is suggested to have originated from a *Drosera*-like predecessor (Givnish, 2015; Murphy et al., 2020). A comparative analysis of the genome of three carnivorous Droseraceae species showed that the evolution of predation in this family could be facilitated by a whole-genome duplication (WGD) in their last common ancestor, followed by diversification of the multiplied genes, including recruitment of root-specific genes to developing traps, massive loss of genes involved in non-carnivorous nutrition, and expansion of gene families associated with the attracting, catching, digesting, and utilizing prey (Palfalvi et al., 2020). However, the molecular mechanism employed by carnivorous plants to develop such a novel structure is still unclear.

In this study, we aimed to identify transcriptional signatures of the transition from the leaf to the mature pitfall trap by performing comparative transcriptomics of *Nepenthes* × *ventrata* leaves and pitchers at three developmental stages. The results revealed activation of genes associated with defense response, shoot apical meristem (SAM) organization, anthocyanin biosynthesis, and flowering, including MADS-domain transcription factors (TFs), in pitchers. Considering that MADS-box genes are thought to have been involved in the evolution of plant flower organs, we suggest that the emergence of the pitcher structure was also due to highly conserved MADS-domain TFs, which usually determine the identity of floral meristems and organs in extant plants.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Nepenthes × *ventrata* Hort. ex Fleming (Nepenthaceae Dumort.) is a natural hybrid between two *Nepenthes* clade 2 species: *N. ventricosa* Blanco (Insignes clade) and *N. alata* Blanco (Graciiflora clade) endemic to the northern forests of the Philippines (Murphy et al., 2020). Plants were obtained from a nursery garden and grown in a greenhouse under controlled conditions (16/8-h light-dark cycle with light intensity of 150–200 μmol m^{-2} s^{-1}, 20/24°C night/day temperature cycle, and 75–90% humidity). Samples of mature leaves at the middle section and pitchers at three developmental stages: primordial pitcher (or early pitcher) (∼1 cm), unopened young pitcher (3–4 cm), and open mature pitcher (or late pitcher) (7–11 cm; the first day of opening; there are no insects inside) were collected from the upper part of plant; three biological replicates were analyzed.

RNA Isolation, Library Preparation, and Transcriptome Sequencing

Total RNA was extracted from 300 mg of each of the 12 tissue samples using a modified CTAB-based technique (Filyushin et al., 2019) to facilitate extraction from wax-coated *Nepenthes* leaves and pitchers and treated with the RNase-free DNase Set (Qiagen, United States). RNA was assessed for quality using BioSpectrometer (Eppendorf, United States) and quantity using the Qubit® RNA Assay Kit and Qubit® 2.0 Fluorometer (Life Technologies, United States). According to the RNA integrity number (RIN) measured using Agilent 2100 Bioanalyzer (Agilent Technologies Inc., United States), all 12 RNA samples were of good quality for gene expression analysis by RNA-seq (RIN ∼8.4–9.7).

mRNA libraries were constructed using a NEBNext® mRNA Library Prep Reagent Set for Illumina® according

to the manufacturer's instructions (New England BioLabs, United States). Twelve barcoded libraries were sequenced by MyGene Co. (Moscow, Russia) on Illumina HiSeq2500 (Illumina, United States), generating a total of 387,930,413 single-end reads (200 bp) and the data were verified by additional in-house sequencing on Illumina MiSeq according to the manufacturer's instructions (Illumina). In total, 23,277,185 paired-end reads (2 × 250 bp) were generated.

De novo Transcriptome Assembly, Identification of Protein Coding Regions, and Annotation

Before assembly, adapter sequences were removed from the reads using cutadapt v1.17 (Martin, 2011), low quality reads were trimmed using sickle v1.33[1], and paired overlapping reads were merged using FLASH v1.2.11. *De novo* transcriptome assembly was carried out using Trinity v2.6.5 (Haas et al., 2013) with default parameters, using all reads combined together.

Coding regions in the assembled transcripts were predicted by TransDecoder v 5.1.0[2] within the Trinity software and the putative transcripts were screened for homology to known protein-coding plant genes using Diamond v.0.9 homology search in NCBI-NR database, and the completeness of the resulting assembly was assessed using BUSCO v.5.0.0 (Seppey et al., 2019). Potential contamination was assessed by checking the taxon of the best homolog for each predicted protein using a diamond search in the Uniref90 database which contains most of publicly available protein sequences clustered with 90% identity.

Relative transcription levels of protein-coding genes were calculated by mapping clean reads on the assembled transcripts using Trinity scripts with RSEM (Li and Dewey, 2011).

Functional annotation of the assembled transcripts was performed using NCBI-NR[3], GO[4], KEGG[5], PlantTFDB[6], MapMan Database[7], and GeneMANIA[8] (Warde-Farley et al., 2010). GO and KEGG terms were assigned using Trinotate v.3.1.1[9] and KAAS annotation server (Moriya et al., 2007), respectively. GO and KEGG terms enrichment was analyzed with goseq v. 1.36.0 and clusterProfiler v.3.12.0, respectively, in R package (Young et al., 2010; Yu et al., 2012); terms with $p < 0.05$ were considered significantly enriched. UniProt data were used to characterize possible functions of translated products[10]. Volcano plots were produced using EnhancedVolcano v1.8.0 R package[11]. Venn diagrams were drawn using web server program[12].

[1]https://github.com/najoshi/sickle

[2]http://transdecoder.github.io

[3]https://www.ncbi.nlm.nih.gov/

[4]http://www.geneontology.org/

[5]http://www.genome.jp/kegg/

[6]http://planttfdb.cbi.pku.edu.cn/

[7]http://mapman.gabipd.org/web/guest/mapcave

[8]https://genemania.org/

[9]http://trinotate.github.io

[10]https://www.uniprot.org/

[11]https://github.com/kevinblighe/EnhancedVolcano

[12]http://bioinformatics.psb.ugent.be/webtools/Venn/

Differential expression was analyzed using Trinity scripts with the edgeR method; genes were considered differentially expressed if changes in the expression levels computed by EdgeR were not less than two-fold [$|\log_2(FC)| \geq 1$], and false discovery rate (FDR) was ≤0.05.

Phylogenetic analysis was performed using the Fast Minimum Evolution method in NCBI[13] and the Maximum Likelihood method based on the JTT matrix-based model in MEGA7.0 (Kumar et al., 2016).

Validation of RNA-Seq Data

Real-time quantitative PCR results for 13 selected genes were validated with RT-qPCR. First-strand cDNA was synthesized with the Reverse Transcription System (Promega, Madison, WI, United States) using an oligo-dT and quantified by fluorimetry. RT-qPCR was performed with SYBR Green and ROX RT-PCR mixture (Syntol, Moscow, Russia), 3 ng cDNA, and 10 μM gene-specific primers in a CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories, United States) at the following cycling conditions: initial denaturation at 95°C for 5 min, 40 cycles of denaturation at 95°C for 15 s and annealing/synthesis at 62°C for 50 s. As homologs of the two house-keeping genes, *Ubiquitin* (*UBQ*) and *Elongation Factor* (*ELF*) previously suggested for *N. khasiana* gene expression normalization (Dkhar and Pareek, 2019) did not show uniform expression, we normalized the data using two other reference genes, *UBQ-ligase Praja-2* and *Actin 7*. Experiments were carried out in three biological and three technical replicates and statistical analysis was performed using GraphPad Prism version 7.02 (GraphPad, San Diego, CA, United States[14]). $P \leq 0.05$ was considered to indicate significant difference.

Nucleotide Sequence Accession Numbers

The raw sequences obtained in this study have been deposited in the NCBI Sequence Read Archive under the accession numbers SRX5724495–SRX5724506.

RESULTS

Nepenthes × *ventrata* Transcriptomes

Tissue samples of mature leaves and pitchers at three stages of development–early pitcher, young pitcher, and late pitcher (**Figure 1**) were analyzed by RNA-seq.

A total of 387,930,413 single-end reads (200 bp) and 23,277,185 paired-end reads (2 × 250 bp) were generated (**Supplementary Table 1**).

The reads of the four samples were combined and assembled. Overall, 245,999 transcripts (with size from 201 to 14,825 bp, total length 234.2 Mbp, and N50 size 1,602 bp) were generated. The completeness of the assembly was verified for 1,614 marker genes of the Embryophyta taxon using BUSCO, which revealed

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

[14]https://www.graphpad.com/scientific-software/prism/

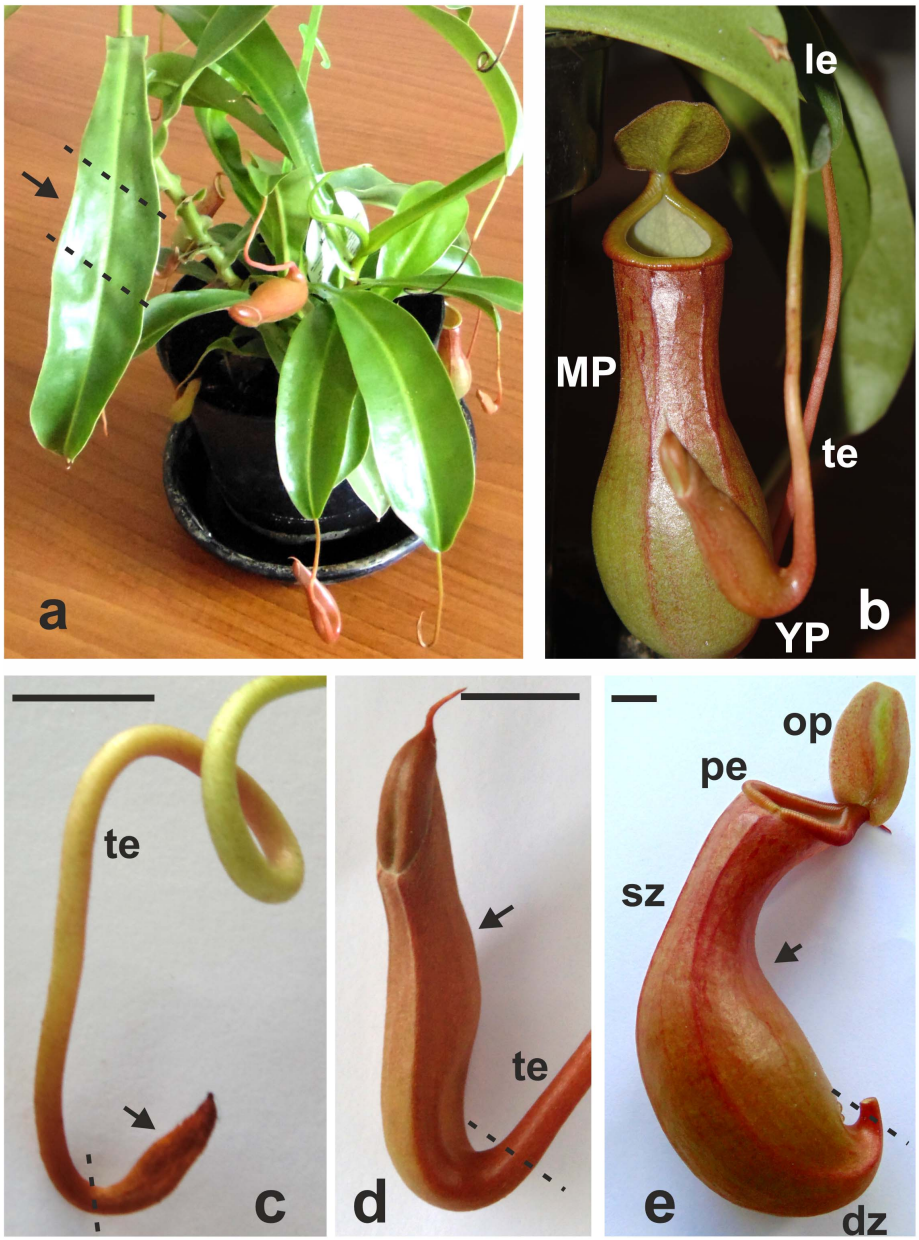


FIGURE 1 | *Nepenthes* × *ventrata* whole plant **(a)**, the leaf with the tendril and the trap **(b)**, and early (primordial) **(c)**, young **(d)**, and late (mature open) **(e)** pitchers. Arrows indicates the part of the organ used for transcriptomics. dz, digestive zone; sz, slippery zone; pe, peristome (rim); op, operculum (lid); te, tendril; le, leaf; MP, late pitcher; YP, young pitcher. Dashed lines indicate organ parts taken for analysis. Scale bar = 1 cm.

that 93.9 and 4.0% of the marker genes were complete and partial, respectively.

Annotation of the assembled transcripts using diamond search (cut-off *E*-value 10^{-5}) against NCBI non-redundant database of plant protein-coding genes revealed that among 85,558 protein-coding sequences, 62,037 had significant homology to non-hypothetical plant proteins (**Supplementary Table 2**). The 85,558 sequences were used for further analysis.

We analyzed the taxonomy of the best homolog for each putative protein; only 135 proteins showed better homology with bacterial proteins. The total expression of these bacterial transcripts ranged from 0.10 to 0.24% depending on the samples; seven and one transcripts were identified as belonging to fungi and insects, respectively. Of the 53,400 proteins that matched in the Uniref90 database, 53,065 had plant proteins as the best homologs, which indicated an extremely low level of contamination.

Analysis with the Trinity software identified many groups of transcripts that may belong to the same gene or its isoforms, alternative splice variants, and paralogs, which could be explained

by multiple paleopolyploidy events (at least seven genome duplications, according to Walker et al., 2017) in the evolutionary history of non-core Caryophyllales, including *Nepenthes*, as well as by the hybrid origin of *N.* × *ventrata*.

The *Nepenthes* × *ventrata* sequences showed the highest matches with *Vitis vinifera* (rosids; 12.38%), *Beta vulgaris* (11.12%), *Chenopodium quinoa* (10.77%), and *Spinacia oleracea* (Caryophyllales; 6.22%) sequences (**Supplementary Table 3**). According to the current taxonomy, *Nepenthes* (non-core Caryophyllales) should be closer to the asterid clade than to rosids (Yao et al., 2019); however, the top transcript homologs belonged primarily to rosids (60.52%), followed by Caryophyllales (28.17%) and asterids (6.36%) (**Supplementary Table 3**). This discrepancy may be attributed to a much higher number of assemblies for rosids (511) than for asterids and Caryophyllales (194 and 45, respectively) available in GenBank.

Gene Ontology Annotation

Gene ontology classification of the transcripts according to the function of the translated products revealed that 4,827 of them belonged to Biological Process (BP), 2,752–to Molecular Function (MF), and 1,020–to Cellular Component (CC) categories representing 97 sub-aspects (**Supplementary Tables 4**, **5**). The BP sub-aspects with the maximum number of transcripts were linked to stress response, cellular component organization, and biosynthetic processes (**Supplementary Table 4**). Most transcripts belonged to BP terms associated with general regulation of plant development, especially leaf development, senescence, and morphogenesis of leaves, and with photosynthesis, including carotenoid metabolism (**Supplementary Table 5**). The pitcher primordium, a red-colored organ with waxy inner walls, defines the efficiency of insect catching and processing; therefore, transcripts were classified in terms related to SAM development, biosynthesis of wax and anthocyanins, and plant-insect interactions. Of a particular interest are transcripts relevant to flowering initiation and formation of inflorescences and flowers, including reproductive organs, nectaries, pollen, ovules, and seeds (**Supplementary Table 5**).

The transcripts of the most enriched MF category were associated with utilization of ATP energy and the uptake and transfer of metal ions to specific locations, which is one of the basic requirements in photosynthetic plants (Yruela, 2013). The CC-classified transcripts were found to encode mainly nuclear, membrane, cytoplasmic, and plasma membrane proteins (**Supplementary Table 5**).

To obtain preliminary information about activation or suppression of various functions in leaves or pitchers, differentially expressed genes (DEGs) were identified based on the relative expression change of >2 times between the transcriptomes (**Supplementary Table 6**) and classified in GO terms (**Supplementary Table 7**). The most enriched BP/MF/CC terms in early and young pitchers vs. leaves were related to embryo development and active developmental and growth processes, whereas in late pitcher vs. leaves they were linked to plant-type cell wall biogenesis and organization, and in leaves vs. early/young/mature pitcher–to photosynthesis (**Supplementary Table 7** and **Supplementary Figure 1**). Late pitcher was similar to leaves in the upregulation of genes involved in oxidation-reduction and downregulation of those related to DNA replication. Late pitcher differed from all other transcriptomes by the downregulation of defense response genes. **Figure 2** shows transcriptional suppression of genes associated with photosynthesis and leaf developmental processes (early/young/mature pitcher), and induction of those involved in cell division, differentiation, and development of the growing (early/young pitcher) pitcher.

Kyoto Encyclopedia of Genes and Genomes Annotation

Kyoto Encyclopedia of Genes and Genomes classification of transcripts in terms of KEGG PATHWAY (KP) and KEGG BRITE (KB; protein families) revealed 26,235 significant matches in the KP database, which were assigned to seven main classes (**Supplementary Figure 2a**), 45 sub-categories, and 401 pathways (**Supplementary Table 8**). Among the 26,235 matches, 15,204 were presented as 26 hierarchical classifications (incorporating many different types of relationships) in three sub-categories: "Genetic Information Processing," "Signaling and Cellular Processes," and "Protein families: Metabolism" (**Supplementary Figure 2b** and **Supplementary Table 8**). The presence of multi-transcript KP sub-categories "Environmental adaptation" and "Immune system" (**Supplementary Figure 2c**), as well as pathways "Plant hormone signal transduction," "Plant-pathogen interaction," and "MAPK signaling pathway–plant" (**Supplementary Table 8**) may reflect defense mechanisms and adaptation to nutrient deficiency in the carnivorous plant.

It should be mentioned that the plant data used by KEGG do not include characteristics of the carnivorous syndrome. Digestive fluid in *Nepenthes* pitcher contains proteases (such as nepenthesin), chitinases, glucanases, phosphatases, proteases, and pathogenesis-related proteins (Fukushima et al., 2017; Ravee et al., 2018; Pavlovič and Mithöfer, 2019). In view of this, we briefly reviewed KP "Digestive system" and found that all our genes of interest were absent there and scattered throughout different other pathways such as "Amino sugar and nucleotide sugar metabolism," "MAPK signaling pathway–plant," "Plant hormone signal transduction," "Plant-pathogen interaction," etc. (**Supplementary Table 8**). Venn diagrams from KEGG gene ontology were used to visualize which tissues have similar gene expression and which are the most differentiated. Comparison of the number of activated and suppressed transcripts in KEGG categories showed the greatest similarity between early and young pitchers (**Figures 3A,B**).

Comparative analysis of KEGG terms enriched (by ≥2 times upregulated transcripts, **Supplementary Table 6**) in each transcriptome revealed that in leaves (vs. early pitcher), they were mostly associated with aging and primary and secondary metabolism, whereas in early and young pitcher (vs. leaves)–with genetic information processing, cell growth and death, development and regeneration, immune system, and

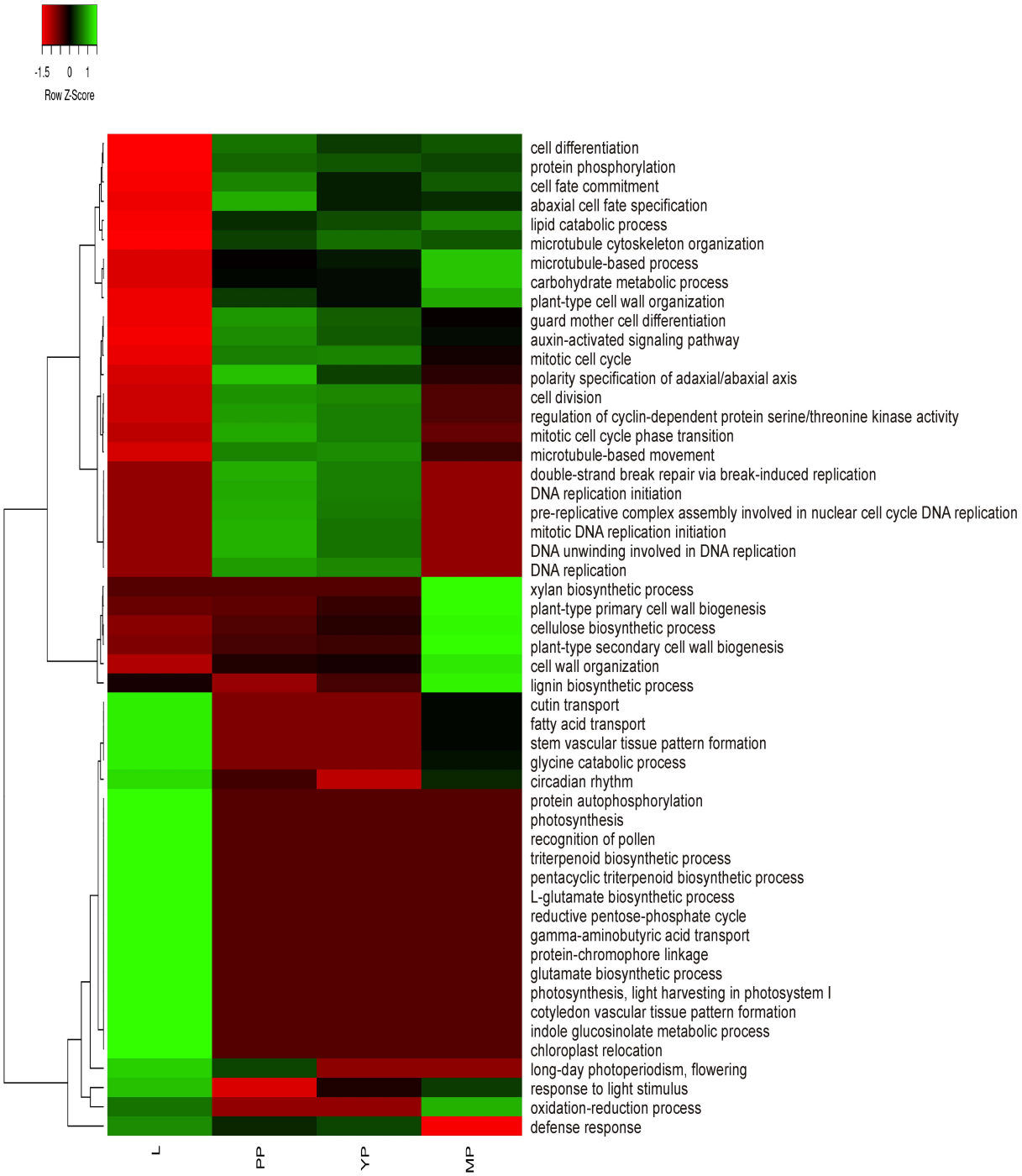


FIGURE 2 | Heatmap based on relative number of ≥2 times upregulated transcripts in gene ontology (GO) terms in leaves (L) vs. early pitcher (PP) and in PP, young pitcher (YP), and late pitcher (MP) vs. leaves.

signal transduction (**Supplementary Table 9**). Late pitcher was similar to leaves in the downregulation of genetic information processing, cell growth and development, immune system, and signal transduction, and to early/young pitcher–in the suppression of genes assigned to primary and secondary metabolism, and activation of transport and catabolism, immune system, and signaling (**Supplementary Table 9** and **Figure 3**). Compared to the other transcriptomes, late pitcher showed transcriptional induction in the categories of "Digestive system" and "Biosynthesis of other secondary metabolites," and reduction in the "Endocrine system" and "Drug resistance" (**Supplementary Table 9**

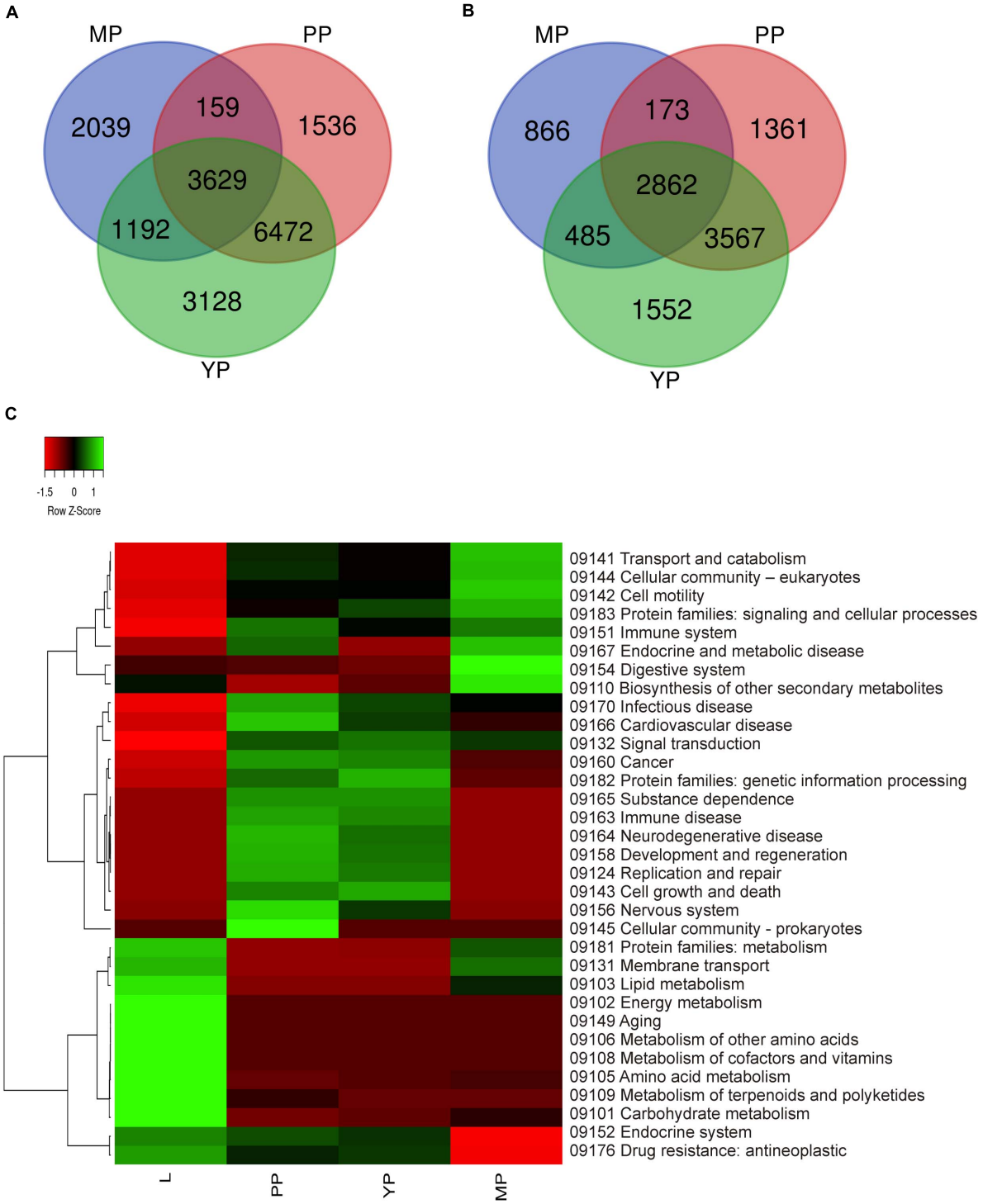


FIGURE 3 | Comparison of differential expression of leaf and pitcher transcripts in Kyoto encyclopedia of genes and genomes (KEGG) terms. Number of transcripts upregulated **(A)** and downregulated **(B)** in early pitcher (PP), young pitcher (YP), and late pitcher (MP) vs. leaves (L) (**Supplementary Table 8**). Heatmap **(C)** based on relative expression of transcripts associated with KEGG terms (**Supplementary Table 9**) in leaves vs. early pitcher, and in PP, YP, and MP vs. leaves.

and **Figure 3**); the latter is possibly correlated with the downregulation of "Defense response" in GO enrichment analysis (**Figure 2**).

Plant Transcription Factor Database Annotation

Nepenthes transcripts encoding TFs were additionally characterized using plant-specific PlantTFDB. In total, 3,298 TF-encoding transcripts were classified into 48 families, among which the most represented (≥20 members) were MYB/MYB-related, bHLH, AP2/ERF, C2H2, C3H, bZIP, NAC, WRKY, G2-like, HD-ZIP, B3/ARF, GRAS, Dof, and MADS-domain protein families (**Supplementary Table 10**). Volcano plots were used to visualize the differences in expression levels between the transcripts, including TF-coding transcripts (**Figure 4**). Analysis of TF differential expression revealed similarity of the regulatory machinery between leaves and late pitcher and between early and young pitchers (**Figure 4** and **Supplementary Table 10**).

MapMan Annotation

Next, we used MapMan to facilitate comparison of upregulated transcripts (**Supplementary Table 6**) in early, young, and late pitchers vs. leaves, late and young pitchers vs. early pitcher, and late pitcher vs. young pitchers. Secondary metabolism patterns were similar between early pitcher and young pitchers and between leaves and late pitcher. Compared to leaves and late pitcher, early pitcher and young pitchers were found to have several secondary metabolism pathways upregulated, including biosynthesis of phenylpropanoids, anthocyanins, and alkaloids (**Figure 5** and **Supplementary Figure 3**). Compared to leaves, photosynthesis was sharply reduced in pitchers at the early developmental stages (early and young pitchers) (**Figure 5**) but not at the late stage (mature pitcher) (**Supplementary Figure 4**). Genes related to biotic and abiotic stress responses were upregulated in pitchers at all developmental stages compared to leaves (**Figure 5** and **Supplementary Figure 5**).

Overall, these results indicate a loss of leaf identity, i.e., photosynthetic activity, in the pitcher, and the upregulation of defense-related and metabolic processes during its development.

Differentially Expressed Genes in Pitcher Initiation, Growth, and Maturation

Next, we selected DEGs upregulated >4 times and annotated them, using *Arabidopsis thaliana* UniProt data for the pair-wise comparisons of early, young or late pitchers with leaves, and leaves with early pitcher (**Supplementary Table 11**). The results revealed significant upregulation of transcripts related to flowering initiation and flower development, biosynthesis of secondary metabolites such as anthocyanins, terpenes, jasmonate, and wax, as well as stomatal development and movement, water transport, immunity, and stress response (**Table 1**).

In accordance with pitcher initiation on the tendril apical meristem, early pitcher differed from leaves in the activation of processes associated with the initiation and organization of embryonic SAM, and determination of organ abaxial–adaxial polarity (**Table 1** and **Supplementary Table 11**).

Initiation of the pitcher primordium was accompanied by transcriptional activation of genes associated with the redox state, abiotic stress, respiration, and signaling as well as TFs and R-genes (**Figure 5** and **Supplementary Figures 3–5**). Early pitcher was similar to young pitcher but both differed from late pitcher in the expression pattern of genes regulating the redox state and induction of signaling, respiration, and TF activity (especially ethylene-responsive factors) (**Supplementary Figure 5** and **Supplementary Table 11**). The upregulated TFs were: ERFs (ERF1A, RAP2-1, ERF012, ERF061, and WRINKLED 1), ANAC081, ANAC022, ANAC042, MYB2, MYB73, MYB88, ATAF2, ZAT5, ZAT9, HSF24, and WRKY 70, etc. (leaves vs. early pitcher); ERFs (AIL1, AIL5, CRF2, CRF4, ESR2, TINY 2, WIN1, and ERF003), heat shock factors (HsfA2, A3, A7, B2, and B4), ANAC083, MYB88, bHLH98, etc. (early pitcher vs. leaves); ERFs (TINY 2, ERF023, ERF115, SHINE 2, WIN1, and ERF003), bZIP36, MYB88, and ZAT9, etc. (mature pitcher vs. leaves) (**Supplementary Table 11**).

Furthermore, initiated and developing pitchers showed upregulation of stress-response factors encoding pathogenesis-related proteins (PR-1C, Fra a 1.03, PRHA, STH-2, STH-21, PTI5, PRHA, SNI1, and thaumatin-like), LRR receptor-like serine/threonine protein kinases, digestive enzymes (peroxidases, glucanases, alfa- and beta-amylases, phospholipases, chitinases, and aspartic proteinases nepenthesin-1 and -2), ABC transporters, dehydration-responsive elements, heat shock proteins, and jasmonate signaling-related molecules (**Supplementary Tables 2**, **6**, **11**).

Pitchers at different developmental stages also showed upregulation (>2–4 times) of genes related to stomatal and guard cell development and functioning, including those involved in water and CO_2 transport across cell membranes, such as aquaporins (NIP5-1, PIP1-2, PIP2-2, PIP2-7, NIP1-2, and TIP1-2 in late pitcher; PIP1-1 and SIP1-2 in early pitcher and YP; PIP1-3 and PIP1-4 in early and young pitchers; and PIP2-8, SIP1-1, TIP1-1, TIP1-3, and TIP4-1 in pitchers of all stages) and aquaporin level regulators (E3 ubiquitin-protein ligase RMA1H1 in late pitcher), and reversible hydration of CO_2 (carbonic anhydrase in leaves, early pitcher, and late pitcher), as well as genes encoding formate dehydrogenase and ribulose 1,5-bisphosphate carboxylase (in late pitcher) (**Supplementary Tables 2**, **6**, **11**).

Although some genes associated with stomatal patterning were upregulated in leaves (such as those encoding calcium sensing receptor, guard cell S-type anion channel SLAC1, and TFs MYB88, bHLH97, and WRKY70), many more of such genes were induced in pitchers, including microtubule-associated protein 1, ABA receptor PYL9, alkaline ceramidase, CO_2-response secreted protease, respiratory burst oxidase homolog protein F, EPIDERMAL PATTERNING FACTOR-like protein 3, ER, ERL1, STOMATAL DENSITY AND DISTRIBUTION 1, and TFs bHLH93, bHLH1, bHLH98, MYB61, MYB66, MYB88, and others (**Supplementary Table 11**).

Pitchers were also characterized by activation of the terpenoid pathway. Many relevant genes were upregulated

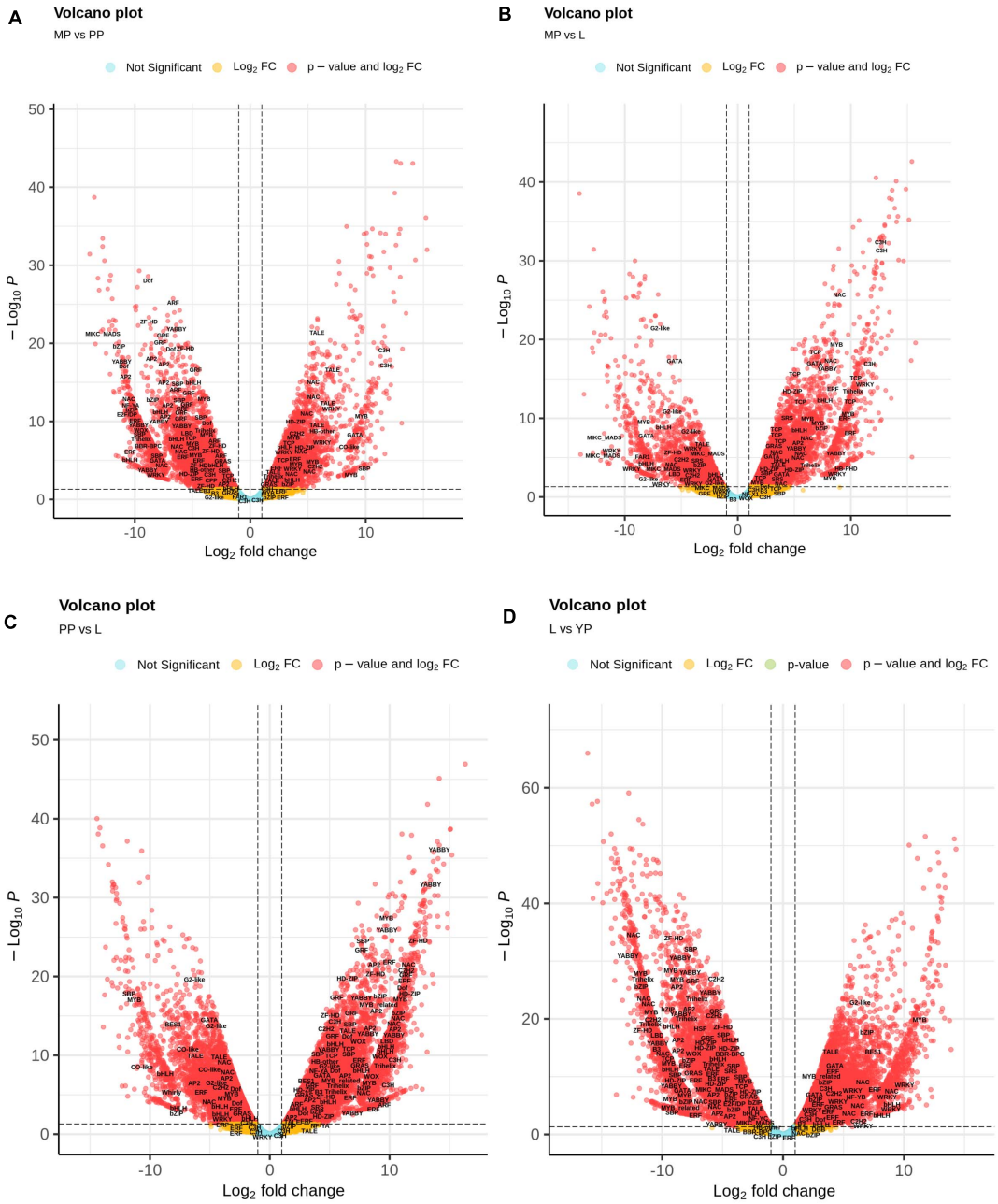


FIGURE 4 | Volcano map analysis of DE transcripts in MP vs. PP **(A)**, in MP vs. L **(B)**, in PP vs. L **(C)**, and in L vs. YP **(D)**. TF-coding DETs are indicated.

in the primordium, including those encoding enzymes involved in terpenoid biosynthesis, such as isopentenyl phosphate kinase, (E,E)-alpha-farnesene synthase, terpene synthase 8, GDSL lipase, tabersonine-19-hydroxy-*O*-acetyltransferase, and sesquiterpene synthase STPS (**Supplementary Table 11**). In MPs with developed glandular tissues, the upregulated transcripts encoded TFs (WRKY 72A and EXPRESSION OF TERPENOIDS 1) and enzymes involved in biosynthesis of germacrene D, germacrene-derived sesquiterpene lactones, triterpene saponins and phytosterols, and monoterpenoid indole alkaloids (**Supplementary Table 11**).

Among the TFs upregulated in early pitcher vs. leaves by more than two times there were multiple homologs of many flowering-related MADS-domain TFs (**Table 1** and **Supplementary Tables 10**, **11**). We performed phylogenetic analysis of the found 45 proteins in comparison with the known MADS-TFs of a model plant *A. thaliana* and a representative of Caryophyllales, *B. vulgaris*; in addition, each of the 45 proteins was separately

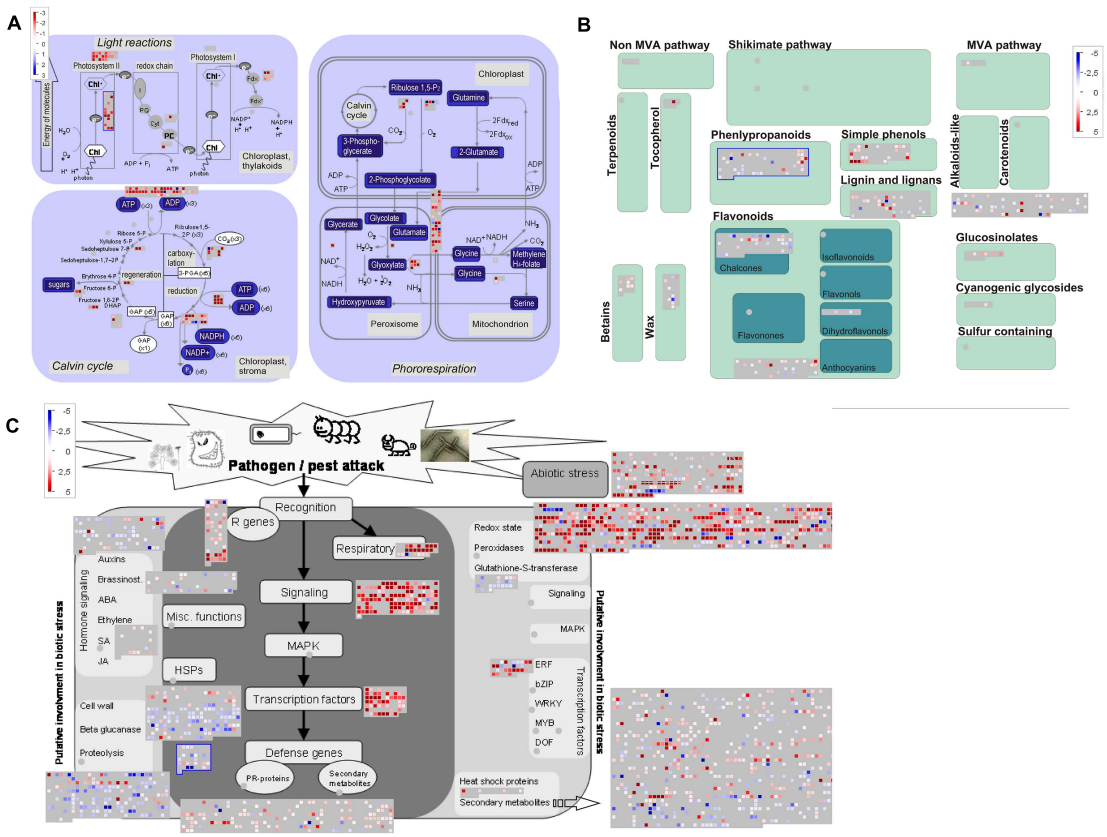


FIGURE 5 | MapMan-based scheme showing differential expression of transcripts associated with photosynthesis **(A)**, secondary metabolism **(B)**, and stress responses **(C)** in PP compared to leaves. The activity of each pathway is presented as blocks, in which cubes represent individual transcripts and their colors reflect the difference in expression levels between the two transcriptomes. Transcripts upregulated in PP vs. leaves are indicated by blue in panel **(A)** and red in panel **(B,C)**.

analyzed with the most similar sequences available from the NCBI GenBank. As a result, the *NveMADS1-45* genes were classified as homologs of *AGL24*, *AGL42*, *SVP*, *SOC1*, *AP1*/*FUL*, *SEP1*/*SEP2*, *SEP3*, *SEP4*, *AG*, *STK*, *AP3*, *AGL6*, *AGL62*, *AGL65*, *AGL104*, *AGL80*, and *AGL82* (**Supplementary Table 12** and **Supplementary Figure 6**).

To verify the presence of the MADS-box transcripts in non-flowering *N.* × *ventrata* tissues (developing pitcher), we searched available transcriptome data on *N.* × *ventrata* late pitcher (24 h after opening) (Wan Zakaria et al., 2016) (NCBI TSA: *Nepenthes ventricosa* × *Nepenthes alata*, transcriptome shotgun assembly; GFAD00000000.1) and male inflorescence of another relative species *N. khasiana* (Scharmann et al., 2019). As a result, we found homologs of *AGL24*, *AGL42*, *SVP*, *SOC1*, *FUL*, *AP1*, *SEP1*, *SEP2*, *SEP3*, *SEP4*, *AG*, *STK*, *AP3*, *AGL6*, *AGL14*, *AGL16*, *AGL19*, *AGL37*, *AGL61*, *AGL66*, *AGL80*, and *AGL104* genes in the *N.* × *ventrata* pitcher transcriptome, which confirms the reliability of the data obtained in this study. The *N. khasiana* male inflorescence transcriptome contained homologs of *SOC1*, *SVP*, *AP1*, *SEP*, *JOINTLESS*, *AG*, B-class genes, *AGL29*, *AGL42*, *AGL61*, *AGL62*, *AGL65*, *AGL80*, *AGL82*, *AGL104*, and *MADS6* (**Supplementary Table 13**).

The identified *N.* × *ventrata* MADS-box genes are known to be involved in plant reproductive development and are expressed in *Arabidopsis* reproductive organs (except for *SVP*) (Smaczniak et al., 2012; Theißen et al., 2016). *AGL24*, *AGL42*, *SVP*, and *SOC1* are also expressed in *Arabidopsis* roots (Parenicová et al., 2003). In addition, *AGL24*, *AGL42*, *SVP*, *SOC1*, *SEP1*, *SEP4*, *AGL6*, *AGL62*, *AGL65*, and *AGL80* are expressed in the vegetative parts of the plant; among them, *AGL24*, *AGL42*, *SVP*, and *SOC1* are associated with flowering initiation (Parenicová et al., 2003; Dorca-Fornell et al., 2011; Hu et al., 2014; Valentim et al., 2015). Thus, the presence of the listed MADS-box genes (like *AP1*, *SEP3*, *AG*, *AGL11*, and *AP3*) in supposedly vegetative *N.* × *ventrata* pitchers and leaves is unusual and deserves special attention.

According to GeneMANIA prediction, *N.* × *ventrata* homologs of type I proteins (AGL65 and AGL104) can participate in gametophyte development, whereas homologs of type II proteins–in the regulation of shoot system development (FUL, AGL6, AGL42, SOC1, and SVP), positive (FUL, AGL6, AGL42, and SOC1) and negative (SVP) regulation of reproductive development, regulation of floral and post-embryonic organ development (AGL11, AGL42, AG, SEP3, and AGL6), and meristem identity and stem cell development and differentiation (FUL, AGL24, AP1, SOC1, and SVP) (**Supplementary Tables 14**, **15**).

Heatmap shows activation of most MADS-box genes during pitcher initiation and development (**Figure 6A**). In early

TABLE 1 | Selected differentially expressed genes (DEGs) upregulated > 4 times in leaves (vs. early pitcher), and early (PP), young (YP), or late (MP) pitchers vs. leaves.

Proteins encoded by DEGs (function–according to UniProt)	L	PP	YP	MP
Negative regulation of vegetative-to-reproductive phase transition and flowering				
TFIID subunit 14b (GAS 41-like)	↑			↑
EARLY BOLTING IN SHORT DAYS	↑	↑		
SPLAYED; RED AND FAR-RED INSENSITIVE 2	↑			
MADS-box protein SHORT VEGETATIVE PHASE (SVP)		↑	↑	↑
EARLY FLOWERING 1; EARLY FLOWERING MYB; and Homeobox protein ATH1		↑	↑	
Prevention of the early vegetative-to-reproductive phase transition				
CHROMATIN REMODELING 11			↑	
Homeobox-DDT domain protein RINGLET 3		↑		
Homeobox-DDT domain protein RINGLET 1	↑			
Flowering promotion				
AG-like MADS-box protein AGL19 (SOC1-like) and VASCULAR PLANT ONE-ZINC FINGER 1	↑			
AGL24-like; AtHB-31; VERNALIZATION 5; APETALA 2; FY; and FD		↑	↑	
AGL42-like; Flowering locus K; and FCA		↑		
Flower, fruit, and seed development				
AGAMOUS-like (AG)		↑	↑	↑
SEPALLATA 2-like (SEP2) and DEFICIENS-like (DEF, AP3)		↑		↑
FRUITFULL-like (FUL)/CAULIFLOWER (CAL, AP1)				↑
AGL6-like and AGL104-like		↑		
AG-like SEP1-like; SEP3-like; AGL61-like (DIANA); AGL37-like (PHERES1); AGL80-like; CMB1-like; JOINTLESS-like (J) AINTEGUMENTA (ANT); Auxin response factor 8 (FRUIT WITHOUT FERTILIZATION); and SPATULA (bHLH024)		↑	↑	
AGL65-like and AGL62-like			↑	
AGL11-like (SEEDSTICK and STK)	↑			
Anthocyanin biosynthesis and accumulation				
MYB1 (MYB113) and Anthocyanidin 5,3-*O*-glucosyltransferase (EC 2.4.1.)			↑	↑
LNK1 and Flavone 3'-*O*-methyltransferase 1 (EC 2.1.1.42)				↑
MYB3; BHLH42; GLABRA 3 (bHLH 001); ANTHOCYANINLESS 2; Flavonol synthase/flavanone 3-hydroxylase (EC 1.14.11.9) (EC 1.14.20.6) (FLS) and Anthocyanidin reductase [(2S)-flavan-3-ol-forming] (ANR) (EC 1.3.1.112)		↑	↑	
Chalcone-flavonone isomerase 3 (EC 5.5.1.6) (CHI)		↑		
Flavonoid 3',5'-hydroxylase CYP75B138 (EC 1.14.14.81) (F3'5'H)	↑	↑	↑	↑
2-oxoglutarate-dependent dioxygenase At5g05600 (EC 1.14.11.) (ANS)	↑	↑	↑	
Chalcone synthase 2 (EC 2.3.1.74) (CHS); Naringenin,2-oxoglutarate 3-dioxygenase (EC 1.14.11.9) (FHT) (Flavanone-3-hydroxylase) (F3H); 2-oxoglutarate-dependent dioxygenase ANS (EC 1.14.11.-) (Anthocyanidin synthase); Dihydroflavonol 4-reductase (DFR) (EC 1.1.1.219) / (Flavanone 4-reductase) (FNR) (EC 1.1.1.234); Flavonoid 3',5'-hydroxylase 1 (F3'5'H) (EC 1.14.14.81); Flavonol 3-*O*-glucosyltransferase UGT89B1 (EC 2.4.1.91); Leucoanthocyanidin reductase (EC 1.17.1.3); and Anthocyanidin 3-*O*-glucosyltransferase (EC 2.4.1.115) (UFGT)		↑	↑	↑
Digestive enzymes				
Aspartic proteinase nepenthesin-2 (EC 3.4.23.12)	↑			↑
Aspartic proteinases nepenthesin-1 (EC 3.4.23.-) and PCS1; Chitinase-like protein 1 (AtCTL1)		↑	↑	↑
Aspartyl protease family protein At5g10770 and Cysteine protease RD21A			↑	↑
Aspartic proteinase A2 (Aspartic protease 57); Aspartic proteinase-like protein 2; and Cysteine proteases RD19A, RD19B (EC 3.4.22.-)				↑
Organization of SAM during embryogenesis and organ separation				
WUSCHEL-related homeobox (WOX) 1; WOX2; WOX3; WOX4; YABBY1; ASYMMETRIC LEAVES 2 (AS2); MAINTENANCE OF MERISTEMS; FASCIATA1; FASCIATA2; TORNADO2; CUP-SHAPED COTYLEDON 2 (CUC2); CUC3; TEBICHI; PHABULOSA (PHB, ATHB14); KANADI 2 (KAN2); and GROWTH-REGULATING FACTOR (GRF) 1–4, 6–9		↑	↑	
AS2-like proteins 1 and 38; WOX9; and DEK1		↑		
WOX8 and WOX13				↑
YABBY4			↑	
YABBY2; YABBY5; and ASYMMETRIC LEAVES 1 (PHANTASTICA)		↑	↑	↑
REVOLUTA (REV); ERECTA (ER); ER-like kinase 1 (ERL1); TONSOKU; SEMI-ROLLED LEAF 2 (SRL2); and GRF8		↑	↑	↑

and young pitchers vs. leaves/late pitcher, the *NveMADS11* (AP1/FUL clade), *NveMADS19* and *NveMADS45* (AGL80), *NveMADS31* (SVP), *NveMADS36* (AGL24), and *NveMADS39*, *43,* and *44* (SOC1) genes were downregulated, whereas the remaining genes were upregulated (**Figure 6A**). In early pitcher vs. leaves, transcript isoforms homologous to *SOC1* (*NveMADS39*, *43*, *44* / *NveMADS37*, *38*, *40*, *41*, *42*), *SVP* (*NveMADS29-30, 32, 33* / *NveMADS31*, *36*), and *AP1*/*FUL* (*NveMADS10, 12, 13* / *NveMADS11*) were, respectively, up-/downregulated (**Figure 6A**).

Consistent with the red color acquired by developing pitchers, we observed activation of anthocyanin biosynthesis in early and young pitchers, which is evidenced by the upregulation of many relevant enzymes and TFs (**Table 1** and **Supplementary Tables 10**, **11**). Compared to leaves, all pitcher stages showed activation of most of the analyzed genes for anthocyanin biosynthesis with a maximum transcription level in young pitcher; one exception was *MYB113* whose expression was constant in leaves, early pitcher, and young pitcher, but increased sharply in late pitcher (**Figure 6B**). Furthermore, late pitcher showed a two-fold upregulation of the gene encoding 4,5-DOPA-extradiol dioxygenase (**Supplementary Table 6**), one of the main enzymes involved in the biosynthesis of betalains–a class of red and yellow pigments (Timoneda et al., 2019).

Regarding other important pigments, carotenoids, it should be mentioned that very few genes of the carotenoid pathway, associated with fruit-specific sequestration of pigment in chromoplasts and abscisic acid synthesis, rather than photosynthesis and photoprotection, were found upregulated: chromoplast-specific carotenoid-associated protein CHRC (in leaves and late pitcher vs. early pitcher), 9-*cis*-epoxycarotenoid dioxygenase NCED1 (in leaves vs. late pitcher and in early pitcher vs. leaves), and carotenoid-cleavage dioxygenase NCED4 (in late pitcher vs. leaves and early pitcher) (**Supplementary Table 11**).

Validation of RNA-Seq Data by Real-Time Quantitative PCR

Thirteen DEGs were selected for RT-qPCR validation of the transcriptomic data with specific primers (**Supplementary Table 16**): 10 genes of the anthocyanin biosynthetic pathway (regulatory *MYB113* and *bHLH001* and structural *CHS1*, *CHS2, CHI*, *F3H*, *F3'5'H*, *DFR*, *ANS*, and *UFGT*) and three TF genes associated with flowering time and floral organ identity (*MYB17* and MADS-box genes *AG* and *SEP3*). The results showed significant upregulation of all analyzed genes in early pitcher compared to leaves, with the exception of *MYB113* activated in late pitcher (**Supplementary Figure 7**), which was consistent with the transcriptome data (**Figure 6**).

DISCUSSION

The *Nepenthes* pitcher, derived from the tip of the leaf tendril, is thought to emerge from a spontaneously occurred epiascidiate leaf folded into a tubular structure (outer adaxial side inward) with fused margins, which provided selective advantage through improved storage of water and nutrients (Givnish, 2015). To survive in nitrogen-deficient conditions, *Nepenthes* (like other known carnivorous species with similar trap structures) turned the tubular leaf into a trap-pitcher with prey attraction and digestion abilities (Givnish, 2015).

Stress-Response Proteins Trigger Trap Initiation

Comparative genomic and transcriptomic studies of true leaves and leaf traps in different carnivorous species, including *Nepenthes*, have shown that signaling pathways involved in prey catching and digestion are similar to those providing protection against pathogens in non-predatory plants (Renner and Specht, 2013; Fukushima et al., 2017). For example, jasmonates associated with response to stress, microbial pathogens, and pests in non-carnivorous plants, in carnivores can perceive signals from captured prey and participate in the development of digestive cavities loaded with hydrolytic enzymes (Pavlovič and Mithöfer, 2019), including pathogenesis-related proteins associated with stress response, which acquired digestive properties (Pavlovič and Mithöfer, 2019). These data suggest that carnivorous plants evolved using pest defense mechanisms (Fukushima et al., 2017; Pavlovič and Mithöfer, 2019). The results of this study revealed that genes related to stress response and pathogen/pest attacks were upregulated in the early pitcher transcriptome (**Supplementary Tables 6**, **11**, **Figure 5**, and **Supplementary Figures 3–5**), which is consistent with the hypothesis of defense-to-carnivory pitcher evolution.

Pitcher Primordium Organization

In response to defense-related signals, the central vein of the leaf lengthens, producing a tendril with an apical meristem at the tip, which forms the pitcher in the process directed by phyllotaxis signals that promote cell division activity and prevent premature cell differentiation. We found that transcription of the *WOX* genes was upregulated in pitchers compared to leaves (**Table 1**, **Supplementary Table 11**, and **Figure 4**), which may be related to lateral organ formation (*WOX1* and *WOX3*), embryo patterning (*WOX8* and *WOX9*), floral transition (*WOX13*), and determination of vascular stem cell niche (*WOX4*) (van der Graaff et al., 2009; Zhou et al., 2015).

Consistent with cell division patterns in adaxial tissues shown in carnivorous *Sarracenia purpurea* (Fukushima et al., 2015), in *N.* × *ventrata* early pitcher the genes regulating the abaxial–adaxial polarity and initiation/organization of embryonic SAM were activated (**Table 1**, **Supplementary Table 11**, and **Figure 4**), including abaxial–adaxial polarity genes shown to be upregulated in pitcher-bearing shoots of carnivorous *Cephalotus* (Fukushima et al., 2017), as well as *AS1* and *REV* genes activated at the site of *N. khasiana* trap initiation and formation (Dkhar and Pareek, 2019). *KAN* and *PHB* genes upregulated in the *N.* × *ventrata* pitcher may determine its initial asymmetric growth and *YABBY* genes may specify the fate of abaxial cells, thus contributing to the proper abaxial/adaxial organization of the trap as shown in *Arabidopsis* (Eshed et al., 2004). Also similar to *Arabidopsis* (Inagaki et al., 2006; Rast and Simon, 2012; Shpak, 2013; Amanda et al., 2016), the longitudinal growth of *N.* × *ventrata* trap primordium may depend on the mutual activity of the upregulated *ER*, *ERL1*, and *ERL2* genes, whereas cell division and differentiation (including epidermal cells) may rely on *TEBICHI* and *DEK1*, and symmetry–on *AS2*. *GRF* genes are known to be involved in the control of cell expansion in the leaf (Dkhar and Pareek, 2014) and, hence, may be important for *N.* × *ventrata* trap outgrowth at the tip of the leaf.

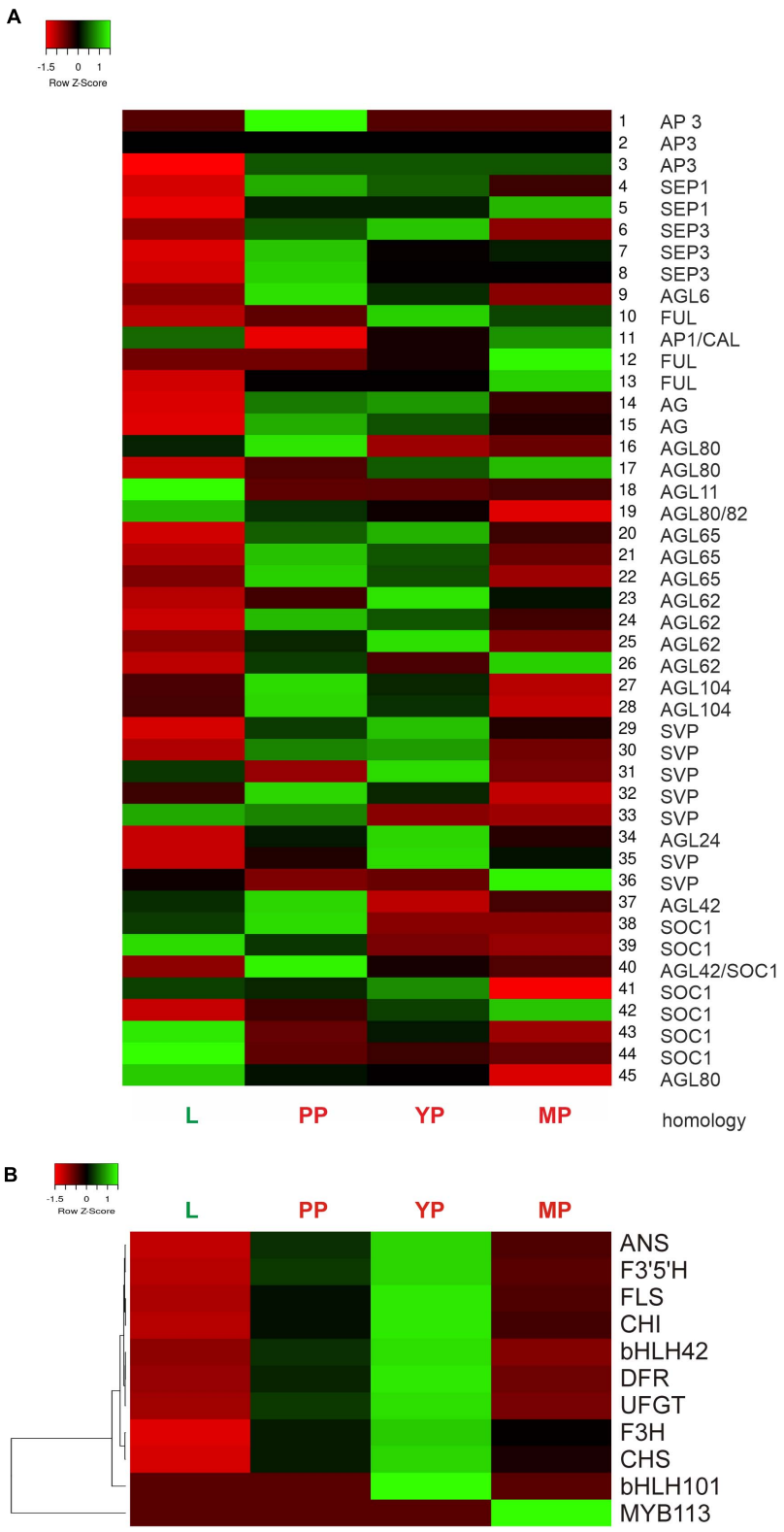


FIGURE 6 | Heatmap of relative expression of the MADS-box genes *NveMADS1–45* **(A)** and genes involved in anthocyanin biosynthesis **(B)**. L vs. PP, and PP, YP and MP vs. leaves.

Folding of the tubular trap may require *SRL2*, which regulate leaf rolling through abaxial cell differentiation (Liu et al., 2016), as well as *bHLH30* involved in upward leaf curling (An et al., 2014) and *FIL* and *YABBY2* related to the twisted leaf phenotype (Bowman, 2000).

Overall, the activation and cooperative activity of these genes in *N.* × *ventrata* early pitcher may contribute to the epiascidiate leaf folding.

Prey Capture

The pitcher of *N. alata* (an *N.* × *ventrata* parent) catches the prey with the help of anisotropic slippery, highly wet peristome (with water and/or nectar films), waxy inner walls, and release of attractants and sugar-rich nectar (Bohn and Federle, 2004; Bauer et al., 2008). The captured prey is then digested in the lower zone fluid enriched with hydrolytic enzymes (Bauer et al., 2008). Consistent with these data, in *N.* × *ventrata* pitcher transcriptomes we observed activation of genes involved in sucrose, wax/cutin, and alkaloid biosynthesis, as well as those encoding proteases and oligosaccharide hydrolases (**Supplementary Table 11** and **Figure 5**).

To attract insects with well-developed CO_2 receptors, closed *Nepenthes* pitchers accumulate CO_2 emitted after opening, which decreases pitcher photosynthetic capacity, while promoting its growth and respiration and increasing carbohydrate synthesis, cuticular wax density, and humidity levels (Baby et al., 2017). Moreover, CO_2 dissolved in the digestive fluid maintains the optimum pH for the activity of hydrolytic enzymes and nutrient absorbance (Baby et al., 2017). Because of the high CO_2 concentrations, "modified stomata" evolved on the inner side wax layers of *Nepenthes* pitchers have altered morphology (Baby et al., 2017). The transcriptome data obtained in this study confirmed the role of CO_2 in the development and functioning of the *N.* × *ventrata* pitcher, as evidenced by the upregulation of genes related to reversible CO_2 hydration and fixation, CO_2 and water transport during transpiration, and stomatal and guard cell development and function (**Supplementary Table 11**).

Traps produced by *Nepenthes* plants have low or no photosynthetic activity and, consequently, lower content of chlorophylls and carotenoids but higher content of anthocyanins (Pavlovič and Saganová, 2015). Accordingly, our comparative transcriptome analysis revealed downregulation of photosynthesis-related genes and upregulation of anthocyanin pathway genes in *N.* × *ventrata* pitchers at all stages compared to leaves (**Supplementary Table 11** and **Figures 5**, **6B**). Another recent study of *Nepenthes* pitchers (Goh et al., 2020) also showed the downregulation of genes associated with photosynthesis and increased biosynthesis of secondary metabolites in a pitcher depleted of fluid proteins.

Anthocyanin Pathway in Prey Attraction

Carnivorous plants, including *Nepenthes* spp., use insects both as pollinators and as food, participating in pollinator-prey conflict (Jürgens et al., 2012). *Nepenthes* pitchers, which do not or weakly reflect UV radiation visible to insects (Schaefer and Ruxton, 2008), acquire the same red color in pitchers and flower tepals by increasing anthocyanin production (Scharmann et al., 2019).

Red coloration could have initially developed in carnivorous plants as an adaptive trait, since anthocyanin accumulation is often associated with stress responses or nutritional deficiency in plants (Schaefer and Ruxton, 2008). At the same time, it increased prey capture efficiency of the traps by providing attractive visual signals. Capture rates are positively correlated with levels of red pigmentation, probably because insects can detect differences in red light intensity compared to the green background (Schaefer and Ruxton, 2008).

The anthocyanin pathway is activated by the MBW complex of TFs, including R2R3-MYB, bHLH, and WD40, which regulate sequential steps in the synthesis of naringenin chalcone, naringenin, and dihydrokaempferol catalyzed by chalcone synthase, chalcone isomerase, and flavanone 3-hydroxylase, respectively (Naing and Kim, 2018). The pathway is then divided into three branches, which ultimately produce three types of anthocyanins (derivatives of cyanidin, delphinidin, and pelargonidin of different colors) as a result of three sequential reactions catalyzed by dihydroflavonol-4-reductase (DFR), anthocyanidin synthase (ANS), and UDP-glucosoflavonoid-3-*O*-glucosyltransferase (UFGT), respectively (Naing and Kim, 2018).

We found that in *N.* × *ventrata* red-burgundy pitchers, most of the anthocyanin pathway genes identified in *Arabidopsis* were expressed, except for *F3'H*, whose transcripts were not detected (**Supplementary Tables 6**, **11**, **Figures 5**, **6B**, and **Supplementary Figure 6**). *F3'H* encodes flavonoid 3'-hydroxylase catalyzing conversion of dihydrokaempferol to dihydroquercetin and its absence may suggest that *N.* × *ventrata* lacks the branch of cyanidin synthesis. Products of the other two branches are blue delphinidins and orange–red pelargonidins (Tanaka and Ohmiya, 2008) and, considering *N.* × *ventrata* pitcher color, we suggest that biosynthesis of pelargonidin may be the most pronounced branch of the anthocyanin pathway in *N.* × *ventrata* pitchers (**Figure 7C**).

In some families of core Caryophyllales, flowers and fruits produce betalains: red-violet betacyanins and yellow betaxanthins, which are structurally and biochemically unrelated to anthocyanins and exclude their synthesis (Timoneda et al., 2019). Nepenthaceae are noncore Caryophyllales that produce anthocyanins (Timoneda et al., 2019); nevertheless, the gene encoding a main enzyme of the betalain pathway, 4,5-DOPA-extradiol-dioxygenase (DODA), was found to be activated in *N.* × *ventrata*, late pitcher (**Supplementary Table 6**). DODA catalyzes the conversion of 3,4-dihydroxy-L-phenyalanine (L-DOPA) to 4,5-seco-DOPA, which spontaneously forms betalamic acid, the precursor of all betalain compounds. Spontaneous conjugation of betalamic acid with cyclo-DOPA produced by cyclization of oxidized L-DOPA generates red-violet betacyanins (Polturak and Aharoni, 2018). Along with DODA activation, transcription of anthocyanin biosynthesis genes was decreased in late pitcher compared to young pitcher (**Figure 6B**), which confirms mutual exclusion of anthocyanins

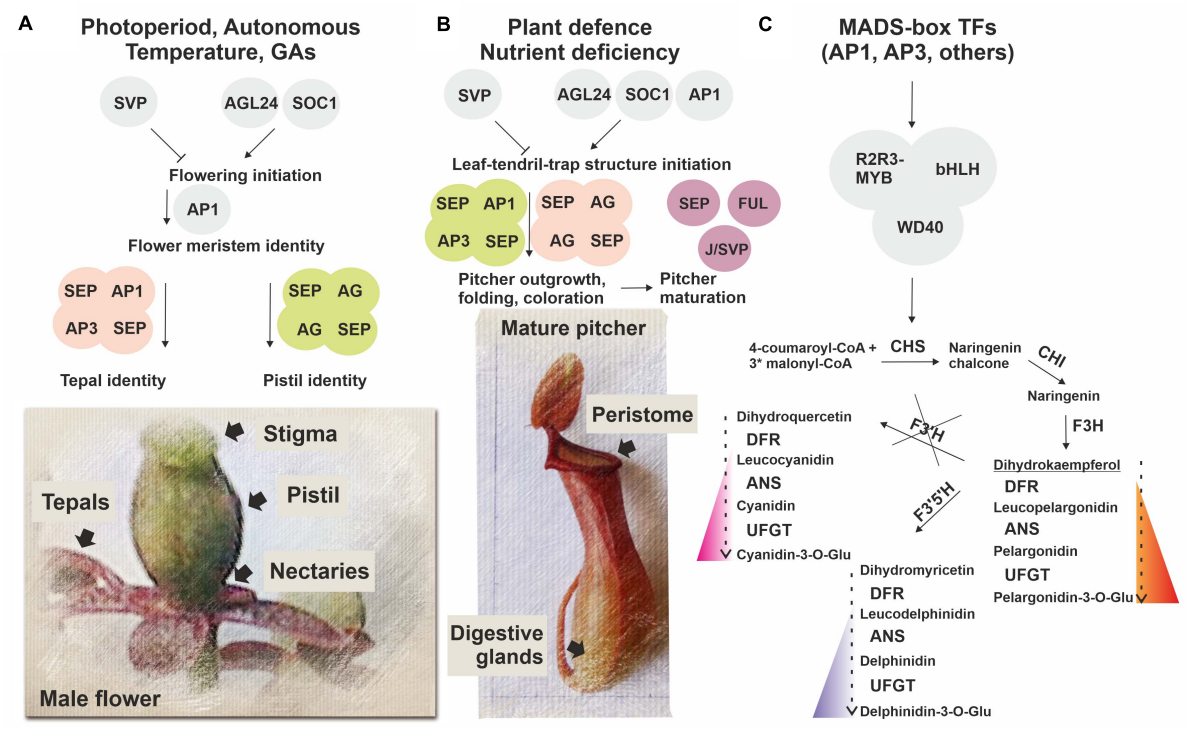


FIGURE 7 | Simplified scheme illustrating a possible role of MADS-domain TFs in *N.* × *ventrata* flower **(A)** and pitcher (leaf-tendril-trap) **(B)** initiation and specification, including anthocyanin pigmentation **(C)**. Homologs of MADS-domain TFs SVP, AGL24, SOC1, and AP1 involved in flowering initiation possibly determine the time of pitcher initiation, whereas homologs of MADS-domain TFs SEP1, 3, AP1, FUL, AP3, AG, and J/SVP involved in the control of floral organ identity and fruit development may also be implicated in conferring pitcher identity signatures. At the same time, MADS-domain TFs may regulate activation of the anthocyanin pathway in developing flowers and pitchers.

and betalains (Timoneda et al., 2019) and indicate a possibility for betalain biosynthesis at later developmental stages of noncore Caryophyllales.

MADS-Domain TFs as Possible Regulators of Pitcher Origin and Development

As part of a long-lasting discussion about the origin of the carnivorous syndrome in plants, Charles Darwin, already in the 19th century, pointed out the similarities between traps and reproductive parts of the flower (stamens and pistils), but since that time, there has been no experimental evidence to support this possibility (Pennisi, 2002). In this study, the MADS-domain TFs (**Table 1** and **Supplementary Tables 2**, **6**, **11**, **12**), which play decisive roles in plant reproductive development (Smaczniak et al., 2012; Theißen et al., 2016) (**Supplementary Tables 14**, **15**), were revealed as the most intriguing set of transcripts upregulated in *N.* × *ventrata* pitchers. In higher plants, MADS-domain TFs regulate flowering initiation (*AGL24*, *AGL42*, *SOC1*, and *AP1*) and repression (*SVP*) (Dorca-Fornell et al., 2011; Hu et al., 2014; Valentim et al., 2015). We found that *AGL24*, two of the four *AP1*/*FUL* homologous genes, and six of the eight *SOC1*/*AGL42* homologous genes were upregulated, whereas three of seven *SVP* homologous genes were downregulated in early pitcher vs. leaf (**Figure 6A**). We also observed activation of genes from other gene families (*FY*, *FCA*, *FLK*, *FD*, and *CONSTANS-like*) known to control flowering time through regulation of MADS-box gene transcription (**Table 1** and **Supplementary Tables 6**, **11**). In *N.* × *ventrata* vegetative SAM, *AGL24*, *AGL42*, and *SOC1* encode proteins which may induce the expression of inflorescence meristem identity gene *LEAFY* (*LFY*) that in turn may activate floral meristem identity gene *AP1* as shown for *Arabidopsis* (Smaczniak et al., 2012). We found all these genes, with the exception of *LFY*, both in the leaves and pitchers (**Table 1** and **Supplementary Table 12**), suggesting that the expression of these genes is independent of the flowering process and that the single leaf-tendril-trap unit may represent a novel structure (adapted for plant nutrition), as has been shown for grapevine tendrils adapted for plant climbing (Calonje et al., 2004). Transcripts of all four genes homologous to *AP1*/*FUL* were found in all types of transcriptomes (**Supplementary Table 12**), which supports the idea that the leaf, tendril, and trap represent a single structure. In this case, *LFY* could be transcribed in the meristem prior to the initiation of leaf-tendril-trap differentiation.

Nepenthes × *ventrata* pitchers were also characterized by the upregulation of MADS-box genes *AG*, *AP3*, *SEP1*, *SEP3*, *SEP4*, and *AGL6* shown to be responsible for floral organ identity (Theißen et al., 2016), as well as genes involved in the development of fruit (*FUL*, and *J*/*SVP*), pollen (*AGL65*,

and *AGL104*), and male gametophytes (*AGL80*, and *AGL62*) (Verelst et al., 2007; Nakano et al., 2015; Figueiredo et al., 2016; Toh et al., 2018; Batista et al., 2019) (**Table 1**, **Supplementary Tables 12**, **14**, **15**, and **Figure 6A**).

According to the ABCDE and Quartet models, floral organs are defined by various combinations of MADS-domain TFs (A, B, C, D, or E-class) assembled into quaternary complexes (Smaczniak et al., 2012). A + E (AP1 + SEP) specify the identity of flower meristem and sepals, A + B + E (AP1 + AP3/PISTILLATA (PI) + SEP)–petals, B + C + E (AP3/PI + AG+SEP)–stamens, C + E (AG + SEP)–carpels (forming pistil), and C + E + D (AG + SEP + STK)–ovules (Theißen et al., 2016). In *N.* × *ventrata* early pitcher and young pitcher, all ABCDE genes, except suppressed *STK* and absent *PI*, were upregulated compared to the leaf; the leaf transcriptome contained all these genes, except *AGL6* (**Supplementary Table 12**).

Duplication, diversification, and neo-functionalization of the MADS-box genes are thought to underlie the origin of flower organs through leaf modification (Melzer et al., 2010; Theißen et al., 2016). As a result, highly conserved MADS-domain TFs in various combinations represent a mechanism controlling flower meristem differentiation (Theißen et al., 2016). The overwhelming variety of existing flower forms indicates the ongoing evolution of the flower MADS-box set through selective changes in protein-protein interactions, downstream target genes, and regulatory patterns (Bartlett, 2017).

Besides traps, dioecious *Nepenthes* species form female and male flowers on separate plants and, therefore, must have a complete set of highly conserved MADS-box genes associated with flowering (Subramanyam and Narayana, 1971). Indeed, the *N. khasiana* male inflorescence transcriptome (Scharmann et al., 2019) contained the similar set of the MADS-box genes as *N.* × *ventrata* leaves and pitchers (**Supplementary Tables 2**, **12**).

It is worth noting the mimicry of pitchers as flowers in the patterns of scent emission (Di Giusto et al., 2010) and the similarity of the pitcher with a petal in color and stomatal patterning and that of the pistil with a tubular structure, extrafloral nectaries in the peristome, and lower glands at the base inside the trap, which produce digestive fluid (Gaume and Forterre, 2007; Bauer et al., 2008). In many plant species, the surface of the carpel stigma is often uneven, tuberous, and wet because of the wax cover and sugar-rich sticky exudate, which contribute to more efficient pollen adhesion (Lau et al., 2017). The peristome also has an anisotropic surface coated with a film of water and/or attractive and slippery sugar-rich nectar for prey capture (Bohn and Federle, 2004).

It is postulated that the carpels, as well as other floral organs, emerged as a result of leaf modifications, which is confirmed by the inter-conversion between leaves and flower organs when the A, B, and C classes or E-class of MADS-box genes are overexpressed or inactivated (Scutt et al., 2006).

We speculate that *N.* × *ventrata* trap development may pass through several transitional steps: (i) initiation of a leaf-tendril-trap structure due to the cooperative activity of flowering time-controlling genes *AGL24* (*NveMADS34*), *AGL42*/*SOC1* (*NveMADS37–44*), *SVP* (for example, isoforms *NveMADS32*, *33*), and *AP1* (*NveMADS11*); (ii) outgrowth, folding, and coloration of the pitcher structure [upregulation of floral organ identity-related genes *AP1* (*NveMADS11*), *AG* (*NveMADS14*, *15*), *SEP*s (*NveMADS4, 6–8*), and *AP3* (*NveMADS1–3*)]; (iii) maturation of the tubular structure acquiring differentiated pistil-like attributes, and upregulation of *SEP*s (*NveMADS5*), *FUL* (*NveMADS10*, *12*, *13*) and *J*/*SVP* (for example, isoforms *NveMADS31*, *36*) potentially involved in pitcher ripening. The functions of type I flower-specific MADS-box genes commonly involved in pollen, and male gametophyte development are unclear; they are less studied than type II genes and may have some unknown roles in the development of both flowers and pitchers.

Besides flower specification, flowering-related MADS-box genes are involved in plant responses to various abiotic stresses and wounding (e.g., from insect feeding) (Castelán-Muñoz et al., 2019), which is consistent with carnivory origin through plant defense mechanisms (Pavlovič and Mithöfer, 2019). Moreover, some studies indicate that MADS-box TFs may positively affect biosynthesis of anthocyanins (Jaakola et al., 2010; Zhao et al., 2019) used by traps to attract insects. The presence of highly conserved MADS-box genes associated with flowering and defense in all angiosperms may be the reason why a pitfall trap originated independently six times in diverse plant lineages.

Thus, non-carnivorous *Nepenthes* predecessor plants could use the MADS-box gene set specifying floral meristems and organs for adaptation to stressful conditions and nutrient deficiency. As a result, they develop a new structure, leaf-tendril-trap, based on the vegetative leaf, using the existing pathways for the formation of floral organs (tepal and pistil) (**Figure 7**). A recently shown WGD event in the last common ancestor of Droseraceae carnivorous plants (Palfalvi et al., 2020) led to the MADS-box genes duplication, and it can be speculated that leaves have acquired the ability to express such MADS-box paralogous genes, which, among other mechanisms, resulted in the leaf-tendril-trap development.

It has been shown that MADS-box genes involved in the initiation of flowering and floral organogenesis can also play an important role in the development of leaves (Burko et al., 2013; Gregis et al., 2013) and roots (Gan et al., 2005; Alvarez-Buylla et al., 2019). Therefore, the leaf-tendril-trap structure can be considered a specialized modified organ with its own unique regulatory pathway that evolved through co-option of genes from the networks controlling SAM identity and organization, leaf and root development, and flower morphogenesis.

The results obtained may clarify the genetic patterns of pitcher trap initiation and development and help answer the question of the origin of the plant carnivory syndrome. For example, given that a protocol for *Nepenthes mirabilis in vitro* regeneration and *Agrobacterium*-mediated transformation has been developed (Miguel et al., 2020), it is possible to obtain transgenic *Nepenthes* plants in which individual MADS-box genes are either overexpressed or silenced through various mechanisms such as CRISPR/Cas9-based genome editing. It would also be useful to compare the flower- and trap-specific paralogs of the MADS-box genes, including their regulatory regions. Evaluation of the morphology, genomics, and

proteomics of traps from such transgenic plants could shed light on the trap-specific functions of the analyzed MADS-box genes and their roles in trap evolution. Also, modern methods for analysis of protein-protein interactions and transcription factor target genes should make possible to compare the functional activity of MADS-box genes in the trap with that of MADS-box genes known to be involved in flowering initiation and flower development.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: NCBI, accession number: PRJNA487526 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA487526).

AUTHOR CONTRIBUTIONS

NR and EK: conceptualization. MF, EK, and EG: plant material. NR, AB, and AM: methodology. AB: software. MF and MS: validation. AB and AS: formal analysis. NR: data curation. AS: Writing–original draft preparation. AS, EK, and NR: writing–review and editing. NR: supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Russian Science Foundation (Grant No. 19-16-00016) and by the Ministry of Science and Higher Education of the Russian Federation.

ACKNOWLEDGMENTS

We are sincerely grateful to the late Konstantin Skryabin, who initiated this project and provided the opportunity to explore unusual unique plants at our Institute. We would like to thank Marina Chuenkova for English language editing. This work was performed using the experimental climate control facility in the Institute of Bioengineering (Research Center of Biotechnology, Russian Academy of Sciences).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.643137/full#supplementary-material

Supplementary Figure 1 | Gene Ontology (GO) BP terms most enriched in leaf **(a)** early pitcher, **(b)** DEGs (leaves vs. early pitcher), and in leaf **(c)**, and mature pitcher **(d)**, DEGs (leaves vs. mature pitcher). GO MF terms most enriched in leaf **(e)** early pitcher, **(f)** DEGs (leaves vs. early pitcher), in leaf **(g)**, and mature pitcher **(h)** DEGs (leaves vs. mature pitcher).

Supplementary Figure 2 | *Nepenthes* transcriptomes in Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY **(a,c)**, and KEGG BRITE **(b)** terms.

Supplementary Figure 3 | MapMan scheme of differentially regulated transcripts associated with secondary metabolism pathways: leaves vs. young pitcher **(a)**, mature pitcher vs. leaves **(b)**, mature pitcher vs. early pitcher **(c)**, and mature pitcher vs. young pitcher **(d)**.

Supplementary Figure 4 | MapMan scheme of differentially regulated transcripts associated with photosynthetic pathways: mature pitcher vs. leaves **(a)**, mature pitcher vs. early pitcher **(b)**, and mature pitcher vs. young pitcher **(c)**.

Supplementary Figure 5 | MapMan scheme of differentially regulated transcripts associated with stress response pathways: leaves vs. young pitcher **(a)**, mature pitcher vs. leaves **(b)**, mature pitcher vs. early pitcher **(c)**, and early pitcher vs. young pitcher **(d)**.

Supplementary Figure 6 | Phylogenetic tree constructed based on the sequences of MADS-domain proteins from *Nepenthes* × *ventrata* (NveMADS1-45), *Arabidopsis thaliana*, and *Beta vulgaris* (Bve): AGAMOUS/AGL11 clade **(A)**; APETALA1/FRUITFULL and APETALA3/PISTILLATA **(B)**; SEPALLATA/AGL6 **(C)**; SVP/AGL24 **(D)**; and Type I **(E)**.

Supplementary Figure 7 | Distance trees constructed for each NveMADS1-45 protein sequence using the Fast Minimum Evolution method (NCBI). Green corresponds to eudicots, yellow – to the query protein.

Supplementary Figure 8 | Relative expression of genes related to flowering (*AGAMOUS*, *SEPALLATA 3*, and *MYB17*) and anthocyanin biosynthesis (*MYB113*, *bHLH001*, *CHS1*, *CHS2*, *CHI*, *F3H*, *F3'5'H*, *DFR*, *ANS*, and *UFGT*) determined by real-time quantitative PCR (RT-qPCR).

Supplementary Table 1 | Statistics of *Nepenthes* × *ventrata* transcriptome assembly.

Supplementary Table 2 | Assembled *Nepenthes* × *ventrata* transcripts with read numbers normalized among transcriptomes.

Supplementary Table 3 | Species distribution of the top BLAST hits for assembled *Nepenthes* × *ventrata* transcripts.

Supplementary Table 4 | Hierarchical GO annotation of the assembled *Nepenthes* × *ventrata* transcripts.

Supplementary Table 5 | Annotation of the assembled *Nepenthes* × *ventrata* transcripts in GO terms.

Supplementary Table 6 | Differentially expressed transcripts (upregulation ≥ 2 times) in the assembled *Nepenthes* × *ventrata* transcriptomes.

Supplementary Table 7 | Gene Ontology (GO) enrichment and depletion in the assembled *Nepenthes* × *ventrata* transcriptomes.

Supplementary Table 8 | Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation of the assembled *Nepenthes* × *ventrata* transcripts.

Supplementary Table 9 | Kyoto Encyclopedia of Genes and Genomes enrichment in the assembled *Nepenthes* × *ventrata* transcriptomes.

Supplementary Table 10 | Plant Transcription Factor Database (PlantTFDB) annotation of the assembled *Nepenthes* × *ventrata* transcripts.

Supplementary Table 11 | Differentially expressed unique genes (upregulation ≥ 4 times) in the assembled *Nepenthes* × *ventrata* transcriptomes.

Supplementary Table 12 | MADS-box related transcripts in the *Nepenthes* × *ventrata* transcriptomes.

Supplementary Table 13 | MADS-related transcripts in the assembled transcriptomes of the *Nepenthes* × *ventrata* mature pitcher and *Nepenthes khasiana* inflorescence.

Supplementary Table 14 | GeneMANIA predicted functions and interactions of putative *Nepenthes* × *ventrata* type I MADS-domain proteins.

Supplementary Table 15 | GeneMANIA predicted functions and interactions of putative *Nepenthes* × *ventrata* type II MADS-domain proteins.

Supplementary Table 16 | Real-time quantitative PCR (RT-qPCR) primers for *Nepenthes* × *ventrata* unigene expression analysis.

REFERENCES

Alvarez-Buylla, E. R., García-Ponce, B., Sánchez, M. P., Espinosa-Soto, C., García-Gómez, M. L., Piñyro-Nelson, A., et al. (2019). MADS-box genes underground becoming mainstream: plant root developmental mechanisms. *New Phytol.* 223, 1143–1158. doi: 10.1111/nph.15793

Amanda, D., Doblin, M. S., Galletti, R., Bacic, A., Ingram, G. C., and Johnson, K. L. (2016). DEFECTIVE KERNEL1 (DEK1) regulates cell walls in the leaf epidermis. *Plant Physiol.* 172, 2204–2218. doi: 10.1104/pp.16.01401

An, R., Liu, X., Wang, R., Wu, H., Liang, S., Shao, J., et al. (2014). The over-expression of two transcription factors, ABS5/bHLH30 and ABS7/MYB101, leads to upwardly curly leaves. *PLoS One*. 9:e107637. doi: 10.1371/journal.pone.0107637

Baby, S., Johnson, A. J., Zachariah, E. J., and Hussain, A. A. (2017). Nepenthes pitchers are CO2-enriched cavities, emit CO2 to attract preys. *Sci. Rep.* 12:11281. doi: 10.1038/s41598-017-11414-7

Bartlett, M. E. (2017). Changing MADS-Box transcription factor protein-protein interactions as a mechanism for generating floral morphological diversity. *Integr. Compar. Biol.* 57, 1312–1321. doi: 10.1093/icb/icx067

Batista, R. A., Moreno-Romero, J., Qiu, Y., van Boven, J., Santos-González, J., Figueiredo, D. D., et al. (2019). The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. *bioRxiv* 8:e50541. doi: 10.1101/616698

Bauer, U., Bohn, H. F., and Federle, W. (2008). Harmless nectar source or deadly trap: Nepenthes pitchers are activated by rain, condensation and nectar. *Proc. R. Soc. B Biol. Sci. U. S. A.* 275, 259–265. doi: 10.1098/rspb.2007.1402

Bohn, H. F., and Federle, W. (2004). Insect aquaplanning: Nepenthes pitcher plants capture prey with the peristome, a fully wettable water-lubricated anisotropic surface. *Proc. Natl. Acad. Sci. U. S. A.* 101, 14138–14143. doi: 10.1073/pnas.0405885101

Bowman, J. L. (2000). The YABBY gene family and abaxial cell fate. *Curr. Opin. Plant Biol.* 3, 17–22. doi: 10.1016/S1369-5266(99)00035-7

Burko, Y., Shleizer-Burko, S., Yanai, O., Shwartz, I., Zelnik, I. D., Jacob-Hirsch, J., et al. (2013). A role for APETALA1/fruitfull transcription factors in tomato leaf development. *Plant Cell.* 25, 2070–2083. doi: 10.1105/tpc.113.113035

Calonje, M., Cubas, P., Martínez-Zapater, J. M., and Carmona, M. J. (2004). Floral meristem identity genes are expressed during tendril development in grapevine. *Plant Physiol.* 135, 1491–1501. doi: 10.1104/pp.104.040832

Castelán-Muñoz, N., Herrera, J., Cajero-Sánchez, W., Arrizubieta, M., Trejo, C., García-Ponce, B., et al. (2019). MADS-box genes are key components of genetic regulatory networks involved in abiotic stress and plastic developmental responses in plants. *Front. Plant Sci.* 10:853. doi: 10.3389/fpls.2019.00853

Di Giusto, B., Bessière, J. M., Guéroult, M., Lim, L. B. L., Marshall, D. J., Hossaert-McKey, M., et al. (2010). Flower-scent mimicry masks a deadly trap in the carnivorous plant Nepenthes rafflesiana. *J. Ecol.* 98, 845–856. doi: 10.1111/j.1365-2745.2010.01665.x

Dkhar, J., and Pareek, A. (2014). What determines a leaf's shape? *Evodevo* 5, 1–19. doi: 10.1186/2041-9139-5-47

Dkhar, J., and Pareek, A. (2019). ASYMMETRIC LEAVES1 and REVOLUTA are the key regulatory genes associated with pitcher development in Nepenthes khasiana. *Sci. Rep.* 9:6318. doi: 10.1038/s41598-019-42779-6

Dorca-Fornell, C., Gregis, V., Grandi, V., Coupland, G., Colombo, L., and Kater, M. M. (2011). The Arabidopsis SOC1-like genes AGL42, AGL71 and AGL72 promote flowering in the shoot apical and axillary meristems. *Plant J.* 67, 1006–1017. doi: 10.1111/j.1365-313X.2011.04653.x

Eshed, Y., Izhaki, A., Baum, S. F., Floyd, S. K., and Bowman, J. L. (2004). Asymmetric leaf development and blade expansion in Arabidopsis are mediated by KANADI and YABBY activities. *Development* 131, 2997–3006. doi: 10.1242/dev.01186

Figueiredo, D. D., Batista, R. A., Roszak, P. J., Hennig, L., and Köhler, C. (2016). Auxin production in the endosperm drives seed coat development in Arabidopsis. *Elife* 5:e20542. doi: 10.7554/eLife.20542

Filyushin, M. A., Kochieva, E. Z., Shchennikova, A. V., Beletsky, A. V., Mardanov, A. V., Ravin, N. V., et al. (2019). Identification and Expression Analysis of Chitinase Genes in Pitchers of Nepenthes sp. during Development. *Dokl. Biochem. Biophys.* 484, 29–32. doi: 10.1134/S1607672919010083

Fukushima, K., Fang, X., Alvarez-Ponce, D., Cai, H., Carretero-Paulet, L., Chen, C., et al. (2017). Genome of the pitcher plant Cephalotus reveals genetic changes associated with carnivory. *Nat. Ecol. Evol.* 1:59. doi: 10.1038/s41559-016-0059

Fukushima, K., Fujita, H., Yamaguchi, T., Kawaguchi, M., Tsukaya, H., and Hasebe, M. (2015). Oriented cell division shapes carnivorous pitcher leaves of Sarracenia purpurea. *Nat. Commun.* 6:6450. doi: 10.1038/ncomms7450

Gan, Y., Filleur, S., Rahman, A., Gotensparre, S., and Forde, B. G. (2005). Nutritional regulation of ANR1 and other root-expressed MADS-box genes in Arabidopsis thaliana. *Planta* 222, 730–742. doi: 10.1007/s00425-005-0020-3

Gaume, L., and Forterre, Y. (2007). A viscoelastic deadly fluid in carnivorous pitcher plants. *PLoS One*. 2:e1185. doi: 10.1371/journal.pone.0001185

Givnish, T. J. (2015). New evidence on the origin of carnivorous plants. *Proc. Natl. Acad. Sci. U. S. A.* 112, 10–11. doi: 10.1073/pnas.1422278112

Goh, H. H., Baharin, A., and Mohd Salleh, F. (2020). Transcriptome-wide shift from photosynthesis and energy metabolism upon endogenous fluid protein depletion in young Nepenthes ampullaria pitchers. *Sci. Rep.* 10:6575. doi: 10.1038/s41598-020-63696-z

Gorb, E., Kastner, V., Peressadko, A., Arzt, E., Gaume, L., Rowe, N., et al. (2004). Structure and properties of the glandular surface in the digestive zone of the pitcher in the carnivorous plant Nepenthes ventrata and its role in insect trapping and retention. *J. Exp. Biol.* 207, 2947–2963. doi: 10.1242/jeb.01128

Gregis, V., Andrés, F., Sessa, A., Guerra, R. F., Simonini, S., Mateos, J. L., et al. (2013). Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in Arabidopsis. *Genome Biol.* 14:R56. doi: 10.1186/gb-2013-14-6-r56

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084

Hatano, N., and Hamada, T. (2008). Proteome analysis of pitcher fluid of the carnivorous plant Nepenthes alata. *J. Proteome Res.* 7, 809–816. doi: 10.1021/pr700566d

Hatano, N., and Hamada, T. (2012). Proteomic analysis of secreted protein induced by a component of prey in pitcher fluid of the carnivorous plant Nepenthes alata. *J. Proteomics*. 75, 4844–4852. doi: 10.1016/j.jprot.2012.05.048

Hu, J. Y., Zhou, Y., He, F., Dong, X., Liu, L. Y., Coupland, G., et al. (2014). miR824-regulated AGAMOUS-LIKE16 contributes to flowering time repression in Arabidopsis. *Plant Cell*. 26, 2024–2037. doi: 10.1105/tpc.114.124685

Inagaki, S., Suzuki, T., Ohto, M. A., Urawa, H., Horiuchi, T., Nakamura, K., et al. (2006). Arabidopsis TEBICHI, with helicase and DNA polymerase domains, is required for regulated cell division and differentiation in meristems. *Plant Cell*. 18, 879–892. doi: 10.1105/tpc.105.036798

Jaakola, L., Poole, M., Jones, M. O., Kämäräinen-Karppinen, T., Koskimäki, J. J., Hohtola, A., et al. (2010). A SQUAMOSA MADS box gene involved in the regulation of anthocyanin accumulation in bilberry fruits. *Plant Physiol.* 153, 1619–1629. doi: 10.1104/pp.110.158279

Jürgens, A., Sciligo, A., Witt, T., El-Sayed, A. M., and Suckling, D. M. (2012). Pollinator-prey conflict in carnivorous plants. *Biol. Rev.* 87, 602–615. doi: 10.1111/j.1469-185X.2011.00213.x

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Kurup, R., Johnson, A. J., Sankar, S., Hussain, A. A., Kumar, C. S., and Sabulal, B. (2013). Fluorescent prey traps in carnivorous plants. *Plant Biol.* 15, 611–615. doi: 10.1111/j.1438-8677.2012.00709.x

Lau, J. Y. Y., Pang, C. C., Ramsden, L., and Saunders, R. M. K. (2017). Stigmatic exudate in the Annonaceae: Pollinator reward, pollen germination medium or extragynoecial compitum? *J. Integr. Plant Biol.* 59, 881–894. doi: 10.1111/jipb.12598

Li, B., and Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 12:323. doi: 10.1186/1471-2105-12-323

Liu, X., Li, M., Liu, K., Tang, D., Sun, M., Li, Y., et al. (2016). Semi-Rolled Leaf2 modulates rice leaf rolling by regulating abaxial side cell differentiation. *J. Exp. Bot.* 67, 2139–2150. doi: 10.1093/jxb/erw029

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J.* 17:10. doi: 10.14806/ej.17.1.200

Melzer, R., Wang, Y. Q., and Theißen, G. (2010). The naked and the dead: The ABCs of gymnosperm reproduction and the origin of the angiosperm flower. *Semin. Cell Dev. Biol.* 21, 118–128. doi: 10.1016/j.semcdb.2009.11.015

Miguel, S., Michel, C., Biteau, F., Hehn, A., and Bourgaud, F. (2020). In vitro plant regeneration and Agrobacterium-mediated genetic transformation of a carnivorous plant, Nepenthes mirabilis. *Sci. Rep*. 10:17482. doi: 10.1038/s41598-020-74108-7

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321

Murphy, B., Forest, F., Barraclough, T., Rosindell, J., Bellot, S., Cowan, R., et al. (2020). A phylogenomic analysis of Nepenthes (Nepenthaceae). *Mol. Phylogenet. Evol.* 144:106668. doi: 10.1016/j.ympev.2019.106668

Naing, A. H., and Kim, C. K. (2018). Roles of R2R3-MYB transcription factors in transcriptional regulation of anthocyanin biosynthesis in horticultural plants. *Plant Mol. Biol.* 98, 1–18. doi: 10.1007/s11103-018-0771-4

Nakano, T., Kato, H., Shima, Y., and Ito, Y. (2015). Apple SVP family MADS-box proteins and the tomato pedicel abscission zone regulator JOINTLESS have similar molecular activities. *Plant Cell Physiol.* 56, 1097–1106. doi: 10.1093/pcp/pcv034

Owen, T. P., and Lennon, K. A. (1999). Structure and development of the pitchers from the carnivorous plant Nepenthes alata (Nepenthaceae). *Am. J. Bot.* 86, 1382–1390. doi: 10.2307/2656921

Palfalvi, G., Hackl, T., Terhoeven, N., Shibata, T. F., Nishiyama, T., Ankenbrand, M., et al. (2020). Genomes of the Venus Flytrap and Close Relatives Unveil the Roots of Plant Carnivory. *Curr. Biol.* 30, 2312–2320. doi: 10.1016/j.cub.2020.04.051

Parenicová, L., de Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., et al. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. *Plant Cell.* 15, 1538–1551. doi: 10.1105/tpc.011544

Pavlovič, A., and Mithöfer, A. (2019). Jasmonate signalling in carnivorous plants: Copycat of plant defence mechanisms. *J. Exp. Bot.* 70, 3379–3389. doi: 10.1093/jxb/erz188

Pavlovič, A., and Saganová, M. (2015). A novel insight into the cost-benefit model for the evolution of botanical carnivory. *Ann. Bot.* 115, 1075–1092. doi: 10.1093/aob/mcv050

Pennisi, E. (2002). PLANT EVOLUTION: Elaborate Carnivorous Plants Prove to Be Kin. *Science* 297:1626. doi: 10.1126/science.297.5587.1626a

Polturak, G., and Aharoni, A. (2018). "La Vie en Rose": Biosynthesis. Sources, and Applications of Betalain Pigments. *Mol. Plant*. 11, 7–22. doi: 10.1016/j.molp.2017.10.008

Rast, M. I., and Simon, R. (2012). Arabidopsis JAGGED LATERAL ORGANS Acts With ASYMMETRIC LEAVES2 to Coordinate KNOX and PIN Expression in Shoot and Root Meristems. *Plant Cell*. 24, 2917–2933. doi: 10.1105/tpc.112.099978

Ravee, R., Salleh, F. I. M., and Goh, H. H. (2018). Discovery of digestive enzymes in carnivorous plants with focus on proteases. *PeerJ*. 6:e4914. doi: 10.7717/peerj.4914

Renner, T., and Specht, C. D. (2013). Inside the trap: Gland morphologies, digestive enzymes, and the evolution of plant carnivory in the Caryophyllales. *Curr. Opin. Plant Biol.* 16, 436–442. doi: 10.1016/j.pbi.2013.06.009

Rottloff, S., Miguel, S., Biteau, F., Nisse, E., Hammann, P., Kuhn, L., et al. (2016). Proteome analysis of digestive fluids in Nepenthes pitchers. *Ann. Bot.* 117, 479–495. doi: 10.1093/aob/mcw001

Schaefer, H. M., and Ruxton, G. D. (2008). Fatal attraction: Carnivorous plants roll out the red carpet to lure insects. *Biol. Lett.* 4, 153–155. doi: 10.1098/rsbl.2007.0607

Scharmann, M., Grafe, T. U., Metali, F., and Widmer, A. (2019). Sex is determined by XY chromosomes across the radiation of dioecious Nepenthes pitcher plants. *Evol. Lett.* 3, 586–597. doi: 10.1002/evl3.142

Scholz, I., Bückins, M., Dolge, L., Erlinghagen, T., Weth, A., Hischen, F., et al. (2010). Slippery surfaces of pitcher plants: Nepenthes wax crystals minimize insect attachment via microscopic surface roughness. *J. Exp. Biol.* 213, 1115–1125. doi: 10.1242/jeb.035618

Scutt, C. P., Vinauger-Douard, M., Fourquin, C., Finet, C., and Dumas, C. (2006). An evolutionary perspective on the regulation of carpel development. *J. Exp. Bot.* 57, 2143–2152. doi: 10.1093/jxb/erj188

Seppey, M., Manni, M., and Zdobnov, E. M. (2019). BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol. Biol.* 1962, 227–245. doi: 10.1007/978-1-4939-9173-0_14

Shpak, E. D. (2013). Diverse roles of ERECTA family genes in plant development. *J. Integr. Plant Biol.* 55, 1238–1250. doi: 10.1111/jipb.12108

Smaczniak, C., Immink, R. G. H., Angenent, G. C., and Kaufmann, K. (2012). Developmental and evolutionary diversity of plant MADS-domain factors: Insights from recent studies. *Dev* 139, 3081–3098. doi: 10.1242/dev.074674

Subramanyam, K., and Narayana, L. L. (1971). A contribution to the floral anatomy of Nepenthes khasiana Hook F. *Proc. Indian Acad. Sci. Sect. B.* 73, 124–131. doi: 10.1007/BF03045312

Tanaka, Y., and Ohmiya, A. (2008). Seeing is believing: engineering anthocyanin and carotenoid biosynthetic pathways. *Curr. Opin. Biotechnol.* 19, 190–197. doi: 10.1016/j.copbio.2008.02.015

Theißen, G., Melzer, R., and Ruümpler, F. (2016). MADS-domain transcription factors and the floral quartet model of flower development: Linking plant development and evolution. *Dev* 143, 3259–3271. doi: 10.1242/dev.134080

Thorogood, C. J., Bauer, U., and Hiscock, S. J. (2018). Convergent and divergent evolution in carnivorous pitcher plant traps. *New Phytol.* 217, 1035–1041. doi: 10.1111/nph.14879

Timoneda, A., Feng, T., Sheehan, H., Walker-Hale, N., Pucker, B., Lopez-Nieves, S., et al. (2019). The evolution of betalain biosynthesis in Caryophyllales. *New Phytol.* 224, 71–85. doi: 10.1111/nph.15980

Toh, S. S., Chen, Z., Rouchka, E. C., Schultz, D. J., Cuomo, C. A., and Perlin, M. H. (2018). Pas de deux: An intricate dance of anther smut and its host. *G3* 8, 505–518. doi: 10.1534/g3.117.300318

Valentim, F. L., Van Mourik, S., Posé, D., Kim, M. C., Schmid, M., Van Ham, R. C. H. J., et al. (2015). A quantitative and dynamic model of the arabidopsis flowering time gene regulatory network. *PLoS One*. 10:e0116973. doi: 10.1371/journal.pone.0116973

van der Graaff, E., Laux, T., and Rensing, S. A. (2009). The WUS homeobox-containing (WOX) protein family. *Genome Biol.* 10:248. doi: 10.1186/gb-2009-10-12-248

Verelst, W., Twell, D., de Folter, S., Immink, R., Saedler, H., and Münster, T. (2007). MADS-complexes regulate transcriptome dynamics during pollen maturation. *Genome Biol.* 8:R249. doi: 10.1186/gb-2007-8-11-r249

Walker, J. F., Yang, Y., Moore, M. J., Mikenas, J., Timoneda, A., Brockington, S. F., et al. (2017). Widespread paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales. *Am. J. Bot*. 104, 858–867. doi: 10.3732/ajb.1700083

Wan Zakaria, W. N. A., Aizat, W. M., Goh, H. H., and Mohd Noor, N. (2019). Protein replenishment in pitcher fluids of Nepenthes × ventrata revealed by quantitative proteomics (SWATH-MS) informed by transcriptomics. *J. Plant Res.* 132, 681–694. doi: 10.1007/s10265-019-01130-w

Wan Zakaria, W. N. A., Loke, K. K., Goh, H. H., and Mohd Noor, N. (2016). RNA-seq analysis for plant carnivory gene discovery in Nepenthes×ventrata. *Genomics Data* 7:11 doi: 10.1016/j.gdata.2015.11.007

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38, W214–W220. doi: 10.1093/nar/gkq537

Yao, G., Jin, J. J., Li, H. T., Yang, J. B., Mandala, V. S., Croley, M., et al. (2019). Plastid phylogenomic insights into the evolution of Caryophyllales. *Mol. Phylogenet. Evol.* 134, 74–86. doi: 10.1016/j.ympev.2018.12.023

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11:R14. doi: 10.1186/gb-2010-11-2-r14

Yruela, I. (2013). Transition metals in plant photosynthesis. *Metallomics* 5, 1090–1109. doi: 10.1039/c3mt00086a

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMI. A J. Integr. Biol.* 16, 284–287. doi: 10.1089/omi.2011.0118

Zhao, H. bo, Jia, H. min, Wang, Y., Wang, G. yun, Zhou, C. chao, Jia, H. juan, et al. (2019). Genome-wide identification and analysis of the MADS-box gene family and its potential role in fruit development and ripening in red bayberry (Morella rubra). *Gene* 717:144045. doi: 10.1016/j.gene.2019.144045

Zhou, Y., Liu, X., Engstrom, E. M., Nimchuk, Z. L., Pruneda-Paz, J. L., Tarr, P. T., et al. (2015). Control of plant stem cell function by conserved interacting transcriptional regulators. *Nature* 517, 377–380. doi: 10.1038/nature13853

Zulkapli, M., Mu'izzuddin, Rosli, M. A. F., Salleh, F. I. M., Mohd Noor, N., Aizat, W. M., et al. (2017). Iso-Seq analysis of Nepenthes ampullaria, Nepenthes rafflesiana and Nepenthes × hookeriana for hybridisation study in pitcher plants. *Genomics Data* 12, 130–131. doi: 10.1016/j.gdata.2017.05.003

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 11 October 2021
doi: 10.3389/fpls.2021.743643

On the Species Delimitation of the *Maddenia* Group of *Prunus* (Rosaceae): Evidence From Plastome and Nuclear Sequences and Morphology

Na Su[1,2†], *Bin-bin Liu*[3,4†], *Jun-ru Wang*[1,2†], *Ru-chang Tong*[1,2], *Chen Ren*[5,6], *Zhao-yang Chang*[1,2], *Liang Zhao*[1,2*], *Daniel Potter*[7] *and Jun Wen*[4]

[1] College of Life Sciences, Northwest A&F University, Yangling, China, [2] Herbarium of Northwest A&F University, Yangling, China, [3] State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China, [4] Department of Botany, National Museum of Natural History, MRC 166, Smithsonian Institution, Washington, DC, United States, [5] Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China, [6] Center of Conservation Biology, Core Botanical Gardens, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China, [7] Department of Plant Sciences, MS2, University of California, Davis, Davis, CA, United States

OPEN ACCESS

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

Reviewed by:
Qiang Fan, Sun Yat-sen University, China
Gregory W. Stull, Kunming Institute of Botany (CAS), China

***Correspondence:**
Liang Zhao
biology_zhaoliang@126.com

[†]*These authors have contributed equally to this work*

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

Received: *19 July 2021*
Accepted: *06 September 2021*
Published: *11 October 2021*

Citation:
Su N, Liu BB, Wang JR, Tong RC, Ren C, Chang ZY, Zhao L, Potter D and Wen J (2021) On the Species Delimitation of the Maddenia Group of Prunus (Rosaceae): Evidence From Plastome and Nuclear Sequences and Morphology. Front. Plant Sci. 12:743643. doi: 10.3389/fpls.2021.743643

The recognition, identification, and differentiation of closely related plant species present significant and notorious challenges to taxonomists. The *Maddenia* group of *Prunus*, which comprises four to seven species, is an example of a group in which species delimitation and phylogenetic reconstruction have been difficult, due to the lack of clear morphological distinctions, limited sampling, and low informativeness of molecular evidence. Thus, the precise number of species in the group and the relationships among them remain unclear. Here, we used genome skimming to generate the DNA sequence data for 22 samples, including 17 *Maddenia* individuals and five outgroups in Amygdaloideae of Rosaceae, from which we assembled the plastome and 446 single-copy nuclear (SCN) genes for each sample. The phylogenetic relationships of the *Maddenia* group were then reconstructed using both concatenated and coalescent-based methods. We also identified eight highly variable regions and detected simple sequence repeats (SSRs) and repeat sequences in the *Maddenia* species plastomes. The phylogenetic analysis based on the complete plastomes strongly supported three main subclades in the *Maddenia* group of *Prunus*, while five subclades were recognized based on the nuclear tree. The phylogenetic network analysis detected six hybridization events. Integrating the nuclear and morphological evidence, we proposed to recognize five species within the *Maddenia* group, i.e., *Prunus fujianensis*, *P. himalayana*, *P. gongshanensis*, *P. hypoleuca*, and *P. hypoxantha*. Within this group, the first three species are well-supported, while the gene flow occurring throughout the *Maddenia* group seems to be especially frequent between *P. hypoleuca* and *P. hypoxantha*, eroding the barrier between them. The phylogenetic trees based on eight concatenated hypervariable regions had a similar topology with the complete plastomes, showing their potential as molecular markers and effective barcodes for further phylogeographic studies on *Maddenia*.

Keywords: *Maddenia*, *Prunus*, Rosaceae, barcoding, chloroplast genome, single-copy nuclear genes, species delimitation

INTRODUCTION

Prunus L. is a genus of more than 200 species, widely distributed in the temperate regions of the Northern Hemisphere and in the subtropics and tropics (Rehder, 1956; Yü et al., 1986; Lu et al., 2003; Hodel et al., 2021). Some taxa of *Prunus* (e.g., almonds, sweet cherries, peaches, and plums) are of significant economic value, and other species have also been used as ornamentals, timber, and medicine (Andro and Riffaud, 1995; Lee and Wen, 2001; Wen et al., 2008).

Maddenia Hook. f. & Thoms was established as a genus by Hooker and Thomson (1854) and was later merged with *Prunus* by Chin et al. (2010) based on the phylogenetic analyses of nuclear and plastid DNA sequences. This provided strong support for the monophyly of *Maddenia* but it was resolved as nested within *Prunus*; these conclusions have also been supported by subsequent studies (Chin et al., 2014; Zhao et al., 2016, 2018; Wang et al., 2021). The *Maddenia* group of *Prunus* is characterized by its simple deciduous leaves with a serrate margin, terminal racemose inflorescences, 10 undifferentiated perianth parts at maturity, and drupe fruits (**Figure 1**; Focke, 1894; Yü et al., 1986; Lu et al., 2003; Kalkman, 2004; Wang et al., 2021). The group includes about 4–7 species endemic to East Asia, mainly distributed in the temperate regions of the Himalayas and eastern China, with China as its center of diversity, and one species in Bhutan, Nepal, and Sikkim of India (Rehder, 1956; Yü et al., 1986; Lu et al., 2003; Chin et al., 2010; Wen and Shi, 2012).

Within the *Maddenia* group, *Prunus himalayana* Hook. f. & Thomson was the first species described, followed by six other putative species (i.e., *P. hypoleuca* Koehne, *P. hypoxantha* Koehne, *P. wilsonii* Koehne, *P. fujianensis* Y. T. Chang, *P. incisoserrata* T. T. Yü & T. C. Ku, and *P. gongshanensis* J. Wen; Hooker and Thomson, 1854; Koehne, 1911; Chang, 1985; Yü et al., 1985; Lu et al., 2003; Wen and Shi, 2012). The species in this group were originally described based on morphological traits, especially the abaxial leaf pubescence (Yü et al., 1986; Lu et al., 2003; Wen and Shi, 2012). For example, *P. hypoxantha* and *P. wilsonii* were considered as two separate species based on the denser pubescence on the veins in *P. wilsonii*, and the two were also differentiated based on the size of their winter bud scales (Yü et al., 1986; Lu et al., 2003). However, Wen and Shi (2012) noted a continuous variation in the leaf pubescence between *P. hypoxantha* and *P. wilsonii*, and therefore treated the latter as a synonym of *P. hypoxantha*. This treatment was also supported by Shi et al. (2013), based on pollen morphology. Furthermore, the relationships among *P. fujianensis*, *P. hypoleuca*, and *P. incisoserrata* are poorly understood (Chang, 1985; Yü et al., 1986; Wen and Shi, 2012). *P. hypoleuca* was described based on its abaxially glabrous leaves, while *P. incisoserrata* and *P. fujianensis* both have pubescent abaxial leaf surfaces (Lu et al., 2003). Additionally, *P. incisoserrata* and *P. fujianensis* were recognized by some workers based on their leaf margin morphology (incised doubly serrate in the former vs. margin incised irregular serrate in the latter; Lu et al., 2003). However, in previous observations, we found that there was a continuous variation in the degree of abaxial pubescence as *P. hypoleuca* also has abaxially pubescent leaf blades and that there was a broad variation on the margin shape of *P. incisoserrata* and *P. fujianensis*, which greatly increased the difficulty in identifying them. In the latest revision of the *Maddenia* clade, Wen and Shi (2012) treated *P. fujianensis* and *P. incisoserrata* as synonyms of *P. hypoleuca* and they also recognized the former variety *P. himalaica* var. *glabrifolia* as a distinct species, *P. gongshanensis*.

It has been challenging to identify these species due to the existence of intermediate morphological features in the *Maddenia* clade. Traditional morphological methods alone cannot meet the needs for the species delimitation of the *Maddenia* group. With the rapid development of the phylogenetic analysis of *Prunus* s.l., the relationships within the *Prunus* and the *Maddenia* group have attracted new attention [see Chin et al. (2014)]. Yet to date, interspecific relationships within the *Maddenia* group are still unclear due to the limited taxon sampling and phylogenetically informative sites included in previous studies (Wen et al., 2008; Chin et al., 2010, 2014; Zhao et al., 2016, 2018).

Deoxyribonucleic acid (DNA) barcoding is an effective way to identify species by using a short DNA sequence (Kress et al., 2005; China Plant BOL Group, 2011; Li et al., 2015; Kress, 2017), however, DNA barcodes generally provide a limited number of informative sites among closely related taxa. As an alternative, genome skimming has been employed to generate complete chloroplast genomes (plastomes), an approach that has been dubbed as "super-barcoding" (Erickson et al., 2008; Yang et al., 2013; Li et al., 2015). The maternal inheritance and conservative genome structure of plastomes have rendered them essential markers in studying the evolutionary history of angiosperms (Gitzendanner et al., 2018; Do et al., 2020; Cai et al., 2021); noteworthy examples include the recent applications in Magnoliaceae (Wang et al., 2020), Rosaceae (Liu et al., 2019, 2020a,b), and Vitaceae (Wen et al., 2018). However, the uniparental inheritance of plastomes limits their power to fully elucidate the evolutionary histories of lineages with reticulate evolution, which has been proved to be very common in Rosaceae (Liu et al., 2020a,b; Hodel et al., 2021). In a case study on Vitaceae, Liu et al. (2021) proposed a new method for obtaining single-copy nuclear (SCN) genes from deep genome skimming data (minimum 10× coverage for optimal performance), and this approach provided a good opportunity to infer phylogenetic relationships using the uniparentally inherited plastomes and the biparentally inherited nuclear genes. Additionally, with the rapid development of next-generation sequencing, it has been feasible to obtain genome skimming data efficiently and economically (Zimmer and Wen, 2015; Zhang N. et al., 2017).

In this study, we assembled 22 plastomes and captured 446 SCN genes from seven assumed species of *Maddenia* and five outgroup species in Amygdaloideae of Rosaceae (Xiang et al., 2016; Zhang S. D. et al., 2017). We also examined their morphological and micromorphological characteristics. We identified simple sequence repeats (SSRs) and repeat sequences from the plastomes of *Maddenia* clade species. Additionally, eight highly variable regions were determined from the plastomes. We aim to test the hypotheses on species delimitations and resolve the interspecific relationships within *Maddenia*, integrating the

plastome, nuclear, and morphological evidence. We also aim to provide potential molecular markers and effective barcodes for further population-level studies on the *Maddenia* group.

MATERIALS AND METHODS

Sampling, DNA Extraction, and Sequencing

For this study, 22 individuals were sampled, including 17 ingroup individuals from the *Maddenia* group and five outgroup species from the other clades of the Rosaceae subfamily, Amygdaloideae (**Table 1**), which includes three other species of *Prunus*. The 17 ingroup samples represented the taxonomic and geographic coverage of *Maddenia* (Yü et al., 1986; Lu et al., 2003). Total genomic DNAs were extracted from 15 mg of silica gel dried leaves using the Cetyltrimethylammonium Bromide (CTAB) method (Doyle and Doyle, 1987). The libraries were prepared at the Molecular Biology Experiment Center, Germplasm Bank of Wild Species in Southwest China using a NEBNext® Ultra™ II DNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA). The paired-end (150 bp) sequencing of the DNA libraries was done on a HiSeq 2500 (Illumina, Inc., San Diego, CA, USA) platform in Beijing Genomics Institution (BGI) (Shenzhen, China), generating ∼2 GB of raw data for each sample.

Plastid Genome and Nuclear Ribosomal DNA (nrDNA) Assembly, Annotation, Visualization, and Phylogenetic Inference

The raw Illumina data were filtered for sequence quality using Trimmomatic v. 0.40 (Bolger et al., 2014) under default parameters. The filtered reads were assembled into plastome using the GetOrganelle pipeline (Jin et al., 2020). For a few accessions, Local Blast (Altschul et al., 1997) was used to align the contigs with the reference genomes (*Prunus armeniaca* (KY420025) and *P. salicina* (KY420002); Zhang N. et al., 2017; Zhang S. D. et al., 2017; Zhang X. et al., 2017). Finally, we concatenated each contig based on the orientation of the reference genome and obtained the consensus sequences through Geneious v.11.0.2 (Kearse et al., 2012). We annotated the assembled chloroplast genomes using Plastid Genome Annotator (PGA: Qu et al., 2019) and made minor manual adjustments using Geneious v.11.0.2. The transfer RNA (tRNA) genes were checked using tRNAscan-SE v.2.0 (Lowe and Chen, 2016). The circular plastid genome diagram was generated using the online OGDRAW (Lohse et al., 2013). The newly generated plastome sequence data of *Maddenia* and the other species of Rosaceae from this study have been submitted to GenBank (**Table 1**).

To obtain high-quality nuclear ribosomal DNA (nrDNA), including the Internal Transcribed Spacer (ITS) 1, 5.8S, and ITS2, a modified reference-based and *de novo* method (Zhang et al., 2015; Liu et al., 2019, 2020a,b) was employed for the assembly of the ITS sequences. The clean reads generated by Trimmomatic v. 0.40 (Bolger et al., 2014) were mapped to the reference sequence (*Prunus hypoleuca*: MH711078) using Bowtie2 v. 2.4.2 (Langmead and Salzberg, 2012), and then the draft sequence for each sample was generated. In addition, we conducted a *de novo* assembly using SPAdes v. 3.15.0 (Bankevich et al., 2012); the resulting scaffolds were used to correct the errors and ambiguities in the consensus sequences. Finally, we obtained high-quality nuclear ribosomal DNA (nrDNA) sequences for each sample using reference-based and *de novo* assembly methods.

We aligned the plastome and the ITS sequences using MAFFT (Katoh and Standley, 2013) using the software Geneious v.11.0.2 (Kearse et al., 2012). Based on maximum likelihood (ML) and Bayesian inference (BI) methods, we reconstructed the phylogeny of the *Maddenia* clade using the following nine datasets: (1) complete plastid genomes; (2) large-single-copy (LSC); (3) small-single-copy (SSC); (4) one inverted repeat (IR); (5) coding sequences (CDS); (6) non-coding region; (7) ITS; (8) concatenated sequence of *matK*, *rbcL*, and *trnH-psbA*; and (9) concatenated sequence of identified hypervariable regions. BI analyses were then conducted using MrBayes v.3.2 (Ronquist et al., 2012). The best-fitting models of nucleotide substitutions

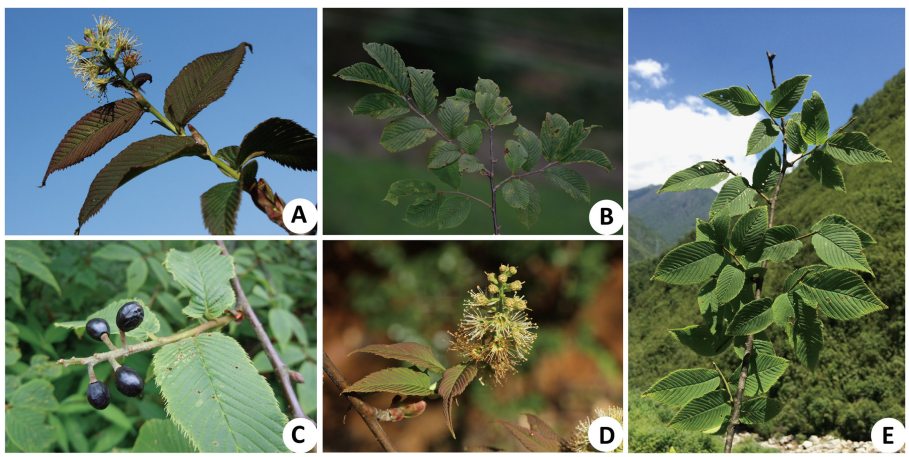


FIGURE 1 | Morphological characteristics of *Maddenia* species. **(A)** *Prunus incisoserrata*; **(B,C)** *P. wilsonii*; **(B)** simple leaf and serrated leaf margin; **(C)** drupe; **(D,E)** *P. hypoxantha*; **(D)** racemose inflorescence; **(E)** simple leaf and serrated leaf margin.

TABLE 1 | Voucher information and GenBank accession numbers of sampled species.

Taxon	Voucher	Location	Latitude (*N*)	Longitude (*E*)	Altitude (*m*)	Raw reads	Sequencing coverage	Mean coverage of plastomes(×)	GenBank accession number	SRA accession number
P. incisoserrata	JR301	Pan tou lu, Qin xu xiang, Min xian County, Gansu, China	34°23′20.02″	103°55′44.28″	2,553	11,635,840	4.79	53.20	MN864470	SRR13868100
P. incisoserrata	JR334	Da dian, Taibai Mt., Shaanxi, China	34°3′39.17″	107°42′20.30″	2,005	14,681,382	6.05	27.20	MN864488	SRR13863260
P. incisoserrata	JR440	Ping he liang, Ning shan County, Shaanxi, China	33°28′59.66″	108°29′46.60″	2,324	20,462,458	8.41	184.73	MN864490	SRR13868097
P. hypoleuca	WX219	Dian bing chang, Taibai Mt., Shaanxi, China	34°4′51.51″	107°42′16.14″	1,310	19,977,616	7.30	70.30	MK911762	SRR12920653
P. hypoleuca	JR324	Shen nong jia Mt., Hubei, China	31°25′32.15″	110°16′51.25″	2,811	17,831,310	7.01	185.40	MN864482	SRR13863263
P. hypoleuca	JR336	Between Da dian and Ping an si, Taibai Mt, Shaanxi, China	34°1′19.28″	107°43′21.75″	2,665	17,536,240	6.85	99.17	MN864483	SRR13863262
P. wilsonii	WX202	Jin ding, Emei Mt., Sichuan, China	29°31′48.11″	103°20′13.01″	2,977	18,022,432	6.90	449.36	MK905682	SRR12920659
P. wilsonii	JR352	Pan tou lu, Qin xu xiang, Min xian County, Gansu, China	34°23′20.02″	103°55′44.28″	2,553	20,192,064	8.27	73.89	MN864491	SRR13863258
P. wilsonii	JR438	Ping he liang, Ning shan County, Sichuan, China	33°28′59.66″	108°29′46.60″	2,324	20,231,114	7.95	261.45	MN864493	SRR13868096
P. hypoxantha	JR372	Cai yuan zi cun, Kang ding County, Sichuan, China	30°3′40.51″	102°0′22.07″	2,498	20,459,760	8.39	35.00	MN864485	SRR13863257
P. hypoxantha	JR426	Tian men shi, Emei Mt., Sichuan, China	29°31′46.26″	103°20′7.75″	2,950	20,149,304	8.26	394.52	MN864486	SRR13868094
P. himalayana	JR377	A sang qiao, Ya dong County, Tibet, China	27°23′43.53″	88°58′32.80″	2,750	19,176,804	8.11	25.42	MN864480	SRR13863252
P. himalayana	JR385	Lin chang, Bo mi County, Tibet, China	29°51′36.25″	95°46′3.39″	2,612	20,387,188	8.27	83.04	MN864481	SRR13863251
P. gongshanensis	JR305	Dan zha cun, Hou qiao zhen, Teng chong County, Yunnan, China	25°32′55.89″	98°13′10.22″	2,600	18,584,668	7.67	12.02	MN864472	SRR13863265
P. gongshanensis	JR381	Lu ma deng xiang, Fu gong County, Yunnan, China	27°10′17.79″	98°45′36.96″	2,276	23,789,540	9.65	21.52	MN864477	SRR13863249
P. fujianensis	JR302	Huang gang Mt., Wuyi Mt., Fujian, China	27°52′18.53″	117°51′36.33″	2,140	21,000,492	8.27	318.80	MN864466	SRR13871574
P. fujianensis	JR386	Wuyi Mt., Qian shan County, Jiang xi, China	27°59′44.73″	117°48′52.92″	2,069	21,705,762	9.09	71.65	MN864475	SRR13863264
P. discadenia	WX203	Ning shan County, Shaanxi, China	33°23′54.95″	108°22′18.90″	2,158	75,563,260	29.52	2076.39	MK905683	SRR12927899
P. serotina	WX204	DC, USA	38°91′	−77°01′	120	18,698,840	7.20	780.63	MK905684	SRR12920648
P. kansuensis	WX207	Mei xian, Shaanxi, China	34°03′43.62″	107°45′34.68″	975	25,613,228	10.53	616.67	MK634746	SRR12920639
Physocarpus amurensis	WX230	Shaanxi, China	34°15′38.35″	108°04′9.03″	355	28,712,176	11.95	701.57	MK911770	SRR12920643
Prinsepia uniflora	WX231	Shaanxi, China	34°15′38.35″	108°04′9.03″	355	18,097,112	7.36	155.54	MK911771	SRR12920642

for BI analyses were determined based on the Akaike Information Criterion (AICc) through the CIPRES Science Gateway website (Miller et al., 2010). MrBayes was run for 10,000,000 generations, sampling every 1,000 generations. The first 25% of trees were discarded as a burn-in and the remaining trees were used to estimate the 50% majority-rule consensus tree and the Bayesian posterior probabilities (PP). For ML, all analyses were performed using the RAxML-HPC Black Box 8.2.12 (Stamatakis, 2014) with 10,000 bootstrap replicates and a GTR + G model at the CIPRES Science Gateway website (Miller et al., 2010).

Plastome Comparisons and Identification of Hypervariable Regions

Gene rearrangement events within the *Maddenia* clade were detected using the Mauve v2.4.0 (Darling et al., 2010) software. We chose one *Maddenia* sequence from each species for plastome comparisons, which were performed online using mVISTA in Shuffle-LAGAN mode (Frazer et al., 2004). The reference sequence used was *P. wilsonii* WX202.

To identify the hypervariable regions, we used 22 plastomes to conduct the sliding window analysis in DnaSP v5 (Librado and Rozas, 2009) using a step size of 200 bp and a window length of 600 bp. We chose the sequences with relatively higher values of nucleotide diversity (Pi) as the hypervariable regions. The Pi refers to the difference of the chloroplast genome sequences among sequenced samples.

Single-Copy Nuclear Marker Development, Gene Assembly, Alignment, and Phylogenetic Inference

As a part of the integrative systematic studies of *Prunus*, Hodel et al. (2021) identified 591 single-copy nuclear exons based on 17 transcriptomes of *Prunus*. Our genome skimming data were sequenced from the whole genomic DNA, which provided the opportunity to capture nuclear genes, including exon and intron sequences. We used three genomes of *Prunus* (*P. dulcis* (Mill.) D. A. Webb (https://www.ncbi.nlm.nih.gov/genome/10947), *P. mume* (Siebold) Sieb. et Zucc. (https://www.ncbi.nlm.nih.gov/genome/13911), and *P. persica* (L.) Batsch (https://www.ncbi.nlm.nih.gov/genome/388) as references to discover the corresponding complete genes (introns and exons) for the 591 exons. The resulting nuclear genes were used as references in the following gene assembly.

For assembling the SCN genes, we followed the pipelines of Liu et al. (2021). Briefly, the adapters and low-quality reads were trimmed using Trimmomatic v. 0.40 (Bolger et al., 2014), and the results were quality-checked using FastQC v. 0.11.9 (Andrews, 2018). The resulting clean reads were counted to calculate the sequencing coverage, assuming the genome size (352.9 Mb: Shirasawa et al., 2017) of *P. avium* (L.) L. HybPiper pipeline v. 1.3.1 (Johnson et al., 2016), with the default settings, was used to target the SCN genes; BWA v. 0.7.1 (Li and Durbin, 2009) was used to align and distribute the reads to the target genes; SPAdes v. 3.15.0 (Bankevich et al., 2012), with a coverage cutoff value of 5, was used to assemble the reads to the contigs; and Exonerate v. 2.2.0 (Slater and Birney, 2005) was used to align the assembled contigs to the target sequences and determine the exon-intron boundaries. To balance the quality and quantity of the captured SCN genes from the uneven sequencing coverage of genome skimming data (cf. **Table 1**), we used a relatively lower coverage cutoff for generating the contigs in SPAdes v. 3.15.0. Python and R scripts included in the HybPiper pipeline (Johnson et al., 2016) were used to retrieve the recovered gene sequences, and to summarize and visualize the recovery efficiency.

The sequences in each SCN gene were aligned using MAFFT v. 7.475 (Nakamura et al., 2018) with the settings: "–localpair –maxiterate1000." Due to the variable sequencing depth in the genome skimming data, we employed three steps to remove the poorly aligned regions. In the first step, we used trimAL v. 1.2 (Capella-Gutiérrez et al., 2009) to trim the alignment of each SCN gene, in which all columns with gaps in more than 20% of the sequences or with a similarity score lower than 0.001 were removed. Considering the low-quality assembly in some regions, we used Spruceup (Borowiec, 2016) to discover, visualize, and remove the outlier sequences in the concatenated multiple sequence alignments with a window size of 50 and an overlap of 25. Because the Spruceup algorithm works better the more data it has, we concatenated all the SCN gene alignments using AMAS v. 1.0 (Borowiec, 2016) before running Spruceup, and we also used AMAS v. 1.0 (Borowiec, 2016) to split the processed/trimmed alignment back into single-locus alignments. The resulting alignments for each SCN gene were trimmed again using trimAL v. 1.2 (Capella-Gutiérrez et al., 2009) with the same parameters described above. At the third step, we excluded the sequences with <250 bp in each alignment using our customized python script (*exclude_short_sequences.py*), as the short sequences in each alignment have limited informative sites for the following coalescent-based species tree inference. Phylogenetic inference of the nuclear data of the *Maddenia* group was performed using both concatenated and coalescent-based methods. To reduce the effect of the missing data, gene alignments with at least 1,000 characters and 18 out of 22 taxa were retained. For the concatenation analysis, the best-fit partitioning schemes and nucleotide substitution models for the nuclear dataset were estimated using PartitionFinder2 (Stamatakis, 2006; Lanfear et al., 2016), under the corrected AICc and linked branch lengths, and with rcluster (Lanfear et al., 2014) algorithm options. The resulting scheme was then used to infer the ML trees using IQ-TREE 2 (Minh et al., 2020) and RAxML 8.2.12 (Stamatakis, 2014), respectively. To estimate the coalescent-based species tree, first, we inferred the individual ML gene trees using RAxML 8.2.12 (Stamatakis, 2014) with a GTRGAMMA model and 100 bootstrap replicates to assess the clade support, in which the low support branches ($\leq$10) of gene trees were contracted by Newick Utilities (Junier and Zdobnov, 2010). The gene trees were then used to infer a species tree with ASTRAL-III v. 5.7.7 (Zhang et al., 2018) using local posterior probabilities (LPP; Sayyari and Mirarab, 2016) to assess clade support.

Retrieving Standard DNA Barcodes

To determine if standard DNA barcodes can resolve the interspecific relationships of *Maddenia* species, we

extracted the gene sequences of *matK*, *rbcL*, and *trnH-psbA* from annotated plastomes, and then concatenated them into a single aligned dataset in Geneious v.11.0.2 (Kearse et al., 2012).

Phylogenetic Network Analyses

To explore the possibility of gene flow as a cause of discordance in the *Maddenia* group, we utilized 18 samples, including 17 *Maddenia* ingroups and one outgroup (*Prunus davidiana* (Carrière) Franch) for the phylogenetic network analyses. Species Networks applying Quartets (SNaQ: Solís-Lemus and Ané, 2016), as implemented in the Julia package PhyloNetworks (Solís-Lemus et al., 2017), was used to examine the contribution of the incomplete lineage sorting (ILS) and reticulation to the phylogenetic history of the *Maddenia* group. We used the ML tree inferred by RAxML for calculating the Concordance Factors (CFs), and the ASTRAL species tree was used as the input tree for SNaQ. We first tested the fit of the models, allowing from 0 to 8 reticulation events (h), and compared the models using their pseudolikelihood scores. For each number of hybrid nodes, we ran 50 SNaQ searches using the best topology from the previous run as a starting tree and retained the highest pseudolikelihood value. To distinguish the best fitting model, we used the log pseudolikelihood profile with h. A sharp improvement is expected until h reaches the best value and a slower, linear improvement thereafter. The best network was visualized in Julia using R.

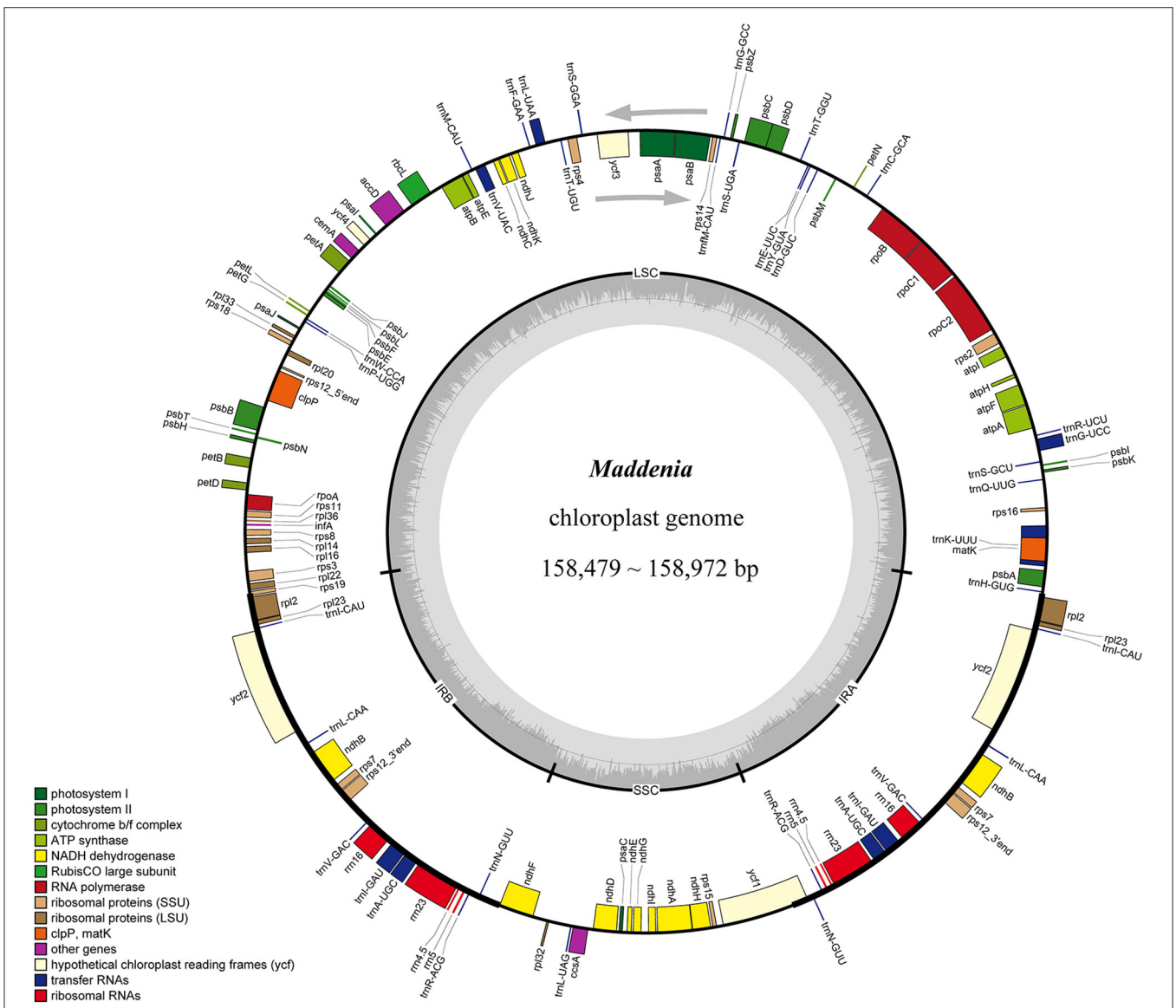


FIGURE 2 | Gene map of *Maddenia* chloroplast genome. The two gray arrows indicate the direction of gene transcription. The dashed area in the inner circle indicates the GC content of the plastome. LSC, large-single-copy; SSC, small-single-copy; IR, inverted repeat.

Characterization of SSRs and Repeat Sequences in Plastomes

We searched for SSRs in seven *Maddenia* species using MISA (Thiel et al., 2003) with the settings at 10, 5, 4, 3, 3, 3 repeat units for mono-, di-, tri-, tetra-, penta-, and hexanucleotide SSRs, respectively. Tandem Repeat Finder (Benson, 1999) was used to analyze the tandem repeat sequences with the default parameters. One inverted repeat sequence was removed before detecting large repeat sequences. We employed REPuter (Kurtz et al., 2001) to identify the large repeat sequences, including forward, reverse, complement, and palindromic repeats. The minimal repeat size and Hamming distance were set at 30 bp and 3, respectively.

Morphological and Micromorphological Characteristics Detection

Images of mature leaves were taken with a Nikon SM225 Stereo microscope (Japan). To show the micromorphological traits, a scanning electron microscope (SEM) was used. The mature leaves were fixed in Formaldehyde-acetic acid-ethanol (FAA) (methanol: acetic acid: ethanol: water = 10:5:50:35), cut into small pieces, and washed in 70% alcohol. Then, they were dehydrated in an increasing alcohol series and iso-amyl acetate series. Afterward, the material was critical-point dried using liquid CO_2 with a K850 critical-point dryer (Quorum). The leaf pieces were then mounted on aluminum stubs and sputter-coated with gold using a JS-1600 sputter coater (HTCY). Photos were taken with a Hitachi S-3400 SEM (Hitachi, Tokyo, Japan).

RESULTS

Characteristics of *Maddenia* Plastomes

We used genome skimming to generate DNA sequence data for 22 samples, including seven *Maddenia* species (17 individuals) and five outgroup species. The size of the *Maddenia* plastomes ranged from 158,479 to 158,972 bp in length. The plastomes of all the *Maddenia* species had a quadripartite structure (**Figure 2**), including a large single-copy region (LSC, 86,939–87,405 bp), a small single-copy region (SSC, 18,862–18,930 bp), and two inverted repeated regions (IRs, 26,292–26,363 bp) (**Table 2**). The total Guanine-cytosine (GC) content of all the *Maddenia* plastomes was 36.6%, but the GC content in IRs (42.5–42.6%) was higher than that in LSC (34.4%) and SSC (30.4–30.5%). All the *Maddenia* plastomes encoded 113 unique genes, including 79 protein-coding genes (CDS), four ribosomal RNAs (rRNAs), and 30 tRNAs. In addition, 17 genes were duplicated in the IRs, of which 6, 4, and 7 encoded proteins, rRNAs, and tRNAs, respectively (**Table 2**). In *Maddenia* plastomes, 14 unique genes had introns, of which two (*ycf3* and *clpP*) had two introns (**Supplementary Table 1**). The genome size, GC content, gene number, and order in all the *Maddenia* plastomes were relatively conserved in comparison to the outgroups (**Table 2**).

Plastome Comparisons

Overall, *Maddenia* plastomes showed high sequence similarity, and non-coding regions had more divergence than coding regions (**Figure 3**). In the Mauve analysis, no rearrangement event was detected among the *Maddenia* plastomes (**Supplementary Figure 1**).

Nucleotide substitution and sequence distance were used to compare the difference of plastomes between the seven *Maddenia* species. Across all individuals, the number of nucleotide substitutions was 0–266 bp and the pairwise sequence distance percentage among the whole plastome sequences was 0–0.00169. The sequence differences between *P. fujianensis*, *P. himalayana*, and *P. gongshanensis* were much higher than those between *P. hypoleuca*, *P. hypoxantha*, *P. incisoserrata,* and *P. wilsonii* (**Table 3**).

IRs Expansion and Contraction

Given that there were no significant differences among the *Maddenia* plastomes (**Supplementary Figures 1**, **2**), *P. wilsonii* WX202 in *Maddenia* was chosen to conduct border comparisons. Six Rosaceae species, i.e., *Physocarpus amurensis* (Maxim.) Maxim. WX230 and *Prinsepia uniflora* Batalin WX231 from Amygdaloideae; *Rosa multiflora* Thunb. (NC_039989) and *Fragaria vesca* L. (NC_015206) from Rosoideae; *Dryas octopetala* var. *Asiatica* (Nakai) Nakai (KY420029) and *Purshia tridentata* (Pursh) DC. (KY420000) from Dryadoideae, were compared to *P. wilsonii* WX202. Variation was detected in the expansion and contraction of IR regions (**Figure 4**). The LSC/IRb borders of Amygdaloideae species and *Purshia tridentata*were located in the *rps19* gene, which extended 81–134 bp into the IRb. In Rosoideae species, the LSC/IRb borders were in the intergenic spacers, and the intact *rps19* gene in the LSC contracted 12–13 bp from the LSC/IRb border. In addition, the SSC/IRs borders were in the *ndhF*/ψ*ycf1* and *ycf1* genes except for the *Rosa multiflora* (only in *ycf1* gene). The IRb/SSC border was located in the pseudogene *ycf1* for *Fragaria vesca* and *Dryas octopetala* var. *asiatica,* and between the pseudogene *ycf1* and *ndhF* gene for *Rosa multiflora*. The *ndhF* gene extended 9–29 bp into the IRb in the Amygdaloideae species and *Purshia tridentata* but was completely located in the SSC for Rosoideae species and *Dryas octopetala* var. *asiatica*. The SSC/IRa border was located in the *ycf1* gene across the Rosaceae species. The gene *trnH* in the LSC contracted 3–324 bp from the border region of IRa/LSC.

Identification of Hypervariable Regions

The Pi values were used to determine hypervariable regions. The result showed that the Pi values in IRs were less than those in LSC and SSC. We chose the regions with relatively higher Pi values as hypervariable regions. A total of eight hypervariable regions were identified, including seven intergenic spacer regions (*trnS–trnG*, *trnR–atpA*, *trnC–petN*, *trnT–trnL*, *ndhC–trnV*, *ndhF–rpl32*, and *rpl32–trnL*) and one protein-coding region (*ycf1*). These sequences were all located in two single-copy regions and none in IR regions (**Figure 5**). The Pi value of eight hypervariable regions ranged from 0.01619–0.03251 (**Table 4**).

Repeat Analyses

A total of 558 SSRs were identified in seven *Maddenia* species, including mono-, di-, tri-, tetra-, and pentanucleotide, but there was no hexanucleotide in all plastomes (**Figure 6B**). *P. fujianensis*

TABLE 2 | Summary information for the plastome sequences of *Maddenia* species.

Taxon	Number	Size (bp)				GC content (%)				Number of genes			
		Total	LSC	SSC	IR	Total	LSC	SSC	IR	Total	CDS	tRNA	rRNA
P. incisoserrata	JR301	158,768	87,212	18,906	26,325	36.6	34.4	30.4	42.5	130 (17)	85 (6)	37 (7)	8 (4)
P. incisoserrata	JR334	158,882	87,347	18,885	26,325	36.6	34.4	30.5	42.5	130 (17)	85 (6)	37 (7)	8 (4)
P. incisoserrata	JR440	158,842	87,310	18,882	26,325	36.6	34.4	30.5	42.5	130 (17)	85 (6)	37 (7)	8 (4)
P. hypoleuca	WX219	158,873	87,338	18,885	26,325	36.6	34.4	30.5	42.5	130 (17)	85 (6)	37 (7)	8 (4)
P. hypoleuca	JR324	158,811	87,299	18,862	26,325	36.6	34.4	30.5	42.5	130 (17)	85 (6)	37 (7)	8 (4)
P. hypoleuca	JR336	158,881	87,346	18,885	26,325	36.6	34.4	30.5	42.5	130 (17)	85 (6)	37 (7)	8 (4)
P. wilsonii	WX202	158,673	87,134	18,889	26,325	36.6	34.4	30.5	42.5	130 (17)	85 (6)	37 (7)	8 (4)
P. wilsonii	JR352	158,751	87,198	18,903	26,325	36.6	34.4	30.4	42.5	130 (17)	85 (6)	37 (7)	8 (4)
P. wilsonii	JR438	158,908	87,373	18,885	26,325	36.6	34.4	30.5	42.5	130 (17)	85 (6)	37 (7)	8 (4)
P. hypoxantha	JR372	158,701	87,167	18,884	26,325	36.6	34.4	30.5	42.5	130 (17)	85 (6)	37 (7)	8 (4)
P. hypoxantha	JR426	158,760	87,206	18,894	26,330	36.6	34.4	30.5	42.5	130 (17)	85 (6)	37 (7)	8 (4)
P. himalayana	JR377	158,524	86,939	18,907	26,339	36.6	34.4	30.4	42.6	130 (17)	85 (6)	37 (7)	8 (4)
P. himalayana	JR385	158,827	87,208	18,893	26,363	36.6	34.4	30.4	42.5	130 (17)	85 (6)	37 (7)	8 (4)
P. gongshanensis	JR305	158,564	87,045	18,921	26,299	36.6	34.4	30.4	42.6	130 (17)	85 (6)	37 (7)	8 (4)
P. gongshanensis	JR381	158,479	86,965	18,930	26,292	36.6	34.4	30.4	42.6	130 (17)	85 (6)	37 (7)	8 (4)
P. fujianensis	JR302	158,972	87,405	18,873	26,347	36.6	34.4	30.5	42.5	130 (17)	85 (6)	37 (7)	8 (4)
P. fujianensis	JR386	158,954	87,390	18,870	26,347	36.6	34.4	30.5	42.5	130 (17)	85 (6)	37 (7)	8 (4)
P. discadenia	WX203	157,915	85,946	19,119	26,415	36.7	34.6	30.2	42.5	130 (17)	85 (6)	37 (7)	8 (4)
P. serotina	WX204	158,760	87,180	18,862	26,358	36.6	34.4	30.4	42.6	130 (17)	85 (6)	37 (7)	8 (4)
P. kansuensis	WX207	157,660	85,764	19,122	26,387	36.8	34.6	30.3	42.6	130 (17)	85 (6)	37 (7)	8 (4)
Physocarpus amurensis	WX230	159,110	87,563	18,827	26,360	36.4	34.1	30.0	42.6	130 (17)	85 (6)	37 (7)	8 (4)
Prinsepia uniflora	WX231	159,186	87,236	19,174	26,388	36.6	34.4	30.2	42.7	130 (17)	85 (6)	37 (7)	8 (4)

The number in braces means duplicated gene number in the IR region.

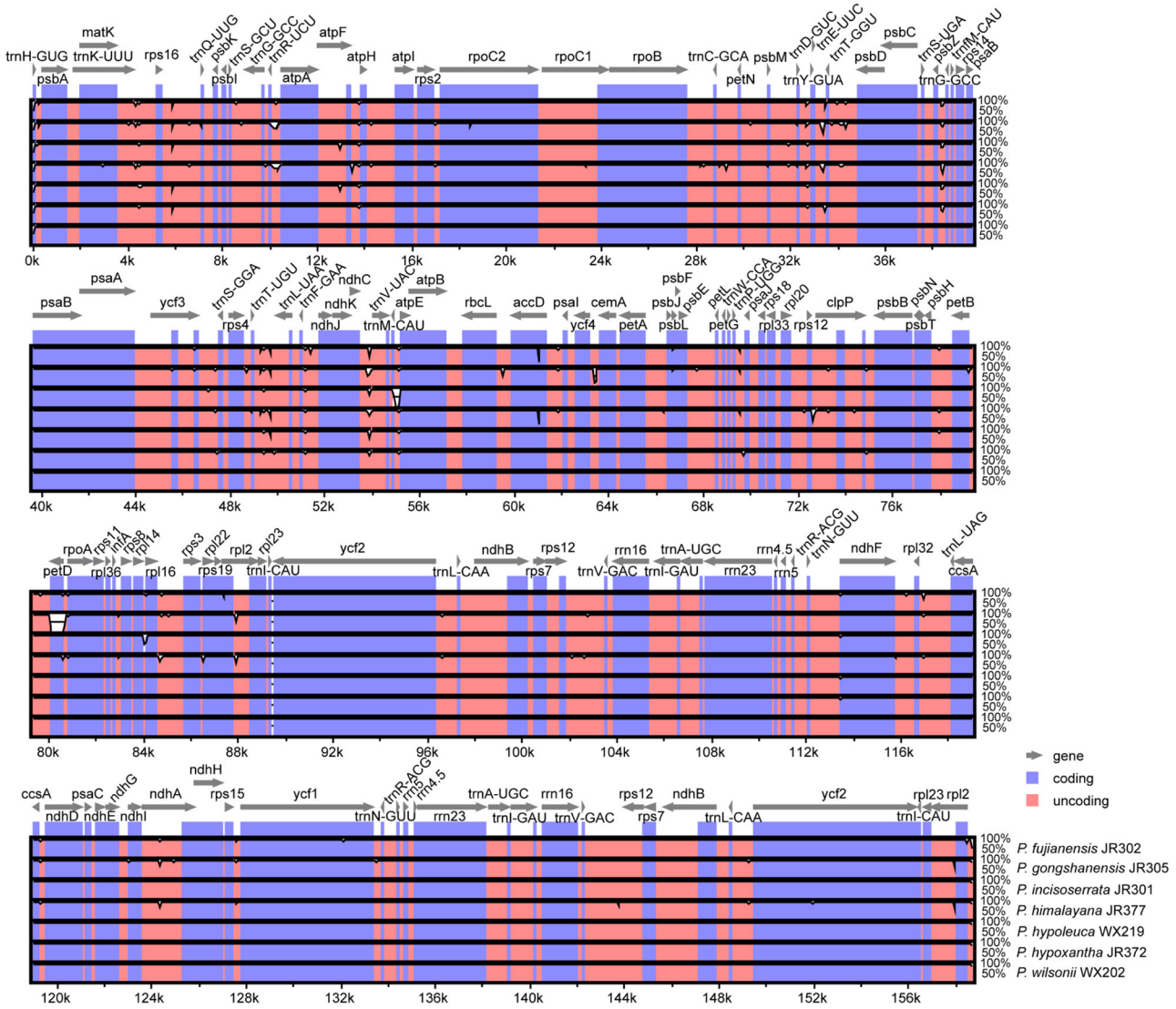


FIGURE 3 | Visualization of alignment of the seven *Maddenia* chloroplast genome sequences using mVISTA. *P. wilsonii* WX202 was used as a reference sequence. Blue represents coding regions, and pink represents non-coding regions.

showed the most SSRs, followed by *P. hypoxantha*, *P. wilsonii*, *P. hypoleuca*, *P. incisoserrata,* and *P. himalayana*, while the SSR numbers of *P. gongshanensis* were the least (**Figure 6B**). There were many SSR motif types in the *Maddenia* plastomes, but most of them had few SSRs and only two types (A/T and AT/TA) contained more SSRs (**Supplementary Figure 3A**). The lengths of SSRs ranged from 10 to 24 bp (**Figure 6C**).

In *Maddenia* plastomes, the repeat sequences included forward, reverse, palindromic, complement, and tandem repeats. *P. incisoserrata* contained the most repeat sequences, *P. himalayana* had the least, and *P. gongshanensis* had no tandem repeats (**Figure 6D**). The most common type of repeat sequences by length was 30–34 bp (**Figure 6A**).

Most of the SSRs and repeat sequences were located in the LSC, followed by the SSC and IRs (**Figure 6E**). In addition, repeat sequences were mainly distributed in the intergenic spacers (IGS), but some were also found in the CDS and intron regions (**Supplementary Figure 3B**).

Phylogenetic Analysis

To resolve the phylogenetic relationship of the *Maddenia* clade, different trees were reconstructed based on complete plastomes and SCN genes. For plastome data, all the trees had identical topology except SSC and IR (**Figure 7A** and **Supplementary Figure 4**). In the rest of this section, the tree based on complete plastomes will be used to discuss the phylogenetic relationships of the *Maddenia* group, which was monophyletic and was separated into three subclades with high support values. Subclade I only include one species (*P. fujianensis*) from Fujian Province of eastern China. Subclade II is sister to subclade III, and together they are both sister to subclade I with a posterior probability of 1.00 Subclade II

TABLE 3 | Numbers of nucleotide substitutions and sequence distance in *Maddenia* plastomes.

	***P. fujianensis* JR302**	***P. fujianensis* JR386**	***P. gongshanensis* JR305**	***P. gongshanensis* JR381**	***P. himalayana* JR377**	***P. himalayana* JR385**	***P. hypoleuca* WX219**	***P. hypoleuca* JR324**	***P. hypoleuca* JR336**	***P. hypoxantha* JR372**	***P. hypoxantha* JR426**	***P. incisoserrata* JR301**	***P. incisoserrata* JR334**	***P. incisoserrata* JR440**	***P. wilsonii* WX202**	***P. wilsonii* JR352**	***P. wilsonii* JR438**
P. fujianensis JR302		0.00000629	0.00140	0.00145	0.00163	0.00105	0.00093	0.00095	0.00095	0.00094	0.00107	0.00088	0.00095	0.00095	0.00103	0.00086	0.00096
P. fujianensis JR386	1		0.00139	0.00145	0.00162	0.00102	0.00093	0.00095	0.00094	0.00093	0.00107	0.00087	0.00094	0.00093	0.00102	0.00087	0.00093
P. gongshanensis JR305	222	220		0.00020	0.00160	0.00090	0.00128	0.00139	0.00130	0.00142	0.00142	0.00130	0.00130	0.00093	0.00144	0.00130	0.00135
P. gongshanensis JR381	230	230	32		0.00169	0.00101	0.00138	0.00143	0.00134	0.00145	0.00147	0.00134	0.00134	0.00138	0.00149	0.00134	0.00139
P. himalayana JR377	257	256	253	266		0.00123	0.00156	0.00163	0.00152	0.00161	0.00168	0.00154	0.00152	0.00154	0.00164	0.00154	0.00154
P. himalayana JR385	166	161	143	160	194		0.00100	0.00110	0.00099	0.00101	0.00109	0.00092	0.00099	0.00094	0.00105	0.00090	0.00097
P. hypoleuca WX219	148	147	202	218	246	158		0.00024	0	0.00023	0.00037	0.00015	0	0.00017	0.00103	0.00014	0.00018
P. hypoleuca JR324	151	151	220	226	258	174	38		0.00025	0.00027	0.00046	0.00023	0.00025	0.00020	0.00037	0.00022	0.00021
P. hypoleuca JR336	150	149	205	212	240	157	0	39		0.00023	0.00037	0.00015	0	0.00018	0.00025	0.00014	0.00019
P. hypoxantha JR372	149	147	225	229	254	160	36	43	36		0.00045	0.00021	0.00023	0.00018	0.00036	0.00014	0.00019
P. hypoxantha JR426	170	169	224	233	266	173	58	73	58	72		0.00037	0.00037	0.00040	0.00030	0.00036	0.00042
P. incisoserrata JR301	140	138	206	212	244	145	24	37	24	34	58		0.00015	0.00016	0.00026	0	0.00017
P. incisoserrata JR334	150	149	205	212	240	157	0	39	0	36	58	24		0.00018	0.00025	0.00014	0.00019
P. incisoserrata JR440	150	147	147	218	243	149	27	31	28	28	63	25	28		0.00030	0.00015	0.00001
P. wilsonii WX202	163	162	228	235	259	167	163	58	40	57	48	41	40	47		0.00025	0.00030
P. wilsonii JR352	137	138	205	211	243	143	23	35	23	23	57	0	23	24	40		0.00016
P. wilsonii JR438	152	148	213	220	244	153	29	33	30	30	66	27	30	1	48	25	

The upper triangle shows the number of nucleotide substitutions and the lower triangle indicates the number of sequence distances in the Maddenia plastomes.

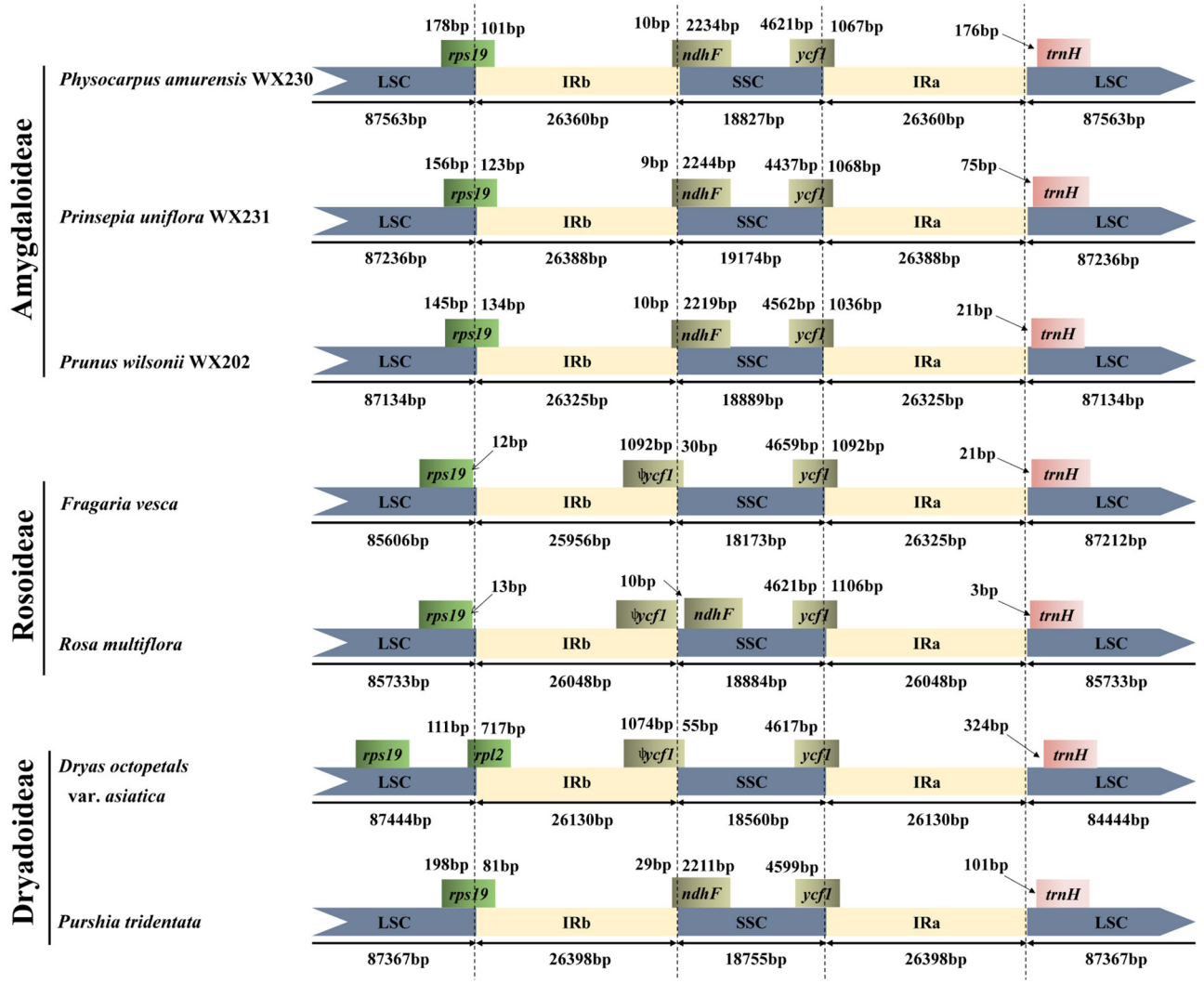


FIGURE 4 | Comparison of the LSC, IRs, and SSC border regions of *Maddenia* group and other Rosaceae plastomes.

consists of *P. gongshanensis* from Yunnan Province (China) and *P. himalayana* from Tibet (China) and the adjacent Himalayan region. For the individuals sampled, the two species are reciprocally monophyletic. Subclade III is composed of four former species with a posterior probability of 1.00 and bootstrap value of 100%, in which samples from the same geographical position were grouped, although the four species were each not clearly identified.

To save costs of sequencing for further investigations on *Maddenia*, we also tried to explore four standard DNA barcodes to identify *Maddenia* species. Concatenated *rbcL*, *matK*, and *trnH-psbA* and ITS datasets were used to construct the phylogenetic trees of *Maddenia*, respectively. The trees constructed by the standard DNA barcodes were not congruent with those reconstructed using the complete plastomes (**Figures 7C,D**). In the phylogenic tree based on concatenated *rbcL*, *matK*, and *trnH-psbA*, although the *Maddenia* species formed a clade, *P. fujianensis* was sister to subclade III rather than to the remaining *Maddenia* species. In addition, the two *P. himalayana* individuals did not group, and subclade III exhibited more polytomies than the tree based on complete plastomes. For the tree based on ITS sequences, there were many polytomies and the interspecific relationships within *Maddenia* were poorly resolved.

Eight concatenated hypervariable regions ("specific barcodes"; see Discussion) were also employed to reconstruct the phylogenetic relationship of the *Maddenia* clade. We found that the topology based on the hypervariable regions was similar to that of complete plastomes, though there were lower support values at some branches (**Figure 7B**).

The recovery efficiency of each SCN gene is shown in **Figure 8**. The quality of the nuclear genes recovered was relatively high. In total, we got 446 SCN genes from raw data. We also filtered out genes with <80% samples, leaving

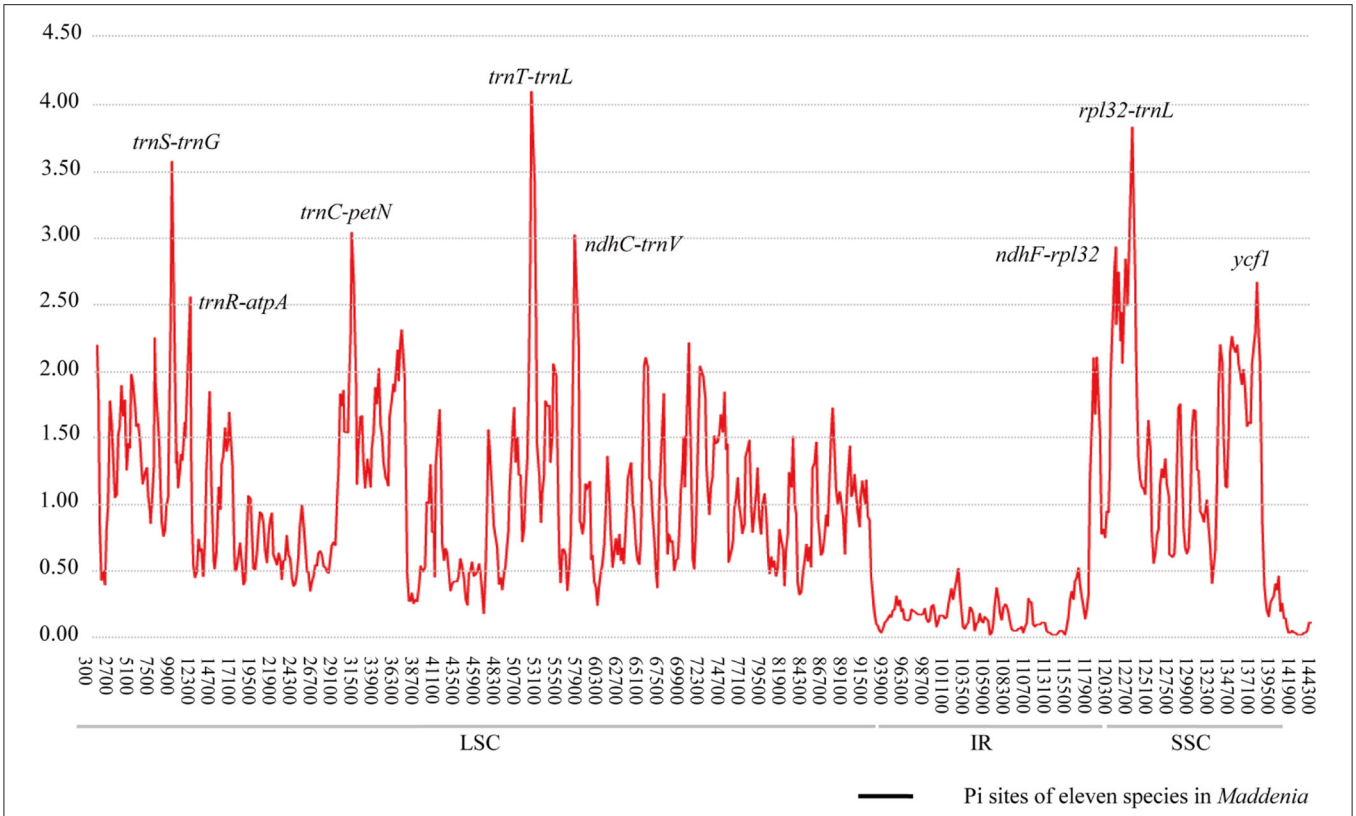


FIGURE 5 | Sliding window analysis of the plastomes of samples.

TABLE 4 | Sequence characteristics of eight high variable regions among 22 plastomes.

Region	Aligned length	Variable sites		Indels		Nucleotide diversity (Pi)
		No.	%	No.	Length range	
trnS-trnG	711	57	8.01	32	1–301	0.02656
trnR-atpA	582	27	3.63	44	1–146	0.03251
trnC-petN	1,075	155	14.41	36	1–44	0.02258
trnT-trnL	1,367	135	9.87	77	1–114	0.02468
ndhC-trnV	772	83	10.75	49	1–160	0.02798
ndhF-rpl32	1,241	172	13.85	60	1–174	0.02746
rpl32-trnL	1,774	163	9.18	78	1–263	0.02570
ycf1	5,730	702	12.25	36	1–30	0.01619

413 SCN genes with more than 600 bp in length. For the tree generated by 446 SCN genes data, the *Maddenia* clade was monophyletic but its deep nodes were not resolved well (**Figure 9**). It was obvious that there were five subclades in the *Maddenia* internal clade according to their geographic positions (e.g., Fujian Province, Yunnan Province, Tibet, Sichuan Province, and Qingling Mountain). Subclades A, B, and C comprise *P. fujianensis*, *P. gongshanensis*, and *P. himalayana*, respectively. The monophyly of each of the three species was well-supported, which was congruent with that of 413 SCN genes (**Supplementary Figure 5**). Subclade D consists of samples of *P. hypoxantha* and one individual of *P. wilsonii*. Subclade E contains *P. hypoleuca*, *P. incisoserrata*, and two individuals of *P. wilsonii*. However, these four species are not identified clearly.

Phylogenetic Network Analyses of Nuclear Data With SNaQ

The optimal h_{max} value inferred in the SNaQ was six, which has the highest pseudolikelihood network score (−676.478, **Supplementary Figure 6**). The phylogenetic network analyses

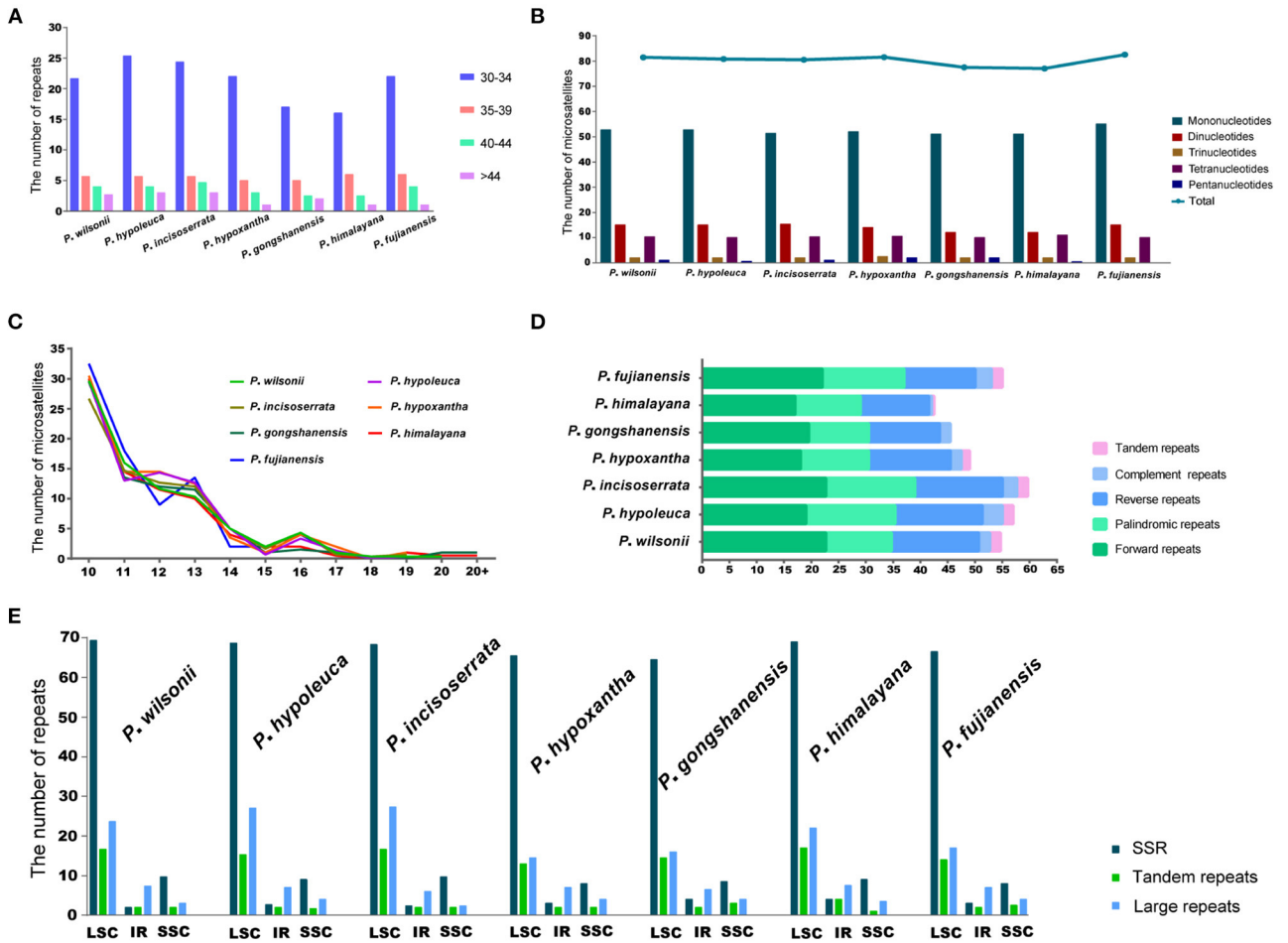


FIGURE 6 | Analyses of simple sequence repeats (SSRs) and repeated sequences in plastomes of *Maddenia* species. **(A)** Frequency of four repeats by length; **(B)** Number of five repeat types; **(C)** Frequency of microsatellites by length; **(D)** Numbers of five different types of repeats; **(E)** Number of all repeats by location.

showed that there were widespread hybridization events within the *Maddenia* group (**Figure 10**). The two *P. himalayana* individuals were 74.4% sister to each other and 25.6% sister to *P. gongshanensis* JR381. The group (*P. fujianensis* JR302+*P. hypoxantha* JR426+ *P. incisoserrata* JR334) was 66.1% sister to (*P. fujianensis* JR386+ *P. incisoserrata* JR301) and 33.9% sister to *P. wilsonii* WX202. *P. incisoserrata* JR334 was 53.5% sister to *P. hypoxantha* JR426 and 46.5% sister to *P. fujianensis* JR302. *P. wilsonii* JR438 was 95.8% sister to *P. hypoleuca* WX219 and 4.21% sister to *P. hypoleuca* JR336.

Morphological and Micromorphological Traits

The mature leaves of the *Maddenia* species are green to deep green adaxially. The shapes of the leaves and leaf bases are multiple (**Figure 11A**$_{1-7}$). The leaf margins are serrulate, irregularly serrate, or doubly serrate. A few glandular teeth were found at the leaf bases of *P. hypoleuca*, *P. incisoserrata*, *P. wilsonii*, *P. hypoxantha*, and *P. fujianensis* (**Figure 11B**$_{1-5}$), while many glandular teeth grow at the lower margins of *P. gongshanensis* (**Figure 11A**$_6$). For *P. himalayana*, leaf margins have less glandular on the foliage branch while, in the reproductive branch, the glandular teeth are distributed abundantly near the bases (**Figure 11B**$_7$).

The most notable morphological character differentiating species in *Maddenia* is the hair distribution on the abaxial leaf surface. In *P. fujianensis*, few hairs were found at the axils between midvein and secondary veins, and there is no hair on the intercostal area (**Figure 11C**$_1$**,D**$_1$). The intercostal area of *P. hypoleuca* and *P. incisoserrata* are also glabrous, but the distribution pattern of hairs at the axils shows high diversity (**Figure 11C**$_{2,3}$**,D**$_{2,3}$). There were no hairs or only a few hairs growing on the bases of secondary veins, or there was a cluster of hairs growing on the axil. Such three situations could be found on a single blade of the leaf. In *P. wilsonii* and *P. hypoxantha* hairs grow all along the veins, but the hairs are present on veinlets in the intercostal area of *P. wilsonii*, which distinguishes it from *P. hypoxantha* (**Figure 11C**$_{4,5}$**,D**$_{4,5}$). Hairs were also observed at the axils

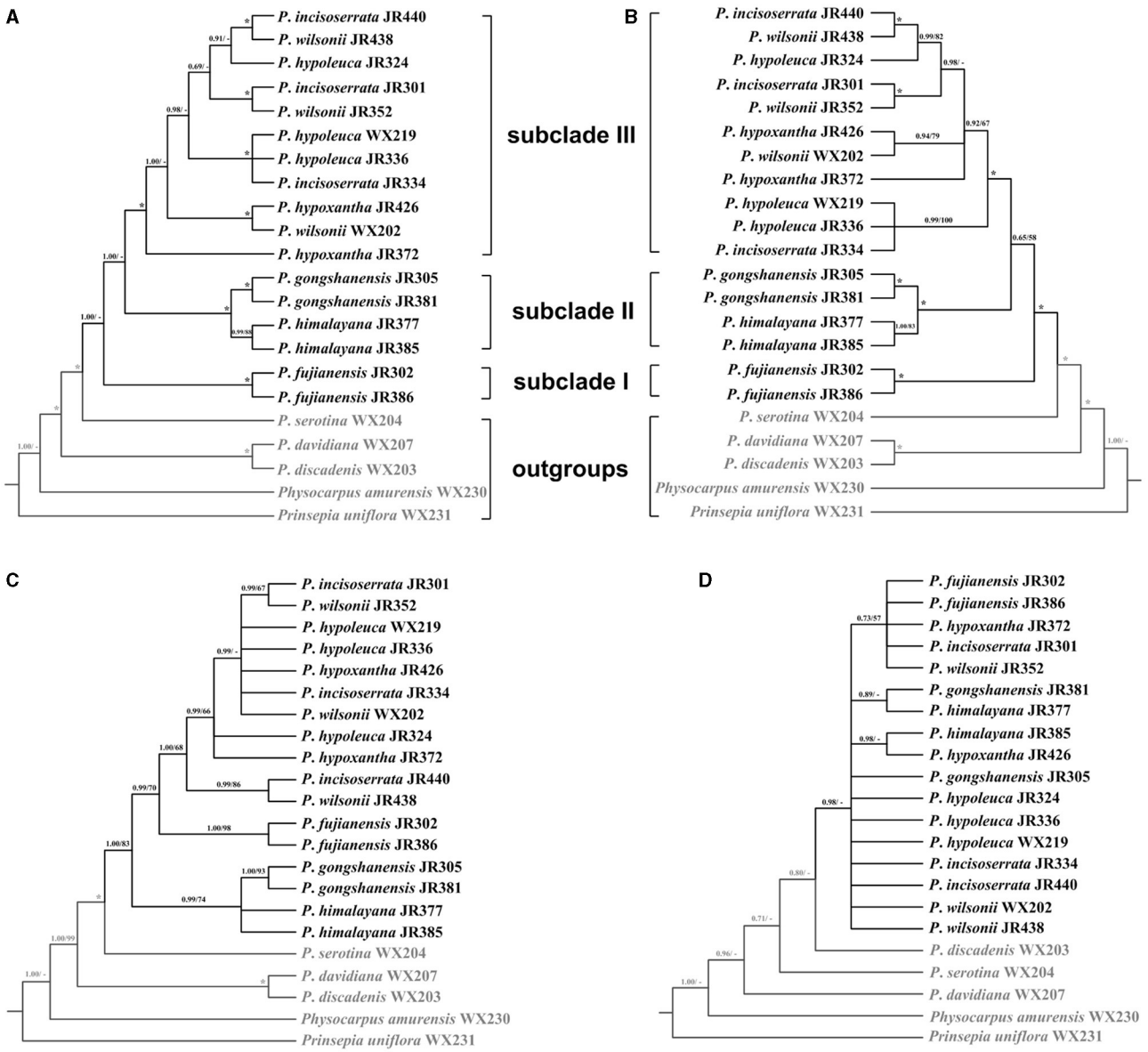


FIGURE 7 | Phylogenetic relationships of the *Maddenia* group of *Prunus* inferred from Bayesian inference (BI) and maximum likelihood (ML) based on four datasets. **(A)** complete plastomes; **(B)** hypervariable regions; **(C)** *rbcL*+*matK*+*trnH-psbA*; **(D)** internal transcribed spacer (ITS) sequence. The support values above the branches show PP (posterior probability) / BS (bootstrap support), and asterisks indicate 1.00/100%. Dashes represent incongruences of BI and ML trees.

of *P. gongshanensis*, and sometimes there are a few hairs on the midvein at the base (**Figure 11C$_6$,D$_6$**). Leaves of *P. himalayana* are densely pubescent on the abaxial side (**Figure 11C$_7$,D$_7$**).

In all seven species, stomata are found only on the abaxial surface, and each of them consists of a pair of guard cells encircled by several other cells (**Figure 11E$_{1-7}$**). Distinct circular ornamentations were found on the cell wall of guard cells in *P. hypoleuca* and *P. incisoserrata* (**Figure 11E$_{2,3}$**). In the other species, such ornamentations are relatively obscure or nearly inexistent (**Figure 11E$_{1,4-7}$**).

DISCUSSION

Comparative Plastomes of *Maddenia*

All sequenced *Maddenia* plastomes share a typical quadripartite structure, which is similar to most photosynthetic angiosperms (Jansen and Ruhlman, 2012; Abdullah et al., 2019; Xu et al., 2019). However, the loss of one complete IR region also occurred in some taxa, such as the inverted-repeat-lacking clade of Fabaceae (Wang et al., 2017), *Erodium* of Geraniaceae (Guisinger et al., 2011), and *Carnegiea* of Cactaceae (Sanderson et al., 2015). In addition, the GC content in the IRs was higher than that in LSC and SSC, which is due to the presence of rRNA genes with high

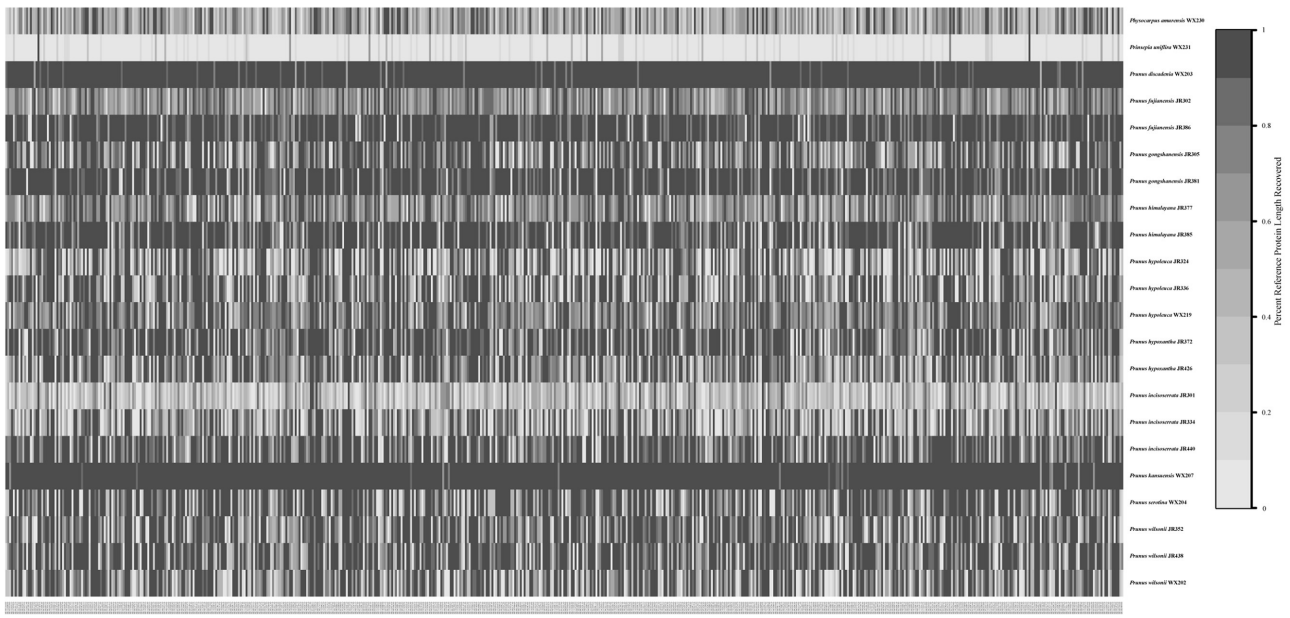


FIGURE 8 | Heat map showing recovery efficiency for 591 genes enriched in the *Maddenia* group recovered by HybPiper. Each column is a gene, and each row is one sample. The shade of gray in the cell is determined by the length of sequence recovered by the pipeline, divided by the length of the reference gene (maximum of 1).

GC content (Kim and Lee, 2004). The conserved genome size, GC content, and gene number of *Maddenia* plastomes resemble other Amygdaloideae species (Wang et al., 2013; Kim et al., 2018). Although gene rearrangement events have been reported in some genera of other families, such as *Lasthenia* of Asteraceae (Walker et al., 2014), *Anemone* of Ranunculaceae (Liu et al., 2018), and *Passiflora* of Passifloraceae (Rabah et al., 2018), we observed no such events in *Maddenia* plastomes (**Supplementary Figure 1**).

The expansion and contraction of the IR region have an impact on the plastome size to some extent (Jansen and Ruhlman, 2012). Expansion events caused several genes in SC regions to enter the IR region. However, small IR expansions and contractions have a much higher frequency than large ones in seed plants (Goulding et al., 1996; Downie and Jansen, 2015). For *Maddenia*, no significant variation and slight IR expansions and contractions exist in every border of plastomes (**Supplementary Figure 2**), demonstrating their conserved traits. Nevertheless, compared with other Rosaceae species, we observe that Rosoideae plastomes have a partial *rps19* gene in the LSC region but an intact *rps19* gene in the LSC of Amygdaloideae. Variations in the location of the *rps19* gene were also documented in other Rosaceae species (e.g., Wang et al., 2013; Kim et al., 2018). However, our results indicate that the location of the *rps19* gene is not useful in distinguishing the three subfamilies, since the two Dryadoideae plastomes we analyzed showed two different locations of the *rps19* gene: one matching Amygdaoideae and the other matching a member of Rosoideae.

The SSRs are effective molecular markers to population genetic and phylogenetic studies in plants (Powell et al., 1995; Doorduin et al., 2011; Zhang X. et al., 2017; Sun et al., 2020). A total of 558 SSRs were identified in seven *Maddenia* species (**Figure 6**). The SSRs motif type (A/T) was quite common and most of SSRs were located in the intergenic spacers, which were similar to other Rosaceae species (Wang et al., 2013). In addition, previous studies supported that the region rich in A/T had the most repeats and indels (Cai et al., 2008). Thus, *Maddenia* SSRs could be further utilized for population genetics research in the future.

Phylogenetic Analyses and Implications for Species Delimitation

Our results provided strong support for the monophyly of the *Maddenia* clade based on both plastome and nuclear datasets. The phylogenetic trees based on complete plastomes and nuclear datasets divided *Maddenia* into three and five major subclades, respectively. Subclade I in the plastome tree (= subclade A in the nuclear tree) included one species (*P. fujianensis*) only distributed at Wuyi Mountain of Fujian Province in Southeast of China. *Maddenia fujianensis* has been treated as a synonym of *P. hypoleuca*, for the differentiating characters between them are continuous (Wen and Shi, 2012). *P. fujianensis* was, however, sister to all remaining species of the *Maddenia* group based on plastome data. Even though this relationship was not congruent with that of SCN genes (**Figure 9**), the monophyly of *P. fujianensis* was well-supported. The SNaQ analyses suggested one hybridization event between *P. fujianensis* JR302 and *P. incisoserrata* JR334 (**Figure 10**). Therefore, it is likely that *P. fujianensis* may be a cryptic species, even though it is morphologically and micromorphologically similar to *P. hypoleuca* (**Figure 11**). More attention should be focused on the origin of *P. fujianensis* in our future studies, sampling *P. hypoleuca* broadly across its entire distribution range.

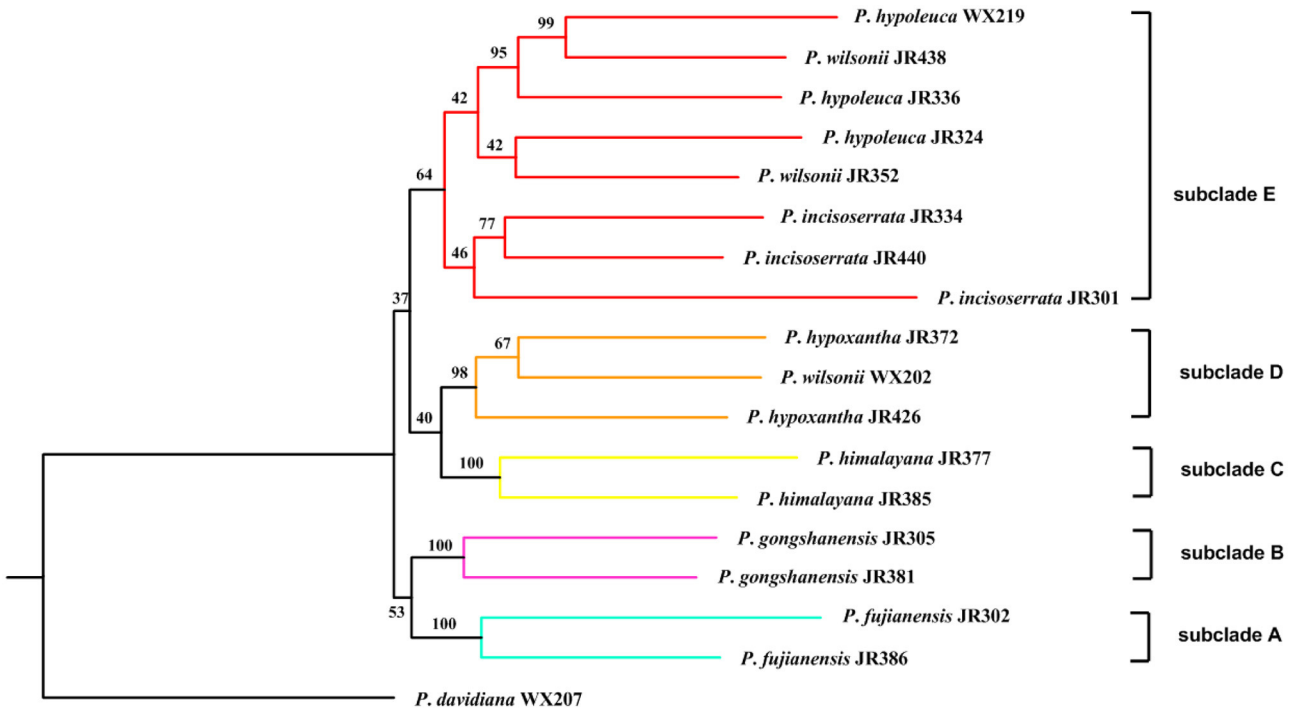


FIGURE 9 | Maximum likelihood inferred from 446 single-copy nuclear (SCN) genes of *Maddenia* group of *Prunus*. The number above branch indicated bootstrap support from the IQ-TREE.

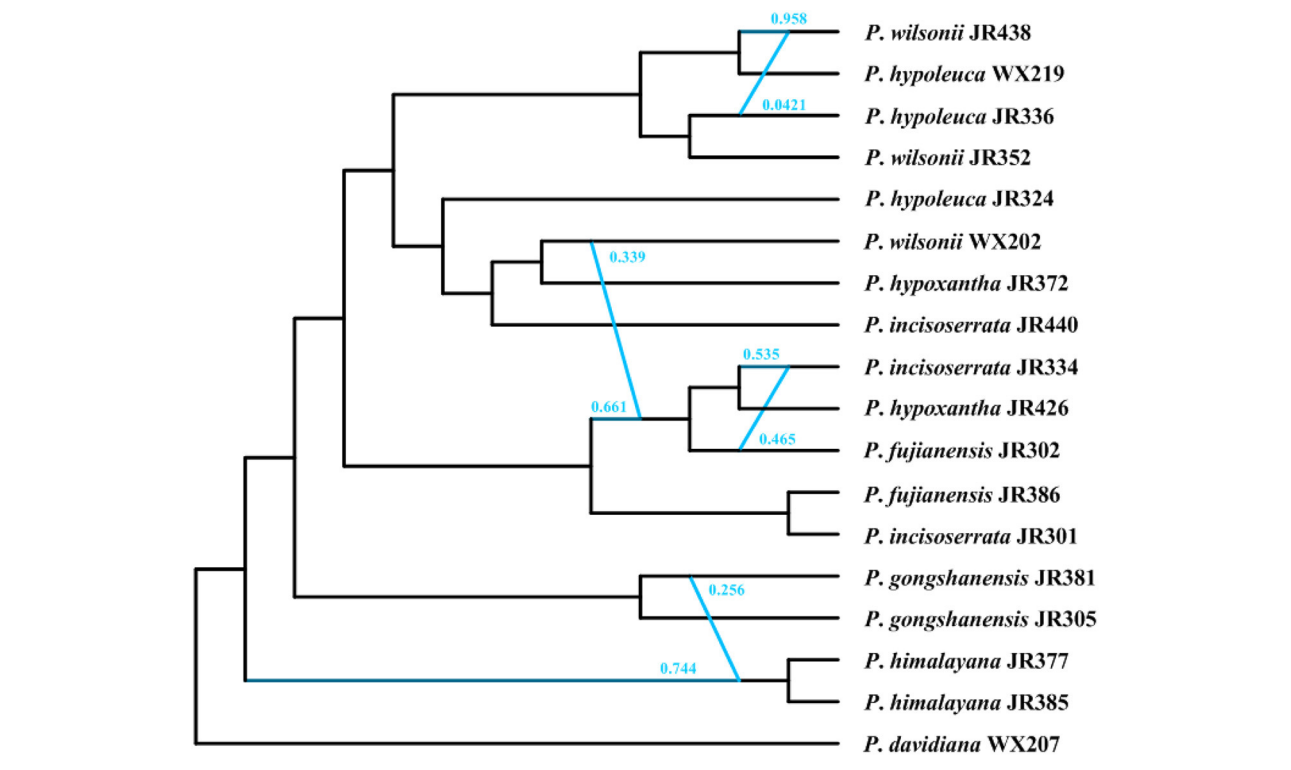


FIGURE 10 | The phylogenetic network inferred in the SNaQ/PhyloNetworks analysis. The blue line indicated the hybridization events in the network.

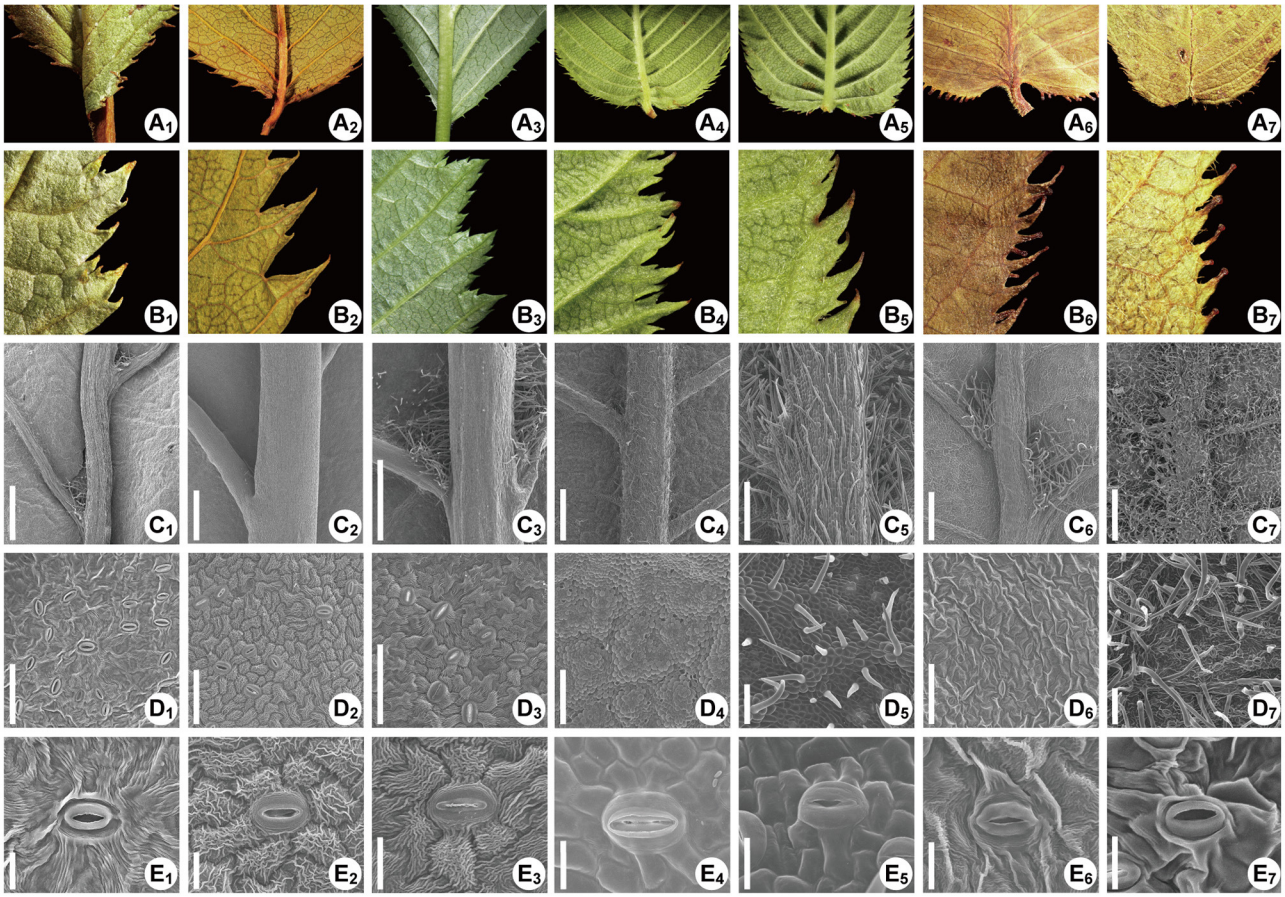


FIGURE 11 | The leaf morphology of the *Maddenia* species. The species are *Prunus fujianensis*, *Prunus hypoleuca*, *P. incisoserrata*, *Prunus hypoxantha*, *P. wilsonii*, *Prunus gongshanensis,* and *Prunus himalayana*, respectively in each row from left to right. **(A_1-A_7)** Leave bases, showing the glandular teeth; **(B_1-B_7)** Partial views of leaf margins, showing the shapes of teeth and glandular teeth; **(C_1-C_7)** Abaxial leaves, showing the pubescence distribution; **(D_1-D_7)** Detailed view of the pubescence on intercostal areas; **(E_1-E_7)** Stomata and guard cells, showing the ornamentation.

Subclade II in the plastome tree consists of *P. gongshanensis* and *P. himalayana* (= subclades B and C, respectively, in the nuclear tree) distributed in Yunnan Province (Northwest Hengduan Mountains) and Southeast Tibet, respectively. *P. gongshanensis* is characterized by subcordate to cordate leaf bases and glabrous leaf surfaces. *P. himalayana* stands out by its abaxially dense pubescent leaf blade. Although these two species can be identified based on both molecular, morphological, and micromorphological evidence, one hybridization event between them was detected in the SNaQ analyses (**Figures 10**, **11**).

Subclade III in the plastome tree (comprising subclades D and E in the nuclear tree) is composed of the remaining four species, but their relationships were not resolved. Interestingly, most different species from the same geographical area were grouped, such as *P. hypoxantha* JR426 and *P. wilsonii* WX202 from the Emei Mountain of Sichuan Province and *P. hypoxantha* JR372 from Kangding of Sichuan Province, and all others from the Qinling Mountain region. Meanwhile, the nuclear tree showed that these four species were divided into two groups according to their geographical distribution. Therefore, we assume that *Maddenia* subclade III might represent two species, i.e. *P. hypoleuca* and *P. hypoxantha*, which is congruent with the treatment of Wen and Shi (2012). *P. incisoserrata* may best be merged with *P. hypoleuca* because they cannot be reliably distinguished by either molecular or morphological evidence (**Figure 11**). Although *P. hypoxantha* and *P. wilsonii* can be identified, to some extent, by the distribution of pubescence on the abaxial leaf surface (i.e., pubescence only on veins vs. denser pubescence on veins (**Figure 11**), we conclude that the latter should be treated as a synonym of the former due to the unsolved relationship between them in the various phylogenetic trees based on our results. Moreover, the sequence differences among species of subclade III of chloroplast tree are minimal (**Table 3**). Gene flow might be widespread within these two species as detected by the SNaQ analyses (**Figure 10**). Future studies should aim to explore the speciation history of subclade III using a broader populational sampling scheme.

On Specific Barcoding of *Maddenia*

Considering the limitations of the standard DNA barcodes and the higher cost of super barcoding, an alternative approach known as "specific barcoding" was proposed, which combined the advantages of the other two (Li et al., 2015). Specific barcoding uses the sequences in target plastomes with high mutation rates. Compared to standard DNA barcodes, specific barcoding is more applicable for differentiation among closely related taxa (Li et al., 2015). We detected eight hypervariable regions among 22 individual plastomes, most of which are located in intergenic spacers (**Figure 5**). This result as well as those from mVISTA, SSRs, and repeat sequence analyses support that the intergenic spacers harbor the highest levels of variation in plastomes. High variability regions in the intergenic spacer have been reported in other studies and shown excellent discriminating ability, such as *Echinacea* of Asteraceae (Zhang N. et al., 2017), *Rhodiola* of Crassulaceae (Zhao et al., 2019), and *Pulsatilla* of Ranunculaceae (Li et al., 2020). Therefore, developing specific barcodes in the intergenic spacers is well-founded and should provide a reliable approach to assess the phylogenetic relationships and identification among *Maddenia* species.

The tree estimated from the specific barcoding regions (**Figure 7B**) had a similar topology with that of the complete chloroplast genomes. However, the sister relationship between subclades II and III was relatively weakly supported (posterior probability of 0.65 and bootstrap support of 58%). At the same time, the gene flow between the species of subclade III within *Maddenia* is active. The utility of these barcodes from chloroplast genomes would be limited for this group. More sampling in population-scale and high-throughput sequencing nuclear data such as RAD or whole-genome resequencing are needed to further explore the relationships of the species of subclade III.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article.

AUTHOR CONTRIBUTIONS

LZ and JW planned and designed the research. NS, B-BL, J-RW, CR, and R-CT performed the experiments and analyzed the data. NS, B-BL, J-RW, R-CT, CR, Z-YC, LZ, DP, and JW wrote the manuscript. All authors approved the final manuscript.

FUNDING

This project was supported by the National Natural Science Foundation of China (Nos. 32170381, 31770200, 32000163, and 31300158) and the Chinese Universities Scientific Fund (No. 2452020179).

ACKNOWLEDGMENTS

We are very grateful to Fuzhen Guo, Xiaohua He, Minrong Luo, Ningjuan Fan, and Guoyun Zhang of Northwest A&F University for their assistance with the SEM and LM. We sincerely thank Prof. Rong Li for the sample collection and Prof. Zhong-hu Li for his help with data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.743643/full#supplementary-material

Supplementary Figure 1 | Structural variation between plastomes of *Maddenia* and outgroups revealed by Mauve.

Supplementary Figure 2 | Comparison of the large-single-copy (LSC), inverted repeats (IRs), and small-single-copy (SSC) border regions of *Maddenia* plastomes.

Supplementary Figure 3 | Microsatellites number of different base compositions and distribution of all repeated sequences. **(A)** Frequency of microsatellites by base composition; **(B)** Number of all repeats by location. CDS, coding sequences; IGS, intergenic spacers.

Supplementary Figure 4 | Phylogenetic relationships of *Maddenia* inferred from Bayesian inference (BI) and maximum likelihood (ML) based on six datasets. **(A)** complete plastomes; **(B)** coding regions; **(C)** LSC region; **(D)** non-coding regions; **(E)** SSC region; **(F)** IR region. The support values above the branches show PP (posterior probability) / BS (bootstrap support), and asterisks indicate 1.00/100%. Dashes represent incongruences of BI and ML trees.

Supplementary Figure 5 | Maximum likelihood inferred from 413 single-copy nuclear (SCN) genes of *Maddenia* group of *Prunus*. The number above branch indicated bootstrap support from the IQ-TREE.

Supplementary Figure 6 | The pseudolikelihood network score from the Species Networks applying Quartets (SNaQ) analysis for each of the maximum number of hybridizations allowed (h_{max}).

Supplementary Table 1 | Genes present in the 17 complete *Maddenia* chloroplast genomes.

REFERENCES

Abdullah, A., Shahzadi, I., Mehmood, F., Ali, Z., Malik, M. S., Wassem, S., et al. (2019). Comparative analyses of chloroplast genomes among three *Firmiana* species: identification of mutational hotspots and phylogenetic relationship with other species of Malvaceae. *Plant Gene* 19:100199. doi: 10.1016/j.plgene.2019.100199

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI–BLAST: a new generation of protein database search program. *Nucl. Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Andrews, S. (2018). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed March 20, 2021).

Andro, M. C., and Riffaud, J. P. (1995). *Pygeum africana* extract for the treatment of patients with benign prostatic hyperplasia: a review of 25 years of published experience. *Curr. Ther. Res.* 56, 796–817. doi: 10.1016/0011-393X(95)85063-5

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucl. Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

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

Cai, C. N., Ma, H., Ci, X. Q., Conran, J. G., and Li, J. (2021). Comparative phylogenetic analyses of Chinese *Horsfieldia* (Myristicaceae) using complete chloroplast genome sequences. *J. Syst. Evol.* 59, 504–514. doi: 10.1111/jse.12556

Cai, Z. Q., Guisinger, M., Kim, H. G., Ruck, E., Blazier, J. C., McMurtry, V., et al. (2008). Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *J. Mol. Evol.* 67, 696–704. doi: 10.1007/s00239-008-9180-7

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. doi: 10.1093/bioinformatics/btp348

Chang, Y. T. (1985). A new species of *Maddenia* (Rosaceae) from Fujian. *Guihaia* 5, 25–26.

Chin, S. W., Shaw, J., Haberle, R., Wen, J., and Potter, D. (2014). Diversification of almonds, peaches, plums and cherries–molecular systematics and biogeographic history of *Prunus* (Rosaceae). *Mol. Phylogenet. Evol.* 76, 34–38. doi: 10.1016/j.ympev.2014.02.024

Chin, S. W., Wen, J., Johnson, G., and Potter, D. (2010). Merging *Maddenia* with the morphologically diverse *Prunus* (Roseaceae). *Bot. J. Linn. Soc.* 163, 236–245. doi: 10.1111/j.1095-8339.2010.01083.x

China Plant BOL Group (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proc. Natl. Acad. Sci. U. S. A.* 108, 19641–19646. doi: 10.1073/pnas.1104551108

Darling, A. E., Mau, B., and Perna, N. T. (2010). Progressive Mauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* 5:e11147. doi: 10.1371/journal.pone.0011147

Do, H. D. K., Kim, C., Chase, M. W., and Kim, J. H. (2020). Implications of plastome evolution in the true lilies (monocot order Liliales). *Mol. Phylogenet. Evol.* 148:106818. doi: 10.1016/j.ympev.2020.106818

Doorduin, L., Gravendeel, B., Lammers, Y., Ariyurek, Y., Chin-A-Woeng, T., and Vrieling, K. (2011). The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. *DNA Res.* 18, 93–105. doi: 10.1093/dnares/dsr002

Downie, S. R., and Jansen, R. K. (2015). A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Syst. Bot.* 40, 336–351. doi: 10.1600/036364415X686620

Doyle, J., and Doyle, J. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.

Erickson, D. L., Spouge, J., Resch, A., Weigt, L. A., and Kress, J. W. (2008). DNA barcoding in land plants: developing standards to quantify and maximize success. *Taxon* 57, 1304–1316. doi: 10.1002/tax.574020

Focke, W. O. (1894). "Rosaceae," in *Dienatürlichen Pflanzenfamilien Volume 8*, eds A. Engler and K. Prantl (Leipzig: Engelmann), 1–61.

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). Vista: computational tools for comparative genomics. *Nucl. Acids Res.* 32, 273–279. doi: 10.1093/nar/gkh458

Gitzendanner, M. A., Soltis, P. S., Yi, T. S., Li, D. Z., and Soltis, D. E. (2018). Plastome phylogenetics: 30 years of inferences into plant evolution. *Adv. Bot. Res.* 85, 293–313. doi: 10.1016/bs.abr.2017.11.016

Goulding, S. E., Olmstead, R. G., Morden, C. W., and Wolfe, K. H. (1996). Ebb and flow of the chloroplast inverted repeat. *Mol. Gen. Genet.* 252,195–206. doi: 10.1007/BF02173220

Guisinger, M. M., Kuehl, J. V., Boore, J. L., and Jansen, R. K. (2011). Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Mol. Biol. Evol.* 28, 583–600. doi: 10.1093/molbev/msq229

Hodel, R. G. J., Zimmer, E., and Wen, J. (2021). A phylogenomic approach resolves the backbone of *Prunus* (Rosaceae) and identifies signals of hybridization and allopolyploidy. *Mol. Phylogenet. Evol.* 160:107118. doi: 10.1016/j.ympev.2021.107118

Hooker, J. D., and Thomson, T. (1854). On *Maddenia* and *Diplarche*, new genera of Himalayan plants. *Hooker's Journal of Botany and Kew Garden Miscellany* 6, 380–384.

Jansen, R. K., and Ruhlman,T. A. (2012). "Plastid genomes of seed plants" in *Genomics of Chloroplasts and Mitochondria*, eds R. Bock and V. Knoop (Dordrecht: Springer Netherlands), 103–126. doi: 10.1007/978-94-007-2920-9_5

Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., de Pamphilis, C. W., Yi, T. S., et al. (2020). Getorganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biol.* 21:241. doi: 10.1186/s13059-020-02154-5

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. *Appl. Plant Sci.* 4:1600016. doi: 10.3732/apps.1600016

Junier, T., and Zdobnov, E. M. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the U_{NIX} shell. *Bioinformatics* 26, 1669–1670. doi: 10.1093/bioinformatics/btq243

Kalkman, C. (2004). "Rosaceae," in *The Families and Genera of Vascular Plants. Flowering Plants–Dicotyledons: Celastrales, Oxalidales, Rosales, Cornales, Ericales*, ed K. Kubitzki (Berlin: Springer), 343–386. doi: 10.1007/978-3-662-07257-8_39

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvement in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kim, H. T., Kim, J. S., Li, Y. M., Mun, J. H., and Kim, J. H. (2018). Molecular markers for phylogenetic applications derived from comparative plastome analysis of *Prunus* species. *J. Syst. Evol.* 57, 15–22. doi: 10.1111/jse.12453

Kim, K. J., and Lee, H. L. (2004). Complete chloroplast genome sequences from Korean Ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* 11, 247–261. doi: 10.1093/dnares/11.4.247

Koehne, E. (1911). "Maddenia," in *Plantae Wilsonianae: An Enumeration of the Woody Plants Collected in Western China for the Arnold Arboretum of Harvard University During the Years 1907, 1908, and 1910*, eds E. H. Wilson and C. S. Sargent (Cambridge: The University Press), 56–59.

Kress, W. J. (2017). Plant DNA barcodes: application today and in the future. *J. Syst. Evol.* 55, 291–307. doi: 10.1111/jse.12254

Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., and Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8369–8374. doi: 10.1073/pnas.0503123102

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucl. Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., and Stamatakis, A. (2014). Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol. Biol.* 14:82. doi: 10.1186/1471-2148-14-82

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2016). Partitionfinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773. doi: 10.1093/molbev/msw260

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9:357. doi: 10.1038/nmeth.1923

Lee, S., and Wen, J. (2001). A phylogenetic analysis of *Prunus* and the Amygdaloideae (Rosaceae) using ITS sequences of nuclear ribosomal DNA. *Am. J. Bot.* 88, 150–160. doi: 10.2307/2657135

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, Q. J., Su, N., Zhang, L., Tong, R. C., Zhang, X. H., Wang, J. R., et al. (2020). Chloroplast genomes elucidate diversity, phylogeny, and taxonomy of *Pulsatilla* (Ranunculaceae). *Sci. Rep.* 10:19781. doi: 10.1038/s41598-020-76699-7

Li, X. W., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y. T., and Chen, S. L. (2015). Plant DNA barcoding: from gene to genome. *Biol. Rev.* 90, 157–166. doi: 10.1111/brv.12104

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187

Liu, B. B., Campbell, C. S., Hong, D. Y., and Wen, J. (2020a). Phylogenetic relationships and chloroplast capture in the *Amelanchier-Malacomeles-Peraphyllum* clade (Maleae, Rosaceae): evidence from chloroplast genome and nuclear ribosomal DNA data using genome skimming. *Mol. Phylogenet. Evol.* 147:106784. doi: 10.1016/j.ympev.2020.106784

Liu, B. B., Hong, D. Y., Zhou, S. L., Xu, C., Dong, W. P., Johnson, G., et al. (2019). Phylogenomic analyses of the *Photinia* complex support the recognition of a new genus *Phippsiomeles* and the resurrection of a redefined *Stranvaesia* in Maleae (Rosaceae). *J. Syst. Evol.* 57, 678–694. doi: 10.1111/jse.12542

Liu, B. B., Liu, G. N., Hong, D. Y., and Wen, J. (2020b). *Eriobotrya* belongs to *Rhaphiolepis* (Maleae, Rosaceae): evidence from chloroplast genome and nuclear ribosomal DNA data. *Front. Plant Sci.* 10:1731. doi: 10.3389/fpls.2019.01731

Liu, B. B., Ma, Z. Y., Ren, C., Hodel, R. G. J., Sun, M., Liu, X. Q., et al. (2021). Capturing single-copy nuclear genes, organellar genomes, and nuclear ribosomal DNA from deep genome skimming data for plant phylogenetics: a case study in Vitaceae. *J. Syst. Evol.* 59, 1124–1138 doi: 10.1111/jse.12806

Liu, H. J., He, J., Ding, C. H., Lyu, R. D., Pei, L. Y., Cheng, J., et al. (2018). Comparative analysis of complete chloroplast genomes of *Anemoclema*, *Anemone*, *Pulsatilla*, and *Hepatica* revealing structural variations among genera in tribe Anemoneae (Ranunculaceae). *Front. Plant Sci.* 9:1097. doi: 10.3389/fpls.2018.01097

Lohse, M., Drechsel, O., Kahlau, S., and Bock, R. (2013). OrganellarGenomeDRAW–a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucl. Acids Res.* 41, 575–581. doi: 10.1093/nar/gkt289

Lowe, T. M., and Chen, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucl. Acids Res.* 44, 54–57. doi: 10.1093/nar/gkw413

Lu, L. D., Gu, C. Z., Li, C. L., Alexander, C., Bartholomew, B., Brach, A. R., et al. (2003). "Rosaceae," in *Flora of China Volume 9*. eds Z. Y. Wu, P. H. Raven, and D. Y. Hong (Beijing: Science Press; St. Louis, MO: Botanical Garden Press), 46–434.

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). "Creating the CIPRES Science Gateway for inference of large phylogenetic trees," in *Proceedings of the Gateway Computing Environments Workshop (GCE)*. New Orleans, LA, 1–8. doi: 10.1109/GCE.2010.5676129

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Nakamura, T., Yamada, K. D., Tomii, K., and Katoh, K. (2018). Parallelization of MAFFT for large-scalemultiple sequence alignments. *Bioinformatics* 34, 2490–2492. doi: 10.1093/bioinformatics/bty121

Powell, W., Morgante, M., Andre, C., McNicol, J. W., Machray, G. C., and Rafalski, J. A. (1995). Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Curr. Biol.* 5, 1023–1029. doi: 10.1016/S0960-9822(95)00206-5

Qu, X. J., Moore, M. J., Li, D. Z., and Yi, T. S. (2019). PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* 15, 1–12. doi: 10.1186/s13007-019-0435-7

Rabah, S. O., Shrestha, B., Hajrah, N. H., Sabir, M. J., Alharby, H. F., Sabir, M. J., et al. (2018). *Passiflora* plastome sequencing reveals widespread genomic rearrangements. *J. Syst. Evol.* 9999, 1–14. doi: 10.1111/jse.12425

Rehder, A. (1956). *Manual of Cultivated Trees and Shrubs Hardy in North America Exclusive of the Subtropical and Warmer Temperate Regions, 2nd Edn*. New York, NY: The MacMillan Company.

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

Sanderson, M. J., Copetti, D., Búrquez, A., Bustamante, E., Charboneau, J. L., Eguiarte, L. E., et al. (2015). Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): loss of the *ndh* gene suite and inverted repeat. *Am. J. Bot.* 102, 1115–1127. doi: 10.3732/ajb.1500184

Sayyari, E., and Mirarab, S. (2016). Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* 33, 1654–1668. doi: 10.1093/molbev/msw079

Shi, W. T., Wen, J., and Lutz, S. S. (2013). Pollen morphology of the *Maddenia* clade of *Prunus* and its taxonomic and phylogenetic implications. *J. Syst. Evol.* 51, 164–183. doi: 10.1111/j.1759-6831.2012.00233.x

Shirasawa, K., Isuzugawa, K., Ikenaga, M., Saito, Y., Yamamoto, T., Hirakawa, H., et al. (2017). The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding. *DNA Res.* 24, 499–508. doi: 10.1093/dnares/dsx020

Slater, G. S. C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31. doi: 10.1186/1471-2105-6-31

Solís-Lemus, C., and Ané, C. (2016). Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genet.* 12:e1005896. doi: 10.1371/journal.pgen.1005896

Solís-Lemus, C., Bastide, P., and Ané, C. (2017). PhyloNetworks: a package for phylogenetic networks. *Mol. Biol. Evol.* 34, 3292–3298. doi: 10.1093/molbev/msx235

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/bioinformatics/btl446

Stamatakis, A. (2014). RaxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Sun, J. H., Wang, Y. H., Liu, Y. L., Xu, C., Yuan, Q. J., Guo, L. P., et al. (2020). Evolutionary and phylogenetic aspects of the chloroplast genome of *Chaenomeles* species. *Sci. Rep.* 10:11466. doi: 10.1038/s41598-020-67943-1

Thiel, T., Michalek, W., Varshney, R. K., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene derived SSR-markers in barley (*Hordeum vulgare* L). *Theor. Appl. Genet.* 106, 411–422. doi: 10.1007/s00122-002-1031-0

Walker, J. F., Zanis, M. J., and Emery, N. C. (2014). Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia burkei* (Madieae, Asteraceae). *Am. J. Bot.* 101, 722–729. doi: 10.3732/ajb.1400049

Wang, S., Shi, C., and Gao, L. Z. (2013). Plastid genome sequence of a wild woody oil species, *Prinsepia utilis*, provides insights into evolutionary and mutational patterns of Rosaceae chloroplast genomes. *PLoS ONE* 8:e73946. doi: 10.1371/journal.pone.0073946

Wang, X., Wang, J. R., Xie, S. Y., Zhang, X. H., Chang, Z. Y., Zhao, L., et al. (2021). Floral morphogenesis of the Maddenia and Pygeum groups of Prunus (Rosaceae), with an emphasis on the perianth. *J. Syst. Evol.* doi: 10.1111/jse.12748. [Epub ahead of print].

Wang, Y. B., Liu, B. B., Nie, Z. L., Chen, H. F., Chen, F. J., Figlar, R. B., et al. (2020). Major clades and a revised classification of *Magnolia* and Magnoliaceae based on whole plastid genome sequences *via* genome skimming. *J. Syst. Evol.* 58, 673–695. doi: 10.1111/jse.12588

Wang, Y. H., Qu, X. J., Chen, S. Y., Li, D. Z., and Yi, T. S. (2017). Plastomes of Mimosoideae: structural and size variation, sequence divergence, and phylogenetic implication. *Tree Genet. Genomes* 13, 1–18. doi: 10.1007/s11295-017-1124-1

Wen, J., Berggren, S. T., Lee, C. H., Ickert-Bond, S., Yi, T. S., Yoo, K. O., et al. (2008). Phylogenetic inferences in *Prunus* (Rosaceae) using chloroplast *ndhF* and ribosomal ITS sequences. *J. Syst. Evol.* 46, 322–332. doi: 10.3724/SP.J.1002.2008.08050

Wen, J., Harris, A. J., Kalburgi, Y., Zhang, N., Xu, Y., Zheng, W., et al. (2018). Chloroplast phylogenomics of the New World grape species (*Vitis*, Vitaceae). *J. Syst. Evol.* 56, 297–308. doi: 10.1111/jse.12447

Wen, J., and Shi, W. T. (2012). Revision of the *Maddenia* clade of *Prunus* (Rosaceae). *Phytokeys* 11, 39–59. doi: 10.3897/phytokeys.11.2825

Xiang, Y. Z., Huang, C. H., Hu, Y., Wen, J., Li, S. S., Yi, T. S., et al. (2016). Well-resolved Rosaceae nuclear phylogeny facilitates feological time and genome duplication analyses and ancestral fruit character reconstruction. *Mol. Biol. Evol.* 34, 262–281. doi: 10.1093/molbev/msw242

Xu, W. Q., Losh, J., Chen, C., Li, P., Wang, R. H., Zhao, Y. P., et al. (2019). Comparative genomics of figworts (*Scrophularia*, Scrophulariaceae), with implications for the evolution of *Scrophularia* and Lamiales. *J. Syst. Evol.* 57, 55–65. doi: 10.1111/jse.12421

Yang, J. B., Tang, M., Li, H. T., Zhang, Z. R., and Li, D. Z. (2013). Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evol. Biol.* 13:84. doi: 10.1186/1471-2148-13-84

Yü, T. T., Lu, L. T., and Ku, T. C. (1985). Taxa nova Rosacearum Sinicarum (V). *Acta Phytotaxonomica Sinica* 23, 209–215.

Yü, T. T., Lu, L. T., Ku, T. C., Li, C. L., and Chen, S. X. (1986). "Rosaceae (3) Prunoideae," in *Flora Reipublicae Popularis Sinicae, Vol. 38*, ed T. T. Yü (Beijing: Science Press), 1–133.

Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* 19:153. doi: 10.1186/s12859-018-2129-y

Zhang, N., Erickson, D. L., Ramachandran, P., Ottesen, A. R., Timme, R. E., Funk, V. A., et al. (2017). An analysis of *Echinacea* chloroplast genomes: implications for future botanical identification. *Sci. Rep.* 7:216. doi: 10.1038/s41598-017-00321-6

Zhang, N., Wen, J., and Zimmer, E. A. (2015). Congruent deep relationships in the grape family (Vitaceae) based on sequences of chloroplast genomes and mitochondrial genes *via* genome skimming. *PLoS ONE* 10 :e0144701. doi: 10.1371/journal.pone.0144701

Zhang, S. D., Jin, J. J., Chen, S. Y., Chase, M. W., Soltis, D. E., and Li, H. T. (2017). Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. *New Phytol.* 214, 1355–1367. doi: 10.1111/nph.14461

Zhang, X., Zhou, T., Kanwal, N., Zhao, Y. M., Bai, G. Q., and Zhao, G. F. (2017). Completion of eight Gynostemma BL. (Cucurbitaceae) chloroplast genomes: characterization, comparative analysis, and phylogenetic relationships. *Front. Plant Sci.* 8:1583. doi: 10.3389/fpls.2017.01583

Zhao, D. N., Ren, Y., and Zhang, J. Q. (2019). Conservation and innovation: plastome evolution during rapid radiation of *Rhodiola* on the Qinghai-Tibetan Plateau. *Mol. Phylogenet. Evol.* 144:106713. doi: 10.1016/j.ympev.2019.106713

Zhao, L., Jiang, X. W., Zuo, Y. J., Liu, X. L., Chin, S. W., Haberle, R., et al. (2016). Multiple events of allopolyploidy in the evolution of the racemose lineages in *Prunus* (Rosaceae) based on integrated evidence from nuclear and plastid data. *PLoS ONE* 11:e0157123. doi: 10.1371/journal.pone.0157123

Zhao, L., Potter, D., Xu, Y., Liu, P. L., Johnson, G., Chang, Z. Y., et al. (2018). Phylogeny and spatio-temporal diversification of *Prunus* subgenus *Laurocerasus* section *Mesopygeum* (Rosaceae) in the Malesian region. *J. Syst. Evol.* 56, 637–651. doi: 10.1111/jse.12467

Zimmer, E. A., and Wen, J. (2015). Using nuclear gene data for plant phylogenetics: progress and prospects II. *Next-gen approaches. J. Syst. Evol.* 53, 371–379. doi: 10.1111/jse.12174

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 03 January 2022
doi: 10.3389/fpls.2021.804662

The Chromosome-Level Genome of Miracle Fruit (*Synsepalum dulcificum*) Provides New Insights Into the Evolution and Function of Miraculin

Zhuang Yang[1,2], Zhenhuan Liu[1,2], Hang Xu[1,2], Yayu Chen[1,2], Pengmeng Du[1,2], Ping Li[1,2], Wenjie Lai[1,2], Haiyan Hu[1,2], Jie Luo[1,2]* and Yuanhao Ding[1,2]**

[1] Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresource, College of Tropical Crops, Hainan University, Haikou, China, [2] Sanya Nanfan Research Institute of Hainan University, Hainan Yazhou Bay Seed Laboratory, Sanya, China

OPEN ACCESS

Edited by:
Magdy S. Alabady,
University of Georgia, United States

Reviewed by:
Liangsheng Zhang,
Zhejiang University, China
Luis Herrera-Estrella,
Texas Tech University, United States

***Correspondence:**
Haiyan Hu
yanhai0987@hainanu.edu.cn
Jie Luo
jie.luo@hainanu.edu.cn
Yuanhao Ding
yhding@hainanu.edu.cn

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

Received: 29 October 2021
Accepted: 29 November 2021
Published: 03 January 2022

Citation:
Yang Z, Liu Z, Xu H, Chen Y, Du P, Li P, Lai W, Hu H, Luo J and Ding Y (2022) The Chromosome-Level Genome of Miracle Fruit (Synsepalum dulcificum) Provides New Insights Into the Evolution and Function of Miraculin. Front. Plant Sci. 12:804662. doi: 10.3389/fpls.2021.804662

Miracle fruit (*Synsepalum dulcificum*) is a rare valuable tropical plant famous for a miraculous sweetening glycoprotein, miraculin, which can modify sour flavors to sweet flavors tasted by humans. Here, we present a chromosome-level high-quality genome of *S. dulcificum* with an assembly genome size of ∼550 Mb, contig N50 of ∼14.14 Mb, and 37,911 annotated protein-coding genes. Phylogenetic analysis revealed that *S. dulcificum* was most closely related to *Camellia sinensis* and *Diospyros oleifera*, and that *S. dulcificum* diverged from the *Diospyros* genus ∼75.8 million years ago (MYA), and that *C. sinensis* diverged from *Synsepalum* ∼63.5 MYA. Ks assessment and collinearity analysis with *S. dulcificum* and other species suggested that a whole-genome duplication (WGD) event occurred in *S. dulcificum* and that there was good collinearity between *S. dulcificum* and *Vitis vinifera*. On the other hand, transcriptome and metabolism analysis with six tissues containing three developmental stages of fleshes and seeds of miracle fruit revealed that Gene Ontology (GO) terms and metabolic pathways of "cellular response to chitin," "plant–pathogen interaction," and "plant hormone signal transduction" were significantly enriched during fruit development. Interestingly, the expression of miraculin (Chr10G0299340) progressively increased from vegetative organs to reproductive organs and reached an incredible level in mature fruit flesh, with an fragments per kilobase of transcript per million (FPKM) value of ∼113,515, which was the most highly expressed gene among all detected genes. Combining the unique signal peptide and the presence of the histidine-30 residue together composed the main potential factors impacting miraculin's unique properties in *S. dulcificum*. Furthermore, integrated analysis of weighted gene coexpression network analysis (WGCNA), enrichment and metabolite correlation suggested that

miraculin plays potential roles in regulating plant growth, seed germination and maturation, resisting pathogen infection, and environmental pressure. In summary, valuable genomic, transcriptomic, and metabolic resources provided in this study will promote the utilization of *S. dulcificum* and in-depth research on species in the Sapotaceae family.

Keywords: genome evolution, miraculin, gene evolution, *Synsepalum dulcificum*, gene function

INTRODUCTION

Miracle fruit (*Synsepalum dulcificum*) is famous for a sweetening glycoprotein, namely, miraculin, which has a taste-modifying activity that converts sour taste to sweet (Kurihara and Beidler, 1968; Misaka, 2013; Niu et al., 2020). Miracle fruit is a native plant of tropical west and central Africa, occasionally distributed from West Africa to the Congo (Achigan-Dako et al., 2015). In the 1960s, miracle fruit was introduced to China as the national gift from the Republic of Ghana to Premier Enlai Zhou. Exports of miracle fruit are protected and prohibited in China and West Africa due to its high prize. Miracle fruit is an evergreen shrub belonging to the *Synsepalum* genus of the Sapotaceae family that has ∼50 genera and 1,100 species. Most species in Sapotaceae are trees and shrubs, which commonly have unique fruit flavors and are widely distributed in tropical and subtropical regions (Khayi et al., 2020).

The height of miracle fruit trees is ∼1.5–4.5 m, with many branches, dense leaves, and small white flowers, which usually flower from February to May and ripen in March, June, and October. The fruit is an ovoid to oblong berry, ∼2 cm long and 1 cm wide, with white and juicy flesh (Tchokponhoué et al., 2021). When the fruit matures, the color of the peel turns an attractive bright red from green in a relatively short time. Miracle fruit can be both autogamic and allogamic, commonly pollinated by insects such as ants, and mainly propagated through seeds (Tchokponhoué et al., 2019a,b; Akinmoladun et al., 2020).

Due to its dramatic taste-modifying activity, miraculin is usually used as a sweetening additive in the beverage and food industry and as an adjuvant for treating diabetic patients with insulin resistance (Chen et al., 2006; Tchokponhoué et al., 2020). Miraculin is a homodimeric protein obtained from the red berries of miracle fruit that exerts its taste-modifying activity by binding hT1R2–hT1R3 and functionally changing into an agonist at acidic pH (Koizumi et al., 2011). On the other hand, the extracts of the various tissues of the miracle fruit are of medicinal importance. For example, the extracts from the leaves show antidiabetic potential in type 2 diabetic rats, the stem has antioxidant bioconstituents that can inhibit human melanoma proliferation, and the extracts from fruit powder may be an effective treatment for acute gouty arthritis (Wang et al., 2011; Shi et al., 2016; Obafemi et al., 2017; Han et al., 2019; Huang et al., 2020). Moreover, miracle fruit also contains many important nutrients, such as proteins, lipids, vitamins, amino acids, and dietary phytochemicals (Njoku et al., 2015; He et al., 2016).

As an important group of tropical plants, there is no high-quality genome reported in the Sapotaceae family, which severely limits the utilization, breeding, and understanding of these species (Niu et al., 2020). Here, we report a high-quality and chromosome-level genome assembly of *S. dulcificum*, opening the gate for well-known of plants in the Sapotaceae family. The assembled genome size of *S. dulcificum* is ∼550 Mb, with contig N50 ∼14.14 Mb. We annotated 37,911 genes that were clustered into 15,799 gene families, and 3,828 and 4,739 gene families were expanded or contracted, respectively. Comparative genomic analysis revealed that miracle fruit underwent a whole-genome duplication (WGD) event, which had a close evolutionary relationship with *Camellia sinensis* and *Diospyros oleifera*. Moreover, metabolome and transcriptome analyses provide further information for understanding the mechanism of metabolite biosynthesis during fruit development. In addition, we found that the expression level of miraculin in the reproductive organs is very high, especially in the flesh of mature fruits with an fragments per kilobase of transcript per million (FPKM) value of ∼113,515, which is the highest expressed gene in the flesh. At the same time, we found that during the evolutionary process, histidine-30 residues and signal peptides were unique to miraculin in miracle fruit, which may have important implications for the production of its function and the evolution of which. We also found that the role of miraculin in miracle fruit is mainly to resist germ infection, defend against environmental pressure and regulate plant growth. As the first reference genome of *S. dulcificum*, this study will provide a valuable genomic resource for research on *S. dulcificum* and facilitate the production and utilization of *S. dulcificum*.

MATERIALS AND METHODS

Plant Materials and Sequencing

The sequenced specimen of *S. dulcificum* was cultivated at Hainan University (20°N, 110°E) in Haikou, Hainan Province, China. Total genomic DNA was extracted from fresh leaves using the plant Genomic DNA Kit (TIANGEN, Beijing, China). The purified DNA was used as input material for the DNA library constructed and sequenced using the PacBio Sequel I and Illumina HiSeq 2,500 platforms at Novogene (Tianjing, China). Total RNA was extracted from root, stem, leaf, flower, flesh, and seed tissues from three developmental stages, including the young stage (green), turning stage (yellowish red), and mature stage (red), and three biological replicates of each tissue were set. RNA sequencing (RNA-seq) was then performed according to the standard process at Novogene (Tianjing, China).

Genome Assembly and Quality Assessment

The *S. dulcificum* genome size was first assessed by flow cytometry (BD FACSCalibur) at OMIX Technologies Corporation (Chengdu, Sichuan, China). Karyotype analysis was performed to identify chromosome ploidy by fluorescence *in situ* hybridization (FISH) at OMIX Technologies Corporation (Chengdu, Sichuan, China). *K*-mer analysis ($K = 19$) was then used for heterozygosity and duplication rate assessment by GenomeScope (v2.0) (Ranallo-Benavidez et al., 2020) using paired-end Illumina data. Clean reads over 5k produced by PacBio sequencing were used for primary genome assembly by NextDenovo (v2.4.0).[1] Preliminarily assembled contigs were polished by Pilon (v1.22) (Walker et al., 2014) using Illumina short reads. Finally, we mapped the high-quality Hi-C to the polished genome using Bowtie (v0.7.8) (Langmead and Salzberg, 2012), and a chromosome-level genome was generated using a 3D-DNA pipeline (180922) (Dudchenko et al., 2017). Benchmarking Universal Single-Copy Orthologs (BUSCOs) (v3.0.2, embryophyta_odb10) (Simao et al., 2015) analysis and the long-terminal repeat assembly index (LAI) (Ou et al., 2018) were used to assess assembly completeness. In addition, PacBio long reads, Illumina short reads, Hi-C data, and RNA-seq data were mapped to the assembly to calculate the coverage through BWA (0.7.17) (Li and Durbin, 2010) and Hisat2 (2.1.0) (Kim et al., 2015).

Genome Annotation and Assessment

Homology-based prediction and *de novo* searching methods were employed to identify repetitive sequences in the *S. dulcificum* genome. RepeatModeler (v2.0.1) (Flynn et al., 2020) and LTR_Finder (v2.9.0) (Xu and Wang, 2007) were used to predict *de novo* repetitive sequences. Combining the *de novo* database and the known Repbase library, repetitive sequences were finally identified by using RepeatMasker (4.1.2-p1) (Tarailo-Graovac and Chen, 2009).

Protein-coding genes were predicted based on ab initio predictions, homology searching and RNA-seq data using the MAKER (v3.01.03) (Cantarel et al., 2008) pipeline. Augustus (v3.0.3) (Stanke et al., 2004) and SNAP (v2006-07-28) (Bromberg and Rost, 2007) were used for ab initio predictions. In homology searches, the protein sequences of *Arabidopsis thaliana*, *Actinidia chinensis*, *D. oleifera*, and *C. sinensis* were provided as protein evidence. For RNA evidence, the clean RNA-seq reads were assembled into inchworm contigs using Trinity (2.1.1) (Grabherr et al., 2011).

tRNAs were identified using tRNAscan-SE (v2.0.7) (Lowe and Eddy, 1997), rRNAs were detected using RNAmmer (v1.2) (Lagesen et al., 2007), and miRNAs and snRNAs were detected by searching the Rfam database (v12.2) (Kalvari et al., 2018) using infernal (v1.1.2) (Nawrocki and Eddy, 2013). The genome protein sequence of *S. dulcificum* was queried against the Plant Transcription Factor Database (PlantTFDB[2]) to identify transcription factors (TFs). The genome features (gene density, GC content, gene expression, repeat sequence density, and collinearity information) of *S. dulcificum* were visualized by Circos (v0.69) (Krzywinski et al., 2009).

Gene Family Identification and Evolution Analysis

Protein sequences from the longest transcripts of each protein-coding gene in 11 species, including *Clethra arborea*, *Roridula gorgonias*, *A. chinensis*, *C. sinensis*, *D. oleifera*, *Aegiceras corniculatum*, *Vaccinium macrocarpon*, *A. thaliana*, *Vitis vinifera*, *Oryza sativa*, and *S. dulcificum*, were used to identify gene families using OrthoFinder (v2.3.11) (Emms and Kelly, 2015) based on an all-versus-all BLASTP (*e*-value $\leq$ 1e−5). The gene families of *D. oleifera*, *V. macrocarpon*, *A. corniculatum*, *A. chinensis*, and *S. dulcificum* were identified by OrthoVenn (v2.0) (Xu et al., 2019).

A total of 293 single-copy orthologous genes from the 11 species were obtained by OrthoFinder. The protein sequences of the 293 single-copy orthologous genes were aligned using Muscle (v3.8.31) (Edgar, 2004), and the results were used to construct a phylogenetic tree by FastTree (v2.1.10) (Price et al., 2010). The MCMCTree program in PAML (v4.9) (Yang, 2007) was used to estimate the divergence time among the 11 species combined with previously published calibration times [*A. thaliana* and *V. vinifera* diverged 107–135 million years ago (MYA) and *A. thaliana* and *O. sativa* diverged 115–308 MYA] in TimeTree[3]. The expansion and contraction of the gene families were calculated using CAFÉ (v4.2) (De Bie et al., 2006) based on gene family count, phylogenetic tree, and species divergence time. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed mainly using ClusterProfiler (v3.13) (Yu et al., 2012) and other in-house R scripts.

Whole-Genome Duplication and Collinearity Analysis

To detect WGD events, the gene coding and protein sequences of *S. dulcificum*, *D. oleifera*, *C. sinensis*, and *A. thaliana* were used for synonymous substitution (Ks) value calculations by using KaKs_calculator (v2.0) (Zhang et al., 2006), and *A. thaliana* was used as a control. Homologous pairs were identified using an all-versus-all search in BLASTP (v2.5.0) (Altschul et al., 1990) and were used to identify syntenic blocks by WGD. The genome annotation information, coding sequence and protein sequence of *S. dulcificum* and *V. vinifera* were used as input files to perform synteny analysis through JCVI (v1.1.8)[4]. Dot plot analysis of *S. dulcificum* was produced by WGDI (v0.4.9) (Sun et al., 2021) using protein sequence and annotation information.

Transcriptome and Metabolome Analysis

RNA sequencing clean data of the 30 samples mentioned above were mapped back to the *S. dulcificum* genome using Hisat2 (v2.1.0) (Kim et al., 2015), and then the results were processed

[1] https://github.com/Nextomics/NextDenovo

[2] http://planttfdb.gao-lab.org/prediction.php

[3] http://www.timetree.org

[4] https://zenodo.org/record/31631#.YUvM4LgzaUk

by Samtools (v1.10-104-g869941a) (Li et al., 2009). The gene counts were calculated by HTSeq (v0.13.5) (Anders et al., 2015) and then normalized by StringTie (v2.1) (Pertea et al., 2015). Differentially expressed genes (DEGs) were then obtained by the DESeq2 (v1.32.0) (Love et al., 2014) package. GO and KEGG enrichment analyses were performed by ClusterProfiler. Weighted gene coexpression network analysis (WGCNA) was performed by the WGCNA (V1.70-3) (Langfelder and Horvath, 2008) package, and the soft threshold power was set to 10. The Pearson's correlation coefficients (PCCs) of genes and metabolites were calculated using the Cor package.

For metabolite detection, the freeze-dried samples were first ground into powder by a grinder (MM 400 Retsch) at 30 Hz for 30 s. Then, 1.0 ml of 70% aqueous methanol was added to 100 mg of powder for metabolite extraction. The metabolites were quantified by the MRM method of LC–MS/MS 6500 (SCIEX, Boston, United States). The relative signal strengths of the metabolites were divided by intensities of the internal standard (0.1 mg/L lidocaine) for normalization (Chen et al., 2013). In the following analysis, log2 transformation was performed for the metabolite content values.

RESULTS

Genome Sequencing and Assembly

To obtain a high-quality chromosomal-level genome of *S. dulcificum* (**Figure 1A**), a combination of short-read Illumina sequencing, long-read PacBio sequencing, and the Hi-C sequence approach was applied. The genome size of *S. dulcificum* was first investigated with ∼565.60 Mb by flow cytometry (**Supplementary Figure 1**). In addition, we found that *S. dulcificum* is a diploid plant containing $2n = 2x = 26$ chromosomes surveyed through FISH (**Supplementary Figure 2**). Relatively low heterozygosity (0.062%) and repetitive sequence content (37.5%) were found by *K*-mer analysis ($K = 19$; **Supplementary Figure 3**). A total of ∼119.57 Gb (226× coverage) of PacBio long reads was obtained, and reads longer than 5k were then used for the primary genome assembly. Subsequently, 104.90 Gb (∼198×) Illumina short reads were used for genome polishing, generating an improved version with a genome size of 568.98 M (contig N50 ∼14.14 M). We further scaffolded the genome to chromosome scale using ∼122.72 Gb (231× coverage) Hi-C clean data. Finally, we successfully clustered 63 contigs into 13 chromosomes by a 3D-DNA pipeline with an anchor ratio of ∼96.63% and generated a chromosome-level *S. dulcificum* genome (∼549.84 Mb) with chromosome lengths ranging from ∼29.30 Mb (Chr13) to ∼68.67 Mb (Chr1) (**Figure 1B**).

To assess the quality of the assembly, we mapped short reads from Illumina and Hi-C back to the final assembled genome, resulting in coverage of 99.29 and 93.64%, respectively. On the other hand, 96.56–98.27% of RNA-seq reads from 30 samples (see section "Materials and Methods") were well aligned to the assembled genome (**Supplementary Table 1**). In addition, 1,326 (96.5%) of the 1,375 complete gene elements in the BUSCOs plant set were covered by the *S. dulcificum* genome (**Supplementary Table 2**). A relatively high LAI assessment score (19.152) indicates the high sequence continuity of assembly, which has met the standard of reference quality (Ou et al., 2018).

Repeat Annotation and Gene Prediction

Using a combination of a *de novo* approach and a homology search, 305,010,596 bp (53.60%) of repetitive sequences were identified in *S. dulcificum*. Class I repeats (43.72%) were the main type; long terminal repeat (LTR) retrotransposons accounted for 40.20%. Gypsy and Copia, two main LTR-type transposons, accounted for 10.52 and 16.62% of the genome, respectively. Class II DNA transposons and unclassified repeats occupied 1.33 and 8.7% of the genome, respectively (**Supplementary Table 3**). For protein-coding gene prediction, a combination of ab initio, homolog, and transcriptome prediction strategies was used. In total, 37,911 protein-coding genes were predicted in *S. dulcificum*, with an average gene length of 3,951.5 bp, coding sequence length of 980.1 bp and exon number per gene of 4.12 (**Table 1**). The corresponding synteny along with GC content, gene expression levels, gene density, and distribution of TEs can be visualized in the CIRCOS diagram (**Figure 2**). In addition, non-coding RNA annotation yielded 117 microRNAs, 761 transfer RNAs, 215 ribosomal RNAs, and 94 small nuclear RNAs. In addition, 1,967 (5.18% in coding genes) TFs from 58 families were identified, and the largest TF families were bHLH, ERF, MYB, C2H2, and NAC (**Supplementary Table 4**). Furthermore, 33,725 protein-coding genes (88.95%) were supported by transcriptome data, and 30,296 protein-coding genes (81.37%) were assigned functions in at least one of the PFAM, COG, KEGG, or GO databases (**Supplementary Table 5**; Ashburner et al., 2000; Kanehisa and Goto, 2000; Mistry et al., 2020; Galperin et al., 2021).

Specific, Expanded, and Contracted Gene Families in *Synsepalum dulcificum*

As a special species in the Ericales order and Sapotaceae family, the special gene families of *S. dulcificum* may have unique functions. To identify the specific gene families in *S. dulcificum*, seven published genomes from Ericales, including *C. arborea* (Hartmann et al., 2020), *R. gorgonias* (Hartmann et al., 2020), *A. chinensis* (kiwi fruit) (Wu et al., 2019), *C. sinensis* (Chinese tea) (Xia et al., 2019, 2020), *D. oleifera Cheng* (wild kaki persimmon) (Zhu et al., 2019), *A. corniculatum* (Ma et al., 2021), and *V. macrocarpon* (American cranberry) (Diaz-Garcia et al., 2021); two representative eudicots [*A. thaliana* (Zapata et al., 2016) and *V. vinifera* (Jaillon et al., 2007)]; and one monocot [*O. sativa* (Sakai et al., 2013)] were chosen for gene family identification, in which *O. sativa* was chosen as the outgroup. A total of 32,771 (86.4%) genes in *S. dulcificum* were clustered into 15,799 gene families, leaving 5,140 genes that did not belong to any gene family (**Supplementary Table 6**). To clearly show the unique gene families of *S. dulcificum*, five important species (*V. vinifera*, *A. chinensis*, *D. oleifera*, *V. macrocarpon*, and *A. corniculatum*) were selected for further analysis. The results showed that 9,005 gene families were shared among miracle and other species, and 1,041 gene families were found to be specific to miracle fruit (**Figure 3A**

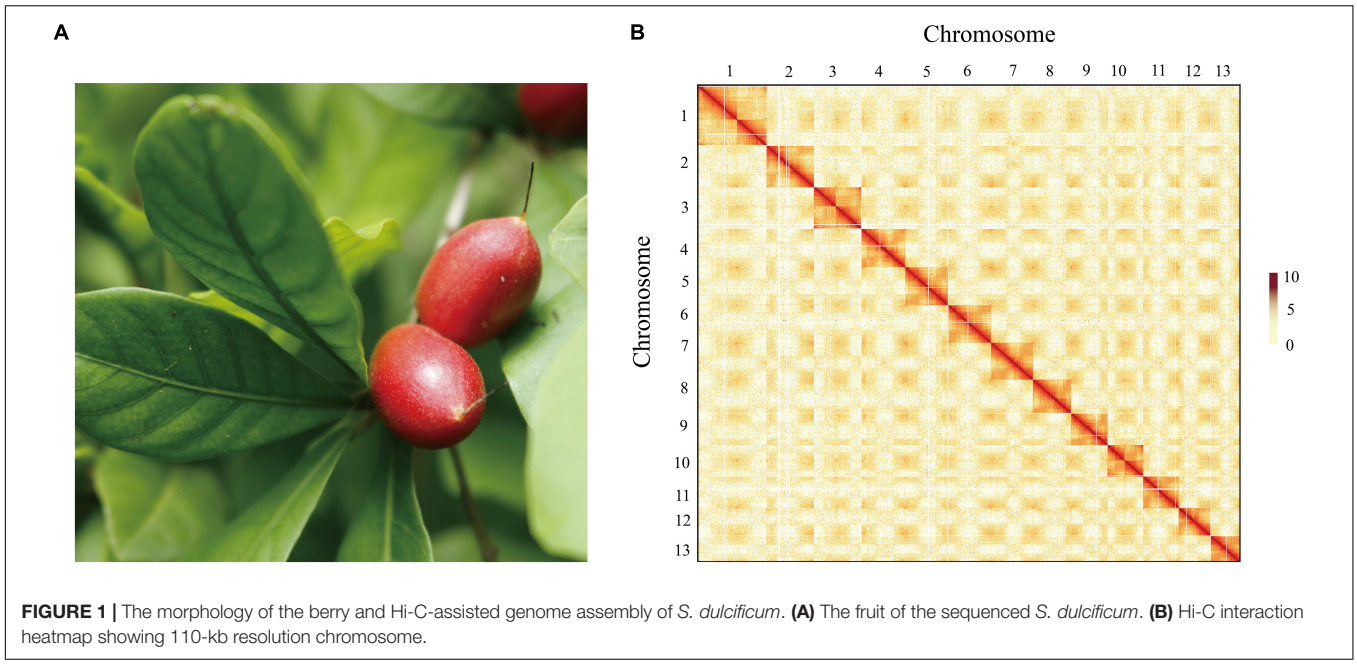


FIGURE 1 | The morphology of the berry and Hi-C-assisted genome assembly of *S. dulcificum*. **(A)** The fruit of the sequenced *S. dulcificum*. **(B)** Hi-C interaction heatmap showing 110-kb resolution chromosome.

and **Supplementary Table 6**). GO and KEGG enrichment analyses revealed that these specific gene families were mainly enriched in "sesquiterpene biosynthetic process," "phytoalexin biosynthetic process," "monoterpenoid biosynthesis," and "brassinosteroid biosynthesis" (**Supplementary Figures 4, 5** and **Supplementary Table 7**). In addition, a total of 18,640 gene families were inferred in the most recent common ancestor (MRCA) from the above 11 species by analyzing gene family expansion and contraction. In *S. dulcificum*, 3,828 gene families expanded, and 4,739 gene families contracted. GO and KEGG enrichment analysis of 102 significantly expanded gene families revealed a marked enrichment in genes involved in "PPAR signaling pathway," "phenylpropanoid biosynthesis," "defense response to fungus," "secondary metabolite biosynthetic process," and "pigment biosynthetic process" (**Supplementary Figures 6, 7** and **Supplementary Table 8**).

TABLE 1 | Statistics of *S. dulcificum* genome assembly and annotation.

Assembly feature	*Synsepalum dulcificum*
Number of contigs	81
Total size of contigs	569,013,336
Longest contig	39,932,979
Contig N50 count	13
Contig N50 length (bp)	14,129,605
Contig N90 count	40
Contig N90 length (bp)	5,031,503
GC content	33.80%
Chr count	13
Mean chr length (MB)	42.31
Number of protein-coding genes	37,911
Mean transcript length (bp)	3,951.5
Mean exon length (bp)	236.5
Mean exon per mRNA	4.12
Total repetitive sequences size (% of genome)	304,991,148 (53.60%)

Phylogenetic Placement of *Synsepalum dulcificum* as a Sister to *Camellia sinensis* and *Diospyros oleifera*

The Ericales order belongs to the Asterids clade, which together with Rosids consists of two main clades of Eudicots. To date, a total of 22 species from eight families have been sequenced in Ericales.[5] Regarding the loss of high-quality genome information from the Sapotaceae family, the phylogenetic position is uncertain. To investigate the phylogenetic placement of *S. dulcificum*, we constructed a phylogenetic tree based on 293 strictly single-copy ortholog gene sets from the 11 species above used for gene family analysis. The results show that *S. dulcificum* is closest to *C. sinensis* and *D. oleifera*, indicating that Sapotaceae family plants are much closer to species of Theaceae and Ebenaceae. The high-confidence phylogenetic tree and calibration points selected from the TimeTree website reflect that the divergence time between *S. dulcificum* and *C. sinensis* is ∼63.5 (45.4–78.5) MYA, and that between *S. dulcificum* and *D. oleifera* is ∼67.8 (50.4–82.6) MYA (**Figure 4**).

Whole-Genome Duplication and Collinearity Analysis of *Synsepalum dulcificum*

Whole-genome duplication events are important driving forces of plant evolution. To assess the WGD events of *S. dulcificum*, we selected a range of species (*D. oleifera*, *C. sinensis*, and *A. thaliana*) to calculate the synonymous substitution rate

[5]https://www.plabipd.de/index.ep

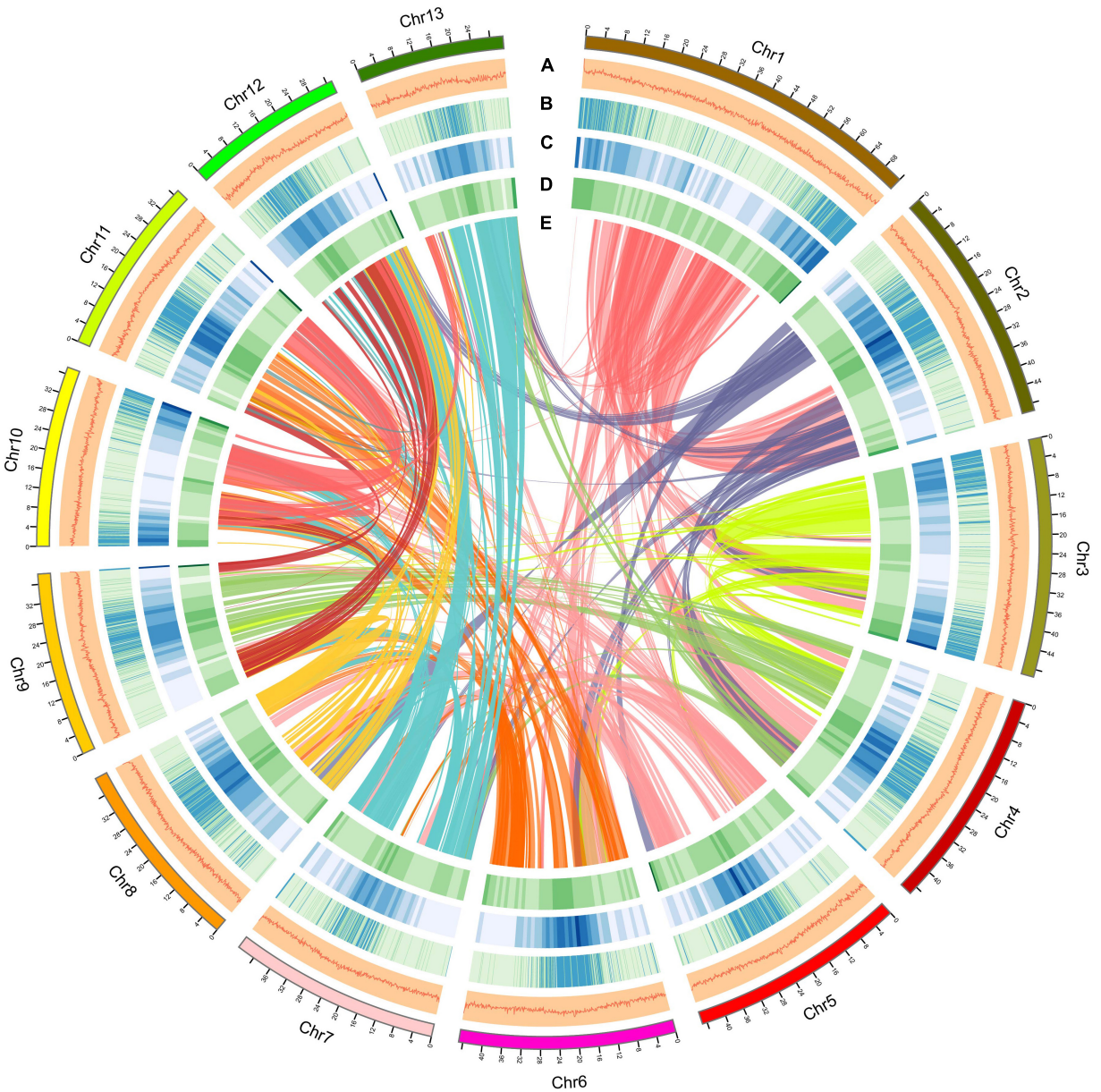


FIGURE 2 | Overview of the *S. dulcificum* genome assembly. The outer layer of colored blocks is a circular representation of the 13 chromosomes, with thick mark labeling each 4 Mb. Layers a–e represent GC content distributions **(A)**, gene expression levels **(B)**, gene density **(C)**, repeat sequence density **(D)**, and inter-chromosomal synteny **(E)**, respectively. More deep mark labeling represents higher expression level, gene density, or repeat sequence density for **(B–D)**.

(Ks) values with duplicate gene pairs. A major peak of *S. dulcificum* was detected at Ks = 0.56 (**Figure 3B**), indicating that *S. dulcificum* underwent a WGD event. The peak value (Ks = 0.45) of orthologs between *S. dulcificum* and *C. sinensis* was lower than the value of *S. dulcificum* (Ks = 0.56) itself, indicating that speciation between *S. dulcificum* and *C. sinensis* occurred later, and the *S. dulcificum* showed one-to-one syntenic relationship with *C. sinensis* (**Supplementary Figure 8**), which suggested that *C. sinensis* might share a common WGD event with *S. dulcificum* (**Figure 3B**). The peak value (Ks = 0.58) of orthologs between *S. dulcificum* and *D. oleifera* was larger than that of *S. dulcificum*, implying that they diverged earlier and that the WGD event occurred soon after they diverged, which corresponds to the phylogenetic relationship in **Figure 4**. The synteny dot plot analysis reveals that duplications within *S. dulcificum* are either interchromosomal (between chromosomes 1 and 2, 1 and 4, 2 and 8, 5 and 8, 6 and 10, 7 and 12, 7 and 13, 8 and 9, and 10 and 11) or intrachromosomal (Chr1 and Chr3) duplications (**Figure 5A**). The intergenomic collinearity between *S. dulcificum* and *V. vinifera* indicated that the *S. dulcificum* genome has a two-to-one syntenic collinearity relationship with *V. vinifera*, which also proved that

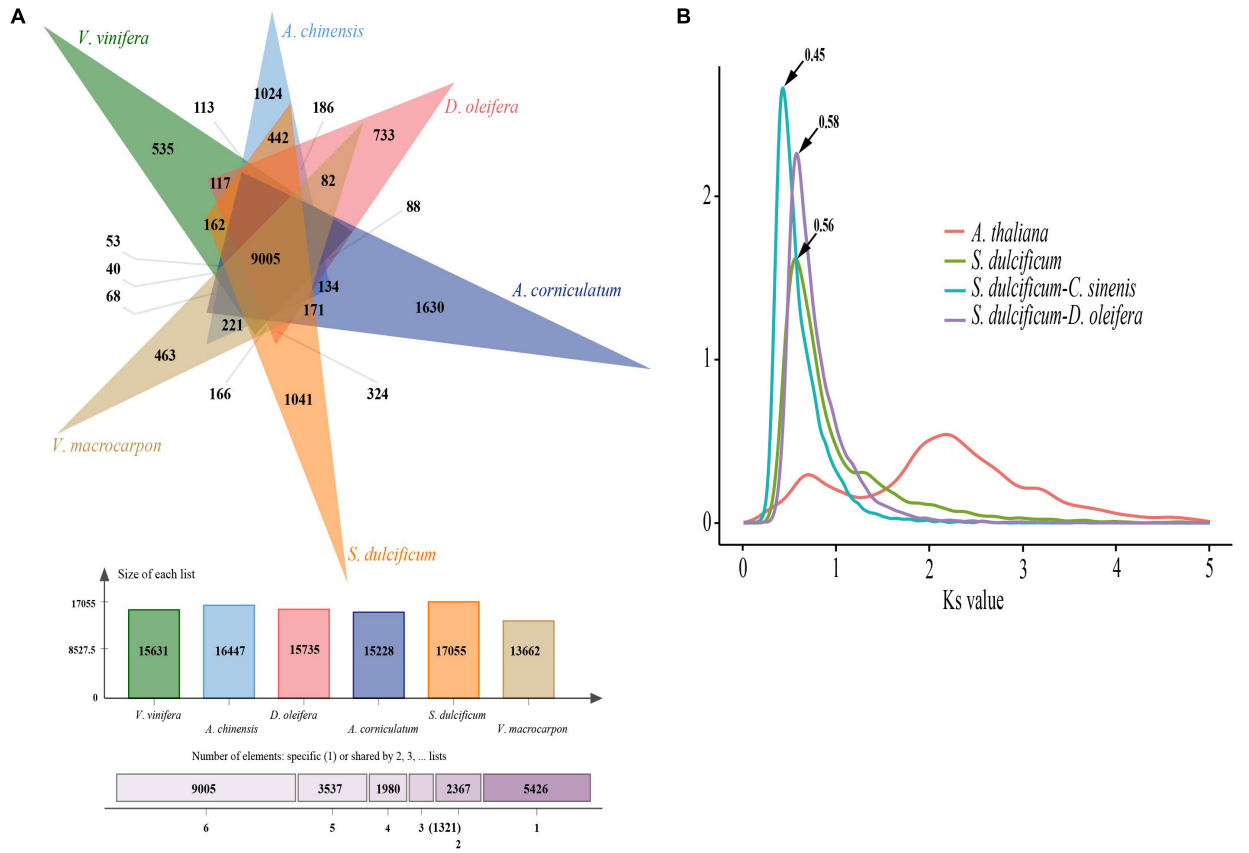


FIGURE 3 | Gene family and WGD analysis of *S. dulcificum*. **(A)** Gene family distributions between *S. dulcificum*, *V. vinifera*, *A. chinensis*, *D. oleifera*, *V. macrocarpon*, and *A. corniculatum*. **(B)** Synonymous substitution rate (Ks) distributions of paralogs and orthologs of *A. thaliana*, *S. dulcificum*, *C. sinensis*, and *D. oleifera*.

S. dulcificum underwent a WGD event after the eudicot ancestor evolved (**Figure 5B**).

Metabolite and Transcript Profiles During Fruit Development Stages

Comprehensive metabolic profiling of roots, stems, leaves, flowers, flesh, and seeds from three developmental stages [T1, young fruit (green); T2, turning-stage fruit (yellowish red); T3, mature fruit (bright red)] of *S. dulcificum* was performed by widely targeted metabolite analysis based on LC–MS/MS. A total of 855 metabolites were identified, with 697 annotated metabolites including 203 lipids, 137 vitamins, 100 amino acid derivatives, 73 organic acids, 63 flavonoids, 36 phenylpropanoids, 28 terpenes, 25 nucleotides and derivatives, 19 saccharides, 6 alkaloids, and 6 phytohormones (**Supplementary Table 12** and **Supplementary Figure 9**). A heatmap with all metabolites showed a good correlation among the three biological replicates of all samples (**Figure 6A**). The metabolite compositions between samples were clearly separated in the PCA diagram, indicating the spatiotemporal specificity of metabolites in miracle fruit (**Supplementary Figure 10**).

To study the changes in metabolites during fruit development, we compared the differences in metabolite content among T3, T2, and T1. Metabolic analysis found that 117 metabolites had higher levels in T2 than in T1, and 63 metabolites had higher levels in T3 than in T2. These increased metabolites were mainly lipids, vitamins, amino acids, and their derivatives (**Supplementary Table 13**). The increased metabolites were subjected to enrichment analysis using the KEGG database. The results of T2 vs. T1 showed that these enriched metabolic pathways were mainly "biosynthesis of plant hormones," "vitamin B6 metabolism," "ubiquinone and other terpenoid-quinone biosynthesis," "glycosylphosphatidylinositol (GPI)-anchor biosynthesis," etc. (**Supplementary Table 14**). The T3 vs. T2 results showed that these enriched metabolic pathways were mainly "phenylpropanoid biosynthesis," "phenylalanine metabolism," "glutathione metabolism," "biosynthesis of phenylpropanoids," etc. (**Supplementary Table 15**). We found that 32 metabolic pathways including "anthocyanin biosynthesis," "biosynthesis of terpenoids and steroids," "phenylpropanoid biosynthesis," and other pathways were enriched in both T2 vs. T1 and T3 vs. T2. This result indicates that these metabolites involved in the common metabolic pathway may play a more important role during fruit development.

RNA sequencing data were generated for six different tissues and three different periods, which correspond to the metabolome samples. We obtained 8.11 Gb of high-quality clean reads

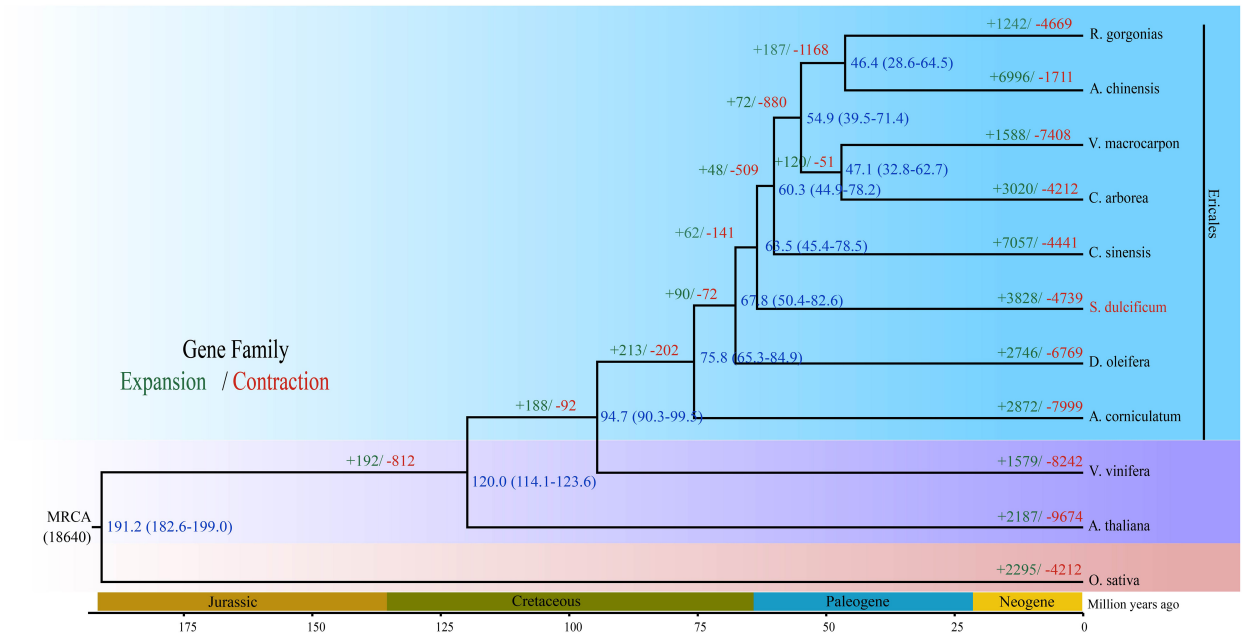


FIGURE 4 | Phylogenetic tree of *S. dulcificum* and 10 other species. The time of species differentiation and the expansion and contraction of gene families are denoted at each node. Green and red font numbers represent expansion and contraction of gene families, blue font words are species differentiation time.

on average after original data filtration. The three biological replicates of each sample were found to have a good correlation by calculating the expression levels indicated by FPKM reads (**Supplementary Figure 11**). In addition, 28,560 genes were found to be expressed (FPKM $\geq$1) in 30 samples. The PCA also supported the classification of the gene expression heatmap (**Supplementary Figure 12**).

To gain further insight into the regulation of the transcriptome dynamic changes throughout fruit development, DEG analysis was performed among T3, T2, and T1 fruit samples. A total of 9,246 and 3,524 DEGs were identified in T2 vs. T1 and T3 vs. T2, respectively, based on p-value < 0.05 and $|\log2\text{Fold Change}| \geq 1$. Among them, 3,824 DEGs were upregulated and 5,418 DEGs were downregulated in T2 vs. T1; in addition, 2,092 DEGs were upregulated and 1,432 DEGs were downregulated in T3 vs. T2. To investigate the specific functions of the upregulated DEGs, GO, and KEGG analyses were performed for T2 vs. T1 and T3 vs. T2. "Phenylpropanoid biosynthetic process," "cellular response to chitin," "flavonoid biosynthetic process," and 388 other GO terms were significantly enriched in T2 vs. T1. "Glucosinolate metabolic process," "lanosterol synthase activity," "response to brassinosteroid," and 358 other GO terms were significantly enriched in T3 vs. T2 (**Supplementary Figures 13, 14** and **Supplementary Table 16**). "Amino sugar and nucleotide sugar metabolism," "fatty acid biosynthesis," "biosynthesis of secondary metabolites," and 118 other metabolic pathways were enriched in T2 vs. T1. "Plant–pathogen interaction," "MAPK signaling pathway," "plant hormone signal transduction," and 107 other metabolic pathways were enriched in T3 vs. T2 (**Supplementary Figures 15, 16** and **Supplementary Table 17**). By comparing the KEGG pathways of the two comparisons (T3 vs. T2 and T2 vs. T1), we found that differential metabolites and DEGs were enriched in the same pathways, such as "steroid biosynthesis," "vitamin B6 metabolism," and "flavonoid biosynthesis" (**Supplementary Table 18**). These enriched pathways suggest that metabolite changes may be positively regulated by genes and provide insights into the genetic basis of metabolic processes underlying different fruit development stages in *S. dulcificum*.

On the other hand, PCCs between miraculin and metabolites were also calculated, and 57 metabolites were found to be highly correlated (PCC >0.6) with miraculin (**Figure 6B** and **Supplementary Table 19**). Among them, there are many metabolites that are particularly important for plant growth and defense. For example, citric acid (MIR032) is involved in fruit growth, vitamin B6 (MIR045) is involved in heat stress, urate (MIR035) is involved in responses to the pathogenic fungus, and inosine 5′-monophosphate (MIR030) is involved in salt stress.

The Evolution and Function of Miraculin

As the most special protein in miracle fruit, miraculin has been well studied in terms of structure, taste-modifying activity mechanisms, subcellular localization, etc., but few reports have focused on its evolution and function in miracle fruit itself (Paladino et al., 2008; Takai et al., 2013; Sanematsu et al., 2016). In this study, an interesting phenomenon was found: the miraculin gene (Chr10G0299340) was the most highly expressed gene in fruit flesh, with an FPKM value of ∼113,515, which is consistent to the report of high protein level of miraculin in the fruit of *S. dulcificum* (Hiwasa-Tanase et al., 2012; **Figure 7A**). To investigate the peculiar properties of miraculin, three homologous genes of miraculin from *S. dulcificum*, seven

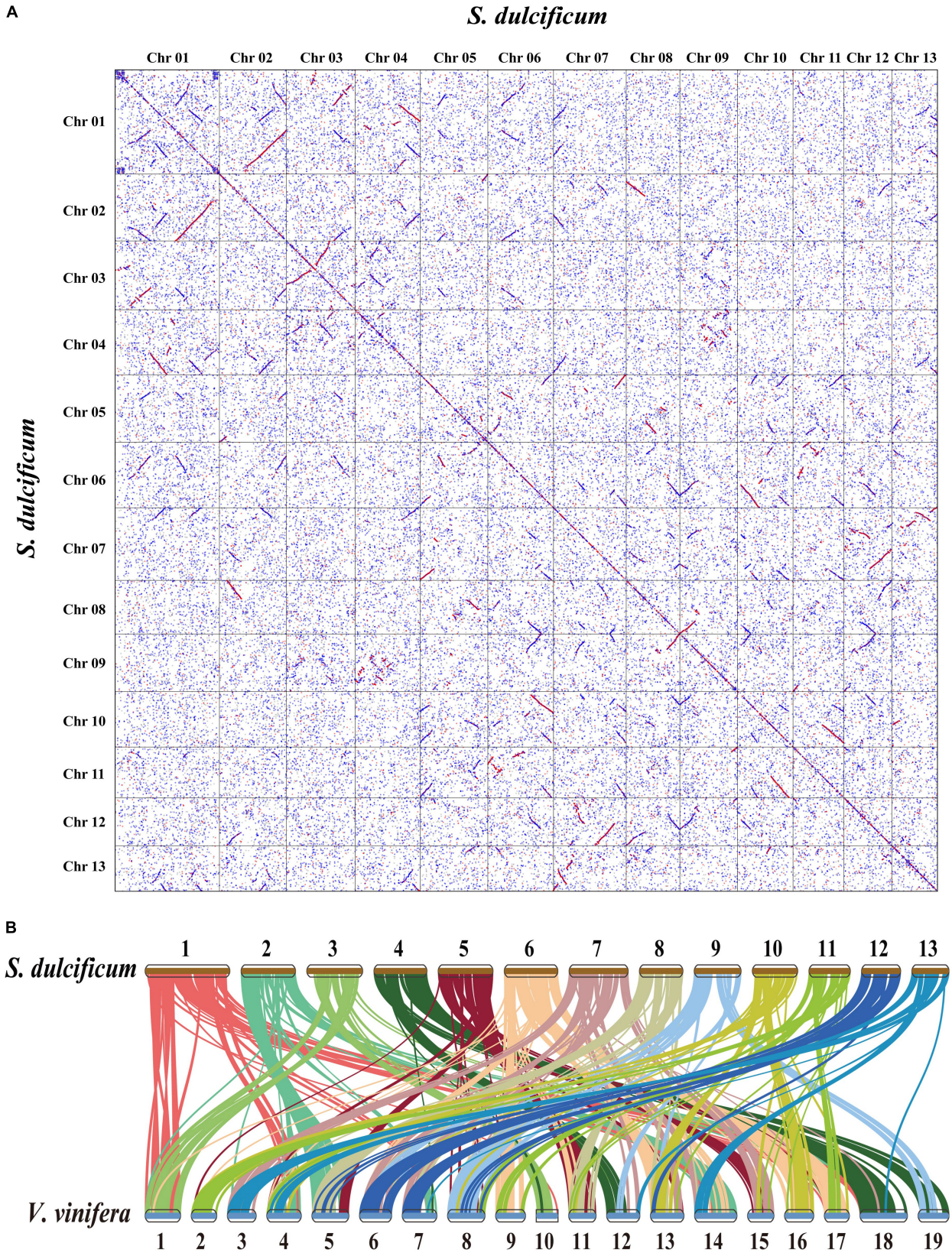


FIGURE 5 | Genomic synteny analysis between chromosomes of *S. dulcificum*, and between *S. dulcificum* and *V. vinifera*. **(A)** Dot plots of paralogs in *S. dulcificum* genome. **(B)** Collinear relationships between *S. dulcificum* and *V. vinifera*.

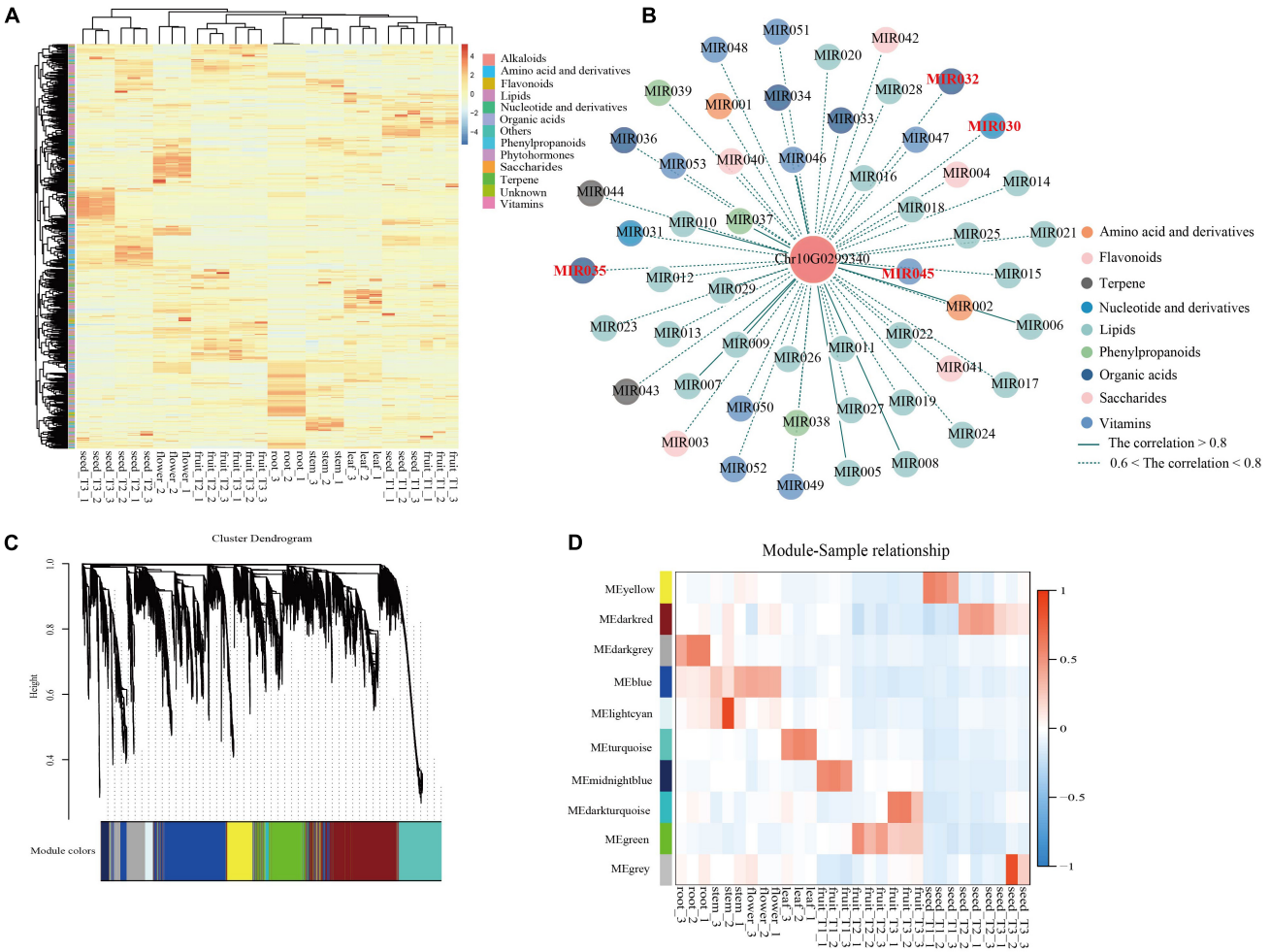


FIGURE 6 | Integrated analysis of metabolome and transcriptome revealed the functions of miraculin in *S. dulcificum*. **(A)** Heatmap of metabolites detected in six tissues including fleshes and seeds from three stages of *S. dulcificum* fruits. **(B)** Regulatory network of metabolites highly related to miraculin. **(C)** The dendrogram of co-expression modules identified by WGCNA across samples. **(D)** The heatmap of the correlation between samples and modules. [MIR030, inosine 5′-monophosphate; MIR032, citric acid; MIR035, urate; MIR045, pyridoxine (vitamin B6)].

from *C. sinensis*, three from *D. oleifera*, and eight from *V. vinifera* were obtained. The expression levels of these homologous genes were determined by a standard RNA-seq analysis process using published data (see **Supplementary Table 9**). The results showed that the expression levels of miraculin in the fruit of *S. dulcificum* were significantly higher than the expression of the homologous genes in other species (**Figure 7B**). Furthermore, a phylogenetic tree was constructed using the protein sequences of these homologous genes. Obviously, the phylogenetic tree divided into four clades in accordance with the species, indicating the relative species specificity of the miraculin family (**Figure 7C**). Ka/Ks analysis showed that the miraculin gene has undergone purifying selection (**Figure 7C**). In addition, we also found that the signal peptide motif (motif 9) of miraculin is unique in *S. dulcificum* and has been reported to function in the transfer of miraculin outside the plasma membrane (Takai et al., 2013; **Figure 7C**). It has been reported that the histidine-30 residue is the key point for the taste-modifying activity of miraculin (Ito et al., 2007). Compared with the homologous genes from other species, the histidine-30 residue is a unique site in *S. dulcificum*. Besides, we compared the gene expression (**Supplementary Figure 17**), protein sequence and functional annotation of the miraculin homologous genes in *S. dulcificum*, the results showed that only Chr10G0299340 has a higher expression level in mature fruits. Thus, we speculated that the extremely high expression level in the flesh of fruit, the unique signal peptides, and the histidine-30 residue together form the specific characteristics of miraculin in *S. dulcificum*.

To dissect the function of miraculin in *S. dulcificum*, WGCNA was performed to investigate the coexpression networks of miraculin. A total of 19,850 genes were screened out after discarding the low expression genes (FPKM <10), which were then used to construct coexpression network modules, resulting in 10 modules (**Figure 6C**). We found that miraculin was located in the yellow module, which contained 2,647 genes (**Figure 6D** and **Supplementary Table 10**). PCCs between

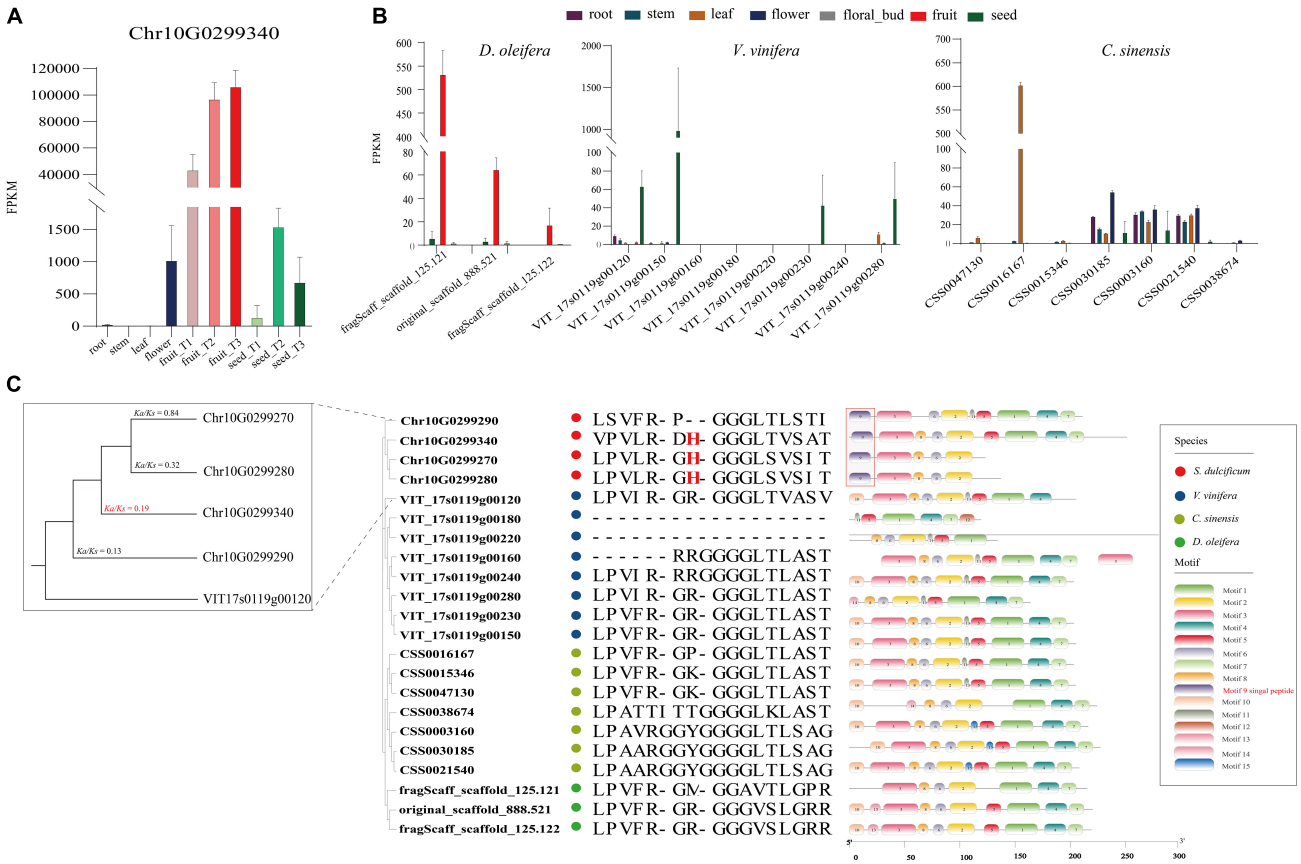


FIGURE 7 | Expression level, phylogenetic tree, and unique protein sites of miraculin and its homologous genes in *S. dulcificum* and genetically close species. **(A)** Expression levels of miraculin in various tissues of *S. dulcificum*. FPKM values were used. **(B)** The expression level of homologous genes of miraculin in tissues of *D. oleifera*, *V. vinifera*, and *C. sinensis*. **(C)** The phylogenetic tree, selective pressure analysis, and protein sequence alignment of miraculin and its homologous genes in *S. dulcificum*, *V. vinifera*, *C. sinensis*, and *D. oleifera*.

genes from the yellow module were calculated, genes highly related to miraculin (PCCs >0.6) were obtained, and 331 genes remained. GO analysis showed that the terms "chitinase activity," "response to fungus," "seed germination," "seed maturation," and other terms related to resistance to adversity were significantly enriched (**Supplementary Figure 18** and **Supplementary Table 11**). The KEGG analysis showed that the pathways "metabolic pathways," "oxidative phosphorylation," and "biosynthesis of secondary metabolites" were significantly enriched (**Supplementary Table 11**).

DISCUSSION

Synsepalum dulcificum originating from western and central Africa is a rare plant distributed in tropical and subtropical regions. In this research, we constructed a high-quality chromosome-level reference genome of *S. dulcificum* by combining Illumina short-read data, PacBio long-read data and Hi-C data. The assembly genome size is ~550 Mb, with a contig N50 of 14.14 Mb. A total of 94.3% complete BUSCOs present in the *S. dulcificum* assembly indicated high genome integrity. In addition, 37,911 protein-coding genes were predicted by a combination of ab initio, homolog, and transcriptome prediction strategies. This is the first chromosome-level reference genome of the Sapotaceae family, which provides important genomic data for *S. dulcificum* studies and provides a reference for studies of other species in the Sapotaceae family.

In this study, 15,799 gene families were identified, 1,041 of which were specific for *S. dulcificum*. Enrichment analysis found that these specific gene families were involved in many important processes, such as "sesquiterpene biosynthetic process," "phytoalexin biosynthetic process," and "monoterpenoid biosynthesis." Phylogenetic analysis based on 293 single-copy orthologous genes revealed that *S. dulcificum* was most closely related to *C. sinensis* and *D. oleifera*, which provides new insight into the evolutionary relationship of species in the Ericales order. The divergence time between *S. dulcificum* and *C. sinensis* was ~63.5 MYA, and that between *S. dulcificum* and *D. oleifera* was ~67.8 MYA. The divergence time between the *D. oleifera* and *C. sinensis* lineages in previous reports is ~85.3 and ~75 MYA, which is slightly larger than that in our study (Akagi et al., 2020). We speculate that this may be because we used more genome data from species of the

order Ericales, especially the genome data of *S. dulcificum* of the Sapotaceae family that we constructed in our study. We also found that 3,828 gene families had expanded and that 4,739 gene families had contracted in the *S. dulcificum* genome, which were mostly involved in the "DNA replication proteins," "PPAR signaling pathway," "defense response to fungus," etc.

As one of the important driving forces for plant evolution, a WGD event was found in *S. dulcificum*. In addition, the *S. dulcificum* genome has good collinearity with *V. vinifera*. However, whether the WGD that occurred in *S. dulcificum* is shared with other species in the Ericales order remains unclear. Previous studies have provided evidence that *C. sinensis* experienced only WGD events after the core-eudicot WGT event, and WGD was shared by at least 17 families in Ericales (Wang et al., 2021). Ebenaceae, Theaceae, and Actinidiaceae may share the same WGD event, which occurred ∼110 MYA. Combined with the evolutionary position of *S. dulcificum* in the order Ericales (Zhang et al., 2020), we speculate that the WGD of *S. dulcificum* may also be shared with other species of Ericales but is not specific to the Sapotaceae family. Interestingly, *C. sinensis* and *D. oleifera* have 15 chromosomes, but *S. dulcificum* has only 13 chromosomes. The ancestral chromosomal base number of the core order Ericales is believed to be 9 (Soza et al., 2019). Thus, research on the evolutionary history of *S. dulcificum* would be meaningful and important for species in the Sapotaceae family, even in Ericales.

Furthermore, we also performed metabolite detection and transcriptome sequencing from six tissues of miracle fruit, including three stages of fruit flesh and seeds. A total of 697 annotated metabolites were detected, and 28,560 genes were expressed. Metabolite difference analysis found that the contents of lipids, vitamins, amino acids, and their derivatives increased during fruit development. The KEGG enrichment analysis of the increased metabolites showed that these pathways were enriched in many important processes. We also performed GO and KEGG enrichment analysis for upregulated DEGs in T3 vs. T2 and T2 vs. T1. The significantly enriched GO terms and KEGG pathways were mainly "phenylpropanoid biosynthetic process," "cellular response to chitin," "sugar metabolism," "fatty acid biosynthesis," "biosynthesis of secondary metabolites," etc. We found that many metabolites and DEGs were enriched in the same pathway, which provides insights for understanding the molecular mechanism of important metabolite biosynthesis.

In this study, an interesting phenomenon was found: the miraculin gene (Chr10G0299340) was the most highly expressed gene in fruit flesh, with an FPKM value of ∼113,515, indicating the potential high protein level of miraculin in *S. dulcificum*. To investigate the peculiar properties of miraculin, three homologous genes of miraculin from *C. sinensis*, *D. oleifera*, and *V. vinifera* were identified. The gene expression of Chr10G0299340 is at least 100 times that of its homologous genes, and we found that it has signal peptides and histidine-30 residues that other homologous genes do not have. Thus, we speculated that the extremely high expression level in the flesh of fruit, the unique signal peptides, and the histidine-30 residue together form the specific characteristics of miraculin in *S. dulcificum*. In previous studies, researchers were only concerned about the benefits of miraculin to the human body and never studied the function of miraculin on the miracle fruit itself (Koizumi et al., 2011; Swamy et al., 2014; Huang et al., 2020). In our research, combining WGCNA, enrichment analysis and metabolite correlation analysis, we believed that miraculin mainly plays a role in regulating seed germination and maturation, resisting pathogen infection, resisting environmental pressure, and regulating plant growth, which is consistent with what has been reported regarding the function of miraculin-like proteins in grape (Ohkura et al., 2018), coffee (Mondego et al., 2011), tomato (Brenner et al., 1998), and rough lemon (Tsukuda et al., 2006). The above results indicated that the peculiar property of miraculin that modifies sour tastes to sweet tastes may be collateral, and the main meaning of its existence is to benefit itself.

In summary, the high-quality reference genome sequence, metabolomic and transcriptomic data of different tissues and periods of *S. dulcificum* obtained in our study provide valuable resources for both functional genomic research and genetic improvement breeding of miracle as well as other economically important plants in the Sapotaceae family.

DATA AVAILABILITY STATEMENT

All sequencing data of the genome have been deposited into the CNGB sequence archive of CNGBdb with accession number CNP0002330 (https://db.cngb.org/). And the sequencing data of RNA-seq have been deposited in the NCBI Sequence Read Archive database under accession number PRJNA778426.

AUTHOR CONTRIBUTIONS

YD and HH designed and supervised the project. ZL, HX, YC, PD, PL, and WL prepared the samples. ZY, ZL, and HX analyzed the data. ZY wrote the manuscript. JL, YD, and HH revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This research was financially supported by the National Natural Science Foundation of China (no. 31960437), the Hainan Natural Science Foundation of High-Level Talents Project (no. 320RC494), the Research Foundation for Advanced Talents of Hainan University [no. KYQD(ZR)1849], and the Innovation Project of Postgraduates of Hainan Province (no. Hyb2020-08).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.804662/full#supplementary-material

REFERENCES

Achigan-Dako, E. G., Tchokponhoué, D. A., N'danikou, S., Gebauer, J., and Vodouhè, R. S. (2015). Current knowledge and breeding perspectives for the miracle plant *Synsepalum dulcificum* (Schum. et Thonn.) Daniell. *Genet. Resour. Crop Evol.* 62, 465–476. doi: 10.1007/s10722-015-0225-7

Akagi, T., Shirasawa, K., Nagasaki, H., Hirakawa, H., Tao, R., Comai, L., et al. (2020). The persimmon genome reveals clues to the evolution of a lineage-specific sex determination system in plants. *PLoS Genet.* 16:e1008566. doi: 10.1371/journal.pgen.1008566

Akinmoladun, A. C., Adetuyi, A. R., Komolafe, K., and Oguntibeju, O. O. (2020). Nutritional benefits, phytochemical constituents, ethnomedicinal uses and biological properties of Miracle fruit plant (*Synsepalum dulcificum* Shumach. & Thonn. Daniell). *Heliyon* 6:e05837. doi: 10.1016/j.heliyon.2020.e05837

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Brenner, E. D., Lambert, K. N., Kaloshian, I., and Williamson, V. M. (1998). Characterization of LeMir, a root-knot nematode-induced gene in tomato with an encoded product secreted from the root. *Plant Physiol.* 118, 237–247. doi: 10.1104/pp.118.1.237

Bromberg, Y., and Rost, B. (2007). SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res.* 35, 3823–3835. doi: 10.1093/nar/gkm238

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., et al. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18, 188–196. doi: 10.1101/gr.6743907

Chen, C. C., Liu, I. M., and Cheng, J. T. (2006). Improvement of insulin resistance by miracle fruit (*Synsepalum dulcificum*) in fructose-rich chow-fed rats. *Phytother. Res.* 20, 987–992. doi: 10.1002/ptr.1919

Chen, W., Gong, L., Guo, Z., Wang, W., Zhang, H., Liu, X., et al. (2013). A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol. Plant* 6, 1769–1780. doi: 10.1093/mp/sst080

De Bie, T., Cristianini, N., Demuth, J. P., and Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22, 1269–1271. doi: 10.1093/bioinformatics/btl097

Diaz-Garcia, L., Garcia-Ortega, L. F., Gonzalez-Rodriguez, M., Delaye, L., Iorizzo, M., and Zalapa, J. (2021). Chromosome-Level Genome Assembly of the American Cranberry (*Vaccinium macrocarpon* Ait.) and Its Wild Relative *Vaccinium microcarpum*. *Front. Plant Sci.* 12:633310. doi: 10.3389/fpls.2021.633310

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., et al. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* 356, 92–95. doi: 10.1126/science.aal3327

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113

Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157. doi: 10.1186/s13059-015-0721-2

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* 117, 9451–9457. doi: 10.1073/pnas.1921046117

Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D., and Koonin, E. V. (2021). COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* 49, D274–D281. doi: 10.1093/nar/gkaa1018

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Han, Y. C., Wu, J. Y., and Wang, C. K. (2019). Modulatory effects of miracle fruit ethanolic extracts on glucose uptake through the insulin signaling pathway in C2C12 mouse myotubes cells. *Food Sci. Nutr.* 7, 1035–1042. doi: 10.1002/fsn3.935

Hartmann, S., Preick, M., Abelt, S., Scheffel, A., and Hofreiter, M. (2020). Annotated genome sequences of the carnivorous plant Roridula gorgonias and a non-carnivorous relative, *Clethra arborea*. *BMC Res. Notes* 13:426. doi: 10.1186/s13104-020-05254-4

He, Z., Tan, J. S., Abbasiliasi, S., Lai, O. M., Tam, Y. J., and Ariff, A. B. (2016). Phytochemicals, nutritionals and antioxidant properties of miracle fruit *Synsepalum dulcificum*. *Indust. Crops Prod.* 86, 87–94. doi: 10.1016/j.indcrop.2016.03.032

Hiwasa-Tanase, K., Hirai, T., Kato, K., Duhita, N., and Ezura, H. (2012). From miracle fruit to transgenic tomato: mass production of the taste-modifying protein miraculin in transgenic plants. *Plant Cell Rep.* 31, 513–525. doi: 10.1007/s00299-011-1197-5

Huang, W., Chung, H. Y., Xuan, W., Wang, G., and Li, Y. (2020). The cholesterol-lowering activity of miracle fruit (*Synsepalum dulcificum*). *J. Food Biochem.* 44:e13185. doi: 10.1111/jfbc.13185

Ito, K., Asakura, T., Morita, Y., Nakajima, K., Koizumi, A., Shimizu-Ibuka, A., et al. (2007). Microbial production of sensory-active miraculin. *Biochem. Biophys. Res. Commun.* 360, 407–411. doi: 10.1016/j.bbrc.2007.06.064

Jaillon, O., Aury, J. M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–467. doi: 10.1038/nature06148

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., et al. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* 46, D335–D342. doi: 10.1093/nar/gkx1038

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.

Khayi, S., Gaboun, F., Pirro, S., Tatusova, T., El Mousadik, A., Ghazal, H., et al. (2020). Complete chloroplast genome of *Argania spinosa*: structural organization and phylogenetic relationships in sapotaceae. *Plants (Basel)* 9:1354. doi: 10.3390/plants9101354

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Koizumi, A., Tsuchiya, A., Nakajima, K., Ito, K., Terada, T., Shimizu-Ibuka, A., et al. (2011). Human sweet taste receptor mediates acid-induced sweetness of miraculin. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16819–16824. doi: 10.1073/pnas.1016644108

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Kurihara, K., and Beidler, L. M. (1968). Taste-modifying protein from miracle fruit. *Science* 161, 1241–1243. doi: 10.1126/science.161.3847.1241

Lagesen, K., Hallin, P., Rodland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955

Ma, D., Guo, Z., Ding, Q., Zhao, Z., Shen, Z., Wei, M., et al. (2021). Chromosome-level assembly of the mangrove plant *Aegiceras corniculatum* genome generated through Illumina, PacBio and Hi-C sequencing technologies. *Mol. Ecol. Resour.* 21, 1593–1607. doi: 10.1111/1755-0998.13347

Misaka, T. (2013). Molecular mechanisms of the action of miraculin, a taste-modifying protein. *Semin. Cell Dev. Biol.* 24, 222–225. doi: 10.1016/j.semcdb.2013.02.008

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

Mondego, J. M., Duarte, M. P., Kiyota, E., Martínez, L., De Camargo, S. R., De Caroli, F. P., et al. (2011). Molecular characterization of a miraculin-like gene differentially expressed during coffee development and coffee leaf miner infestation. *Planta* 233, 123–137. doi: 10.1007/s00425-010-1284-9

Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi: 10.1093/bioinformatics/btt509

Niu, Y. F., Ni, S. B., and Liu, J. (2020). Complete chloroplast genome of *Synsepalum dulcificum* D.: a magical plant that modifies sour flavors to sweet. *Mitochondrial. DNA B Resour.* 5, 3052–3053. doi: 10.1080/23802359.2020.1798299

Njoku, N. E., Ubbaonu, C. N., Alagbaoso, S. O., Eluchie, C. N., and Umelo, M. C. (2015). Amino acid profile and oxidizable vitamin content of *Synsepalum dulcificum* berry (miracle fruit) pulp. *Food Sci. Nutr.* 3, 252–256. doi: 10.1002/fsn3.213

Obafemi, T. O., Akinmoladun, A. C., Olaleye, M. T., Agboade, S. O., and Onasanya, A. A. (2017). Antidiabetic potential of methanolic and flavonoid-rich leaf extracts of *Synsepalum dulcificum* in type 2 diabetic rats. *J. Ayurveda Integr. Med.* 8, 238–246. doi: 10.1016/j.jaim.2017.01.008

Ohkura, S. I., Hori, M., Saitoh, K., Okuzawa, T., Okamoto, I., Furukawa, N., et al. (2018). Structural and functional analysis of miraculin-like protein from *Vitis vinifera*. *Biochim. Biophys. Acta Proteins Proteom.* 1866, 1125–1130. doi: 10.1016/j.bbapap.2018.08.009

Ou, S., Chen, J., and Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* 46:e126. doi: 10.1093/nar/gky730

Paladino, A., Costantini, S., Colonna, G., and Facchiano, A. M. (2008). Molecular modelling of miraculin: structural analyses and functional hypotheses. *Biochem. Biophys. Res. Commun.* 367, 26–32. doi: 10.1016/j.bbrc.2007.12.102

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

Ranallo-Benavidez, T. R., Jaron, K. S., and Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* 11:1432. doi: 10.1038/s41467-020-14998-3

Sakai, H., Lee, S. S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., et al. (2013). Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol.* 54:e6. doi: 10.1093/pcp/pcs183

Sanematsu, K., Kitagawa, M., Yoshida, R., Nirasawa, S., Shigemura, N., and Ninomiya, Y. (2016). Intracellular acidification is required for full activation of the sweet taste receptor by miraculin. *Sci. Rep.* 6:22807. doi: 10.1038/srep22807

Shi, Y. C., Lin, K. S., Yi-Fen, J., Bao-Hong, L., Han, Y., Cui, Z., et al. (2016). Miracle fruit (*Synsepalum dulcificum*) exhibits as a novel anti-hyperuricaemia agent. *Molecules* 21:140. doi: 10.3390/molecules21020140

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Soza, V. L., Lindsley, D., Waalkes, A., Ramage, E., Patwardhan, R. P., Burton, J. N., et al. (2019). The rhododendron genome and chromosomal organization provide insight into shared whole-genome duplications across the heath family (Ericaceae). *Genome Biol. Evol.* 11, 3353–3371. doi: 10.1093/gbe/evz245

Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* 32, W309–W312. doi: 10.1093/nar/gkh379

Sun, P., Jiao, B., Yang, Y., Shan, L., Li, T., Li, X., et al. (2021). WGDI: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *bioRxiv* [Preprint] bioRxiv 2021.2004.2029.441969,

Swamy, K. B., Hadi, S. A., Sekaran, M., and Pichika, M. R. (2014). The clinical effects of *Synsepalum dulcificum*: a review. *J. Med. Food* 17, 1165–1169.

Takai, A., Satoh, M., Matsuyama, T., Ito, A., Nakata, R., Aoyama, T., et al. (2013). Secretion of miraculin through the function of a signal peptide conserved in the Kunitz-type soybean trypsin inhibitor family. *FEBS Lett.* 587, 1767–1772. doi: 10.1016/j.febslet.2013.04.026

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* Chapter 4, Unit4.10.

Tchokponhoué, D. A., Achigan-Dako, E. G., N'danikou, S., Nyadanu, D., Kahane, R., Houéto, J., et al. (2020). Phenotypic variation, functional traits repeatability and core collection inference in *Synsepalum dulcificum* (Schumach & Thonn.) Daniell reveals the Dahomey Gap as a centre of diversity. *Sci. Rep.* 10:19538. doi: 10.1038/s41598-020-76103-4

Tchokponhoué, D. A., Achigan-Dako, E. G., N'danikou, S., Nyadanu, D., Kahane, R., Odindo, A. O., et al. (2021). Comparative analysis of management practices and end-users' desired breeding traits in the miracle plant [*Synsepalum dulcificum* (Schumach & Thonn.) Daniell] across ecological zones and sociolinguistic groups in West Africa. *J. Ethnobiol. Ethnomed.* 17:41. doi: 10.1186/s13002-021-00467-8

Tchokponhoué, D. A., N'danikou, S., and Achigan-Dako, E. G. (2019a). A combination of approaches evidenced seed storage behaviour in the miracle berry *Synsepalum dulcificum* (Schumach. et Thonn.) Daniell. *BMC Plant Biol.* 19:117. doi: 10.1186/s12870-019-1714-1

Tchokponhoué, D. A., N'danikou, S., Houéto, J. S., and Achigan-Dako, E. G. (2019b). Shade and nutrient-mediated phenotypic plasticity in the miracle plant *Synsepalum dulcificum* (Schumach. & Thonn.) Daniell. *Sci. Rep.* 9:5135. doi: 10.1038/s41598-019-41673-5

Tsukuda, S., Gomi, K., Yamamoto, H., and Akimitsu, K. (2006). Characterization of cDNAs encoding two distinct miraculin-like proteins and stress-related modulation of the corresponding mRNAs in *Citrus jambhiri* lush. *Plant Mol. Biol.* 60, 125–136. doi: 10.1007/s11103-005-2941-4

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Wang, H. M., Chou, Y. T., Hong, Z. L., Chen, H. A., Chang, Y. C., Yang, W. L., et al. (2011). Bioconstituents from stems of *Synsepalum dulcificum* Daniell (Sapotaceae) inhibit human melanoma proliferation, reduce mushroom tyrosinase activity and have antioxidant properties. *J. Taiwan Inst. Chem. Eng.* 42, 204–211.

Wang, Y., Chen, F., Ma, Y., Zhang, T., Sun, P., Lan, M., et al. (2021). An ancient whole-genome duplication event and its contribution to flavor compounds in the tea plant (Camellia sinensis). *Hortic. Res.* 8:176. doi: 10.1038/s41438-021-00613-z

Wu, H., Ma, T., Kang, M., Ai, F., Zhang, J., Dong, G., et al. (2019). A high-quality Actinidia chinensis (kiwifruit) genome. *Hortic. Res.* 6:117. doi: 10.1038/s41438-019-0202-y

Xia, E., Tong, W., Hou, Y., An, Y., Chen, L., Wu, Q., et al. (2020). The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation. *Mol. Plant* 13, 1013–1026. doi: 10.1016/j.molp.2020.04.010

Xia, E. H., Li, F. D., Tong, W., Li, P. H., Wu, Q., Zhao, H. J., et al. (2019). Tea plant information archive: a comprehensive genomics and bioinformatics platform for tea plant. *Plant Biotechnol. J.* 17, 1938–1953. doi: 10.1111/pbi.13111

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., et al. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 47, W52–W58. doi: 10.1093/nar/gkz333

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268. doi: 10.1093/nar/gkm286

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591.

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287.

Zapata, L., Ding, J., Willing, E. M., Hartwig, B., Bezdan, D., Jiao, W. B., et al. (2016). Chromosome-level assembly of *Arabidopsis thaliana* Ler reveals the extent of translocation and inversion polymorphisms. *Proc. Natl. Acad. Sci. U.S.A.* 113, E4052–E4060. doi: 10.1073/pnas.1607532113

Zhang, C., Zhang, T., Luebert, F., Xiang, Y., Huang, C.-H., Hu, Y., et al. (2020). Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. *Mol. Biol. Evol.* 37, 3188–3210.

Zhang, Z., Li, J., Zhao, X.-Q., Wang, J., Wong, G. K.-S., and Yu, J. (2006). KaKs_calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* 4, 259–263.

Zhu, Q. G., Xu, Y., Yang, Y., Guan, C. F., Zhang, Q. Y., Huang, J. W., et al. (2019). The persimmon (*Diospyros oleifera* Cheng) genome provides new insights into the inheritance of astringency and ancestral evolution. *Hortic. Res.* 6:138. doi: 10.1038/s41438-019-0227-2

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

MINI REVIEW
published: 13 January 2022
doi: 10.3389/fpls.2021.825289

Genetic Basis of Carnivorous Leaf Development

*Arpita Agrawal[1], Ashwani Pareek[2,3] and Jeremy Dkhar[1,4]**

[1] Plant EvoDevo Laboratory, Agrotechnology Division, CSIR-Institute of Himalayan Bioresource Technology, Palampur, India, [2] Stress Physiology and Molecular Biology Laboratory, School of Life Sciences, Jawaharlal Nehru University, New Delhi, India, [3] National Agri-Food Biotechnology Institute, Mohali, India, [4] Academy of Scientific and Innovative Research, Ghaziabad, India

Plant carnivory is often manifested as dramatic changes in the structure and morphology of the leaf. These changes appear to begin early in leaf development. For example, the development of the *Sarracenia purpurea* leaf primordium is associated with the formation of an adaxial ridge, whose growth along with that of the leaf margin resulted in a hollow structure that later developed into a pitcher. In *Nepenthes khasiana*, pitcher formation occurs during the initial stages of leaf development, although this has not been shown at the primordial stage. The formation of the *Utricularia gibba* trap resulted from the growth of the dome-shaped primordium in both the longitudinal and transverse directions. Recent research has begun to unfold the genetic basis of the development of the carnivorous leaf. We review these findings and discuss them in relation to the flat-shaped leaves of the model plant *Arabidopsis*.

Keywords: carnivorous plants, leaf development, *Nepenthes*, *Utricularia*, *Sarracenia*

OPEN ACCESS

Edited by:
Jeremy D. Rentsch,
Francis Marion University,
United States

Reviewed by:
Andrej Pavlovič,
Palacký University Olomouc, Czechia

***Correspondence:**
Jeremy Dkhar
jeremydkhar@ihbt.res.in

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

Received: *30 November 2021*
Accepted: *23 December 2021*
Published: *13 January 2022*

Citation:
Agrawal A, Pareek A and Dkhar J (2022) Genetic Basis of Carnivorous Leaf Development. Front. Plant Sci. 12:825289. doi: 10.3389/fpls.2021.825289

INTRODUCTION

Carnivorous plants develop a set of morphological features termed "carnivorous syndrome", which facilitate the capture and digestion of attracted prey (Pavlovič et al., 2007). This syndrome is manifested mostly on the leaves – as innovative morphological structures (**Figure 1**) – and has been the focus of research ever since the publication of Charles Darwin's book *Insectivorous plants* (Darwin, 1875; Ellison and Gotelli, 2009). Studies have now shed light on why plant carnivory evolved, leading to a deeper understanding of the underlying mechanisms governing prey entrapment and digestion. It took a 140 years later, through an excellent study involving the purple pitcher plant *S. purpurea* (Fukushima et al., 2015), that the developmental basis of these highly specialized leaves became known. A few years later, the developing leaf transcriptome of *N. khasiana* was reported, hinting at the possible link between leaf polarity genes and pitcher formation (Dkhar and Pareek, 2019). Another outstanding study from the Coen lab provided insights on the formation of the *U. gibba* trap (Whitewoods et al., 2020). These studies have now begun to reveal the molecular mechanisms underpinning the development of the carnivorous leaf, shaped either as a cup (in *U. gibba*) or a pitcher (in *S. purpurea*). These innovative leaf morphologies appear to have evolved independently through changes in the existing genetic mechanisms governing flat-shaped leaf development (e.g., *Arabidopsis*, **Figure 2A**) (Lee et al., 2019).

LEAF ORGANOGENESIS AT THE SHOOT APICAL MERISTEM OF CARNIVOROUS PLANTS

Leaves arise from the peripheral zone of the SAM (Bowman and Floyd, 2008). In *Arabidopsis* and other model plant organisms, the process begins with the recruitment of leaf founder cells (Dkhar and Pareek, 2014). In carnivorous plants, initiation of the leaf primordium might also involve founder cells recruitment. This recruitment is critical to determine the region at the flanks of SAM where primordia will initiate. Initiation is then followed by the establishment of the adaxial/abaxial polarity along with the bulging of the leaf primordia (Tsukaya, 2013). The developmental processes leading to the protrusion of the incipient leaf primordia appear to be common among land plants. As development progresses, structural changes begin to appear on the leaf primordia of carnivorous plants. In *S. purpurea*, the adaxial side of the developing leaf primordium becomes elevated, forming a ridge (Fukushima et al., 2015). Further growth and development of the adaxial ridge resulted in the formation of a hollow structure at the distal part of the primordium, which eventually develops into a pitcher (Fukushima et al., 2015). In *Arabidopsis*, the developing leaf primordium remains flat until maturity (Matsumoto and Okada, 2001) (**Figure 2A**). In *U. gibba*, trap develops laterally on a leaf. But not all *U. gibba* leaf bears a trap; rather, filiform-shaped leaflets develop. What specifies a *U. gibba* leaf to develop a trap instead of a leaflet is inherently genetic (discussed below). Interestingly, both leaflet and trap primordia look alike during the initial stages of development (Whitewoods et al., 2020). Changes appear at later stages: leaflet primordia turn into cylinders with tapered ends whereas trap primordia become spherical in shape, due to growth in both the longitudinal and transverse directions (Whitewoods et al., 2020). In *Nepenthes*, pitchers are attached to a flattened leaf base lamina *via* a tendril. It was previously thought that pitcher initiation occurs at the tip of a tendril (Owen and Lennon, 1999). However, recent evidence suggests that pitcher formation in *N. khasiana* occurs early in leaf development and shares anatomical features with the young in-rolled leaf base lamina (Dkhar and Pareek, 2019). It remains to be seen though, how pitcher initiation occurs on the leaf primordium of *Nepenthes*.

TRAP MORPHOGENESIS IN *Utricularia gibba* IS A RESULT OF RESTRICTED GENE EXPRESSION

How does a *U. gibba* plant direct its leaf to develop a trap, instead of a leaflet? The answer lies in the expression pattern of the *PHAVOLUTA* (*PHV*) gene, a member of the class III *HD-ZIP* gene family known to specify the adaxial identity of lateral organs (McConnell et al., 2001). The *Arabidopsis* genome contains five *HD-ZIPIII* genes namely *PHABOLUSA* (*PHB*), *REVOLUTA* (*REV*), *ATHB8*, *ATHB15,* and *PHV* (Emery et al., 2003). In *U. gibba*, *PHV* showed extended as well as restricted expression patterns on the adaxial side of the developing leaf primordia (Whitewoods et al., 2020). This differential gene expression pattern carries biological significance: primordia showing restricted *PHV* expression developed into traps. When this restricted expression pattern is perturbed, through heat-shock-induced ectopic expression of *PHV* that is altered to prevent recognition by miRNA (preferably miR165), allowing *PHV* expression throughout the leaf tissue, trap development is significantly reduced (Whitewoods et al., 2020). Thus, trap initiation in *U. gibba* is a result of restricted *PHV* expression. In *S. purpurea*, however, no distinct *PHB* or *FIL* (*FILAMENTOUS FLOWER*, a gene specifying abaxial identity) expression patterns were seen between the hollow and the ridge regions, indicating that neither of the two genes play a role in pitcher initiation. Instead, changes in the orientation of cell division in the developing leaf primordium led to the development of the pitcher-shaped leaf in *S. purpurea* (Fukushima et al., 2015).

Arabidopsis REVOLUTA MUTANTS DISPLAY *Nepenthes* LEAF PHENOTYPE

Arabidopsis mutants of the *HD-ZIPIII* genes viz. *PHB*, *PHV* and *REV* display leaf phenotypes similar to those seen in pitcher plants, particularly *Nepenthes*. These gain-of-function mutations transform the flat-shaped leaves of *Arabidopsis* into trumpet-shaped (McConnell and Barton, 1998; McConnell et al., 2001; Emery et al., 2003; Zhong and Ye, 2004). In severe cases, these trumpet-shaped leaves grew out from the abaxial midvein of the slightly narrower leaf lamina (**Figure 2B**), resembling the pitcher-shaped leaves of *Nepenthes*. Unlike *phb* and *phv* gain-of-function mutants, the inside surface of the trumpet-shaped leaf of gain-of-function *rev* mutants is adaxial, as is the case with *Nepenthes* pitchers. The *rev* phenotype is due to a single nucleotide change in the putative lipid/sterol-binding START domain of the *REV* gene, which prevented recognition and thereby negative regulation by miR165 resulting in the expanded expression of *REV* (Emery et al., 2003). In *N. khasiana*, *REV* showed increased expression in the tip of the young *N. khasiana* leaf that later developed into a pitcher (Dkhar and Pareek, 2019). We now know that increased and restricted expression pattern of *PHV* resulted in trap formation in *U. gibba*. It is likely then that the increased and expanded expression of *REV* probably led to the development of the *Nepenthes* pitcher.

AUXIN AND ITS ROLE IN PITCHER DEVELOPMENT

Trumpet-shaped leaves were also seen in *Arabidopsis* plants with defective *PIN1* and *REV* genes, but not on single *pin1* or *rev Arabidopsis* mutants (Qi et al., 2014). *PIN1* encodes for an auxin efflux carrier protein that mediates local auxin accumulation during leaf initiation (Reinhardt et al., 2003). Based on their observations, Qi et al. (2014) suggested that auxin and *REV* act independently to promote leaf polarity. Fukushima et al. (2015) evaluated the effects of auxin on pitcher development in *S. purpurea*. Neither the addition of

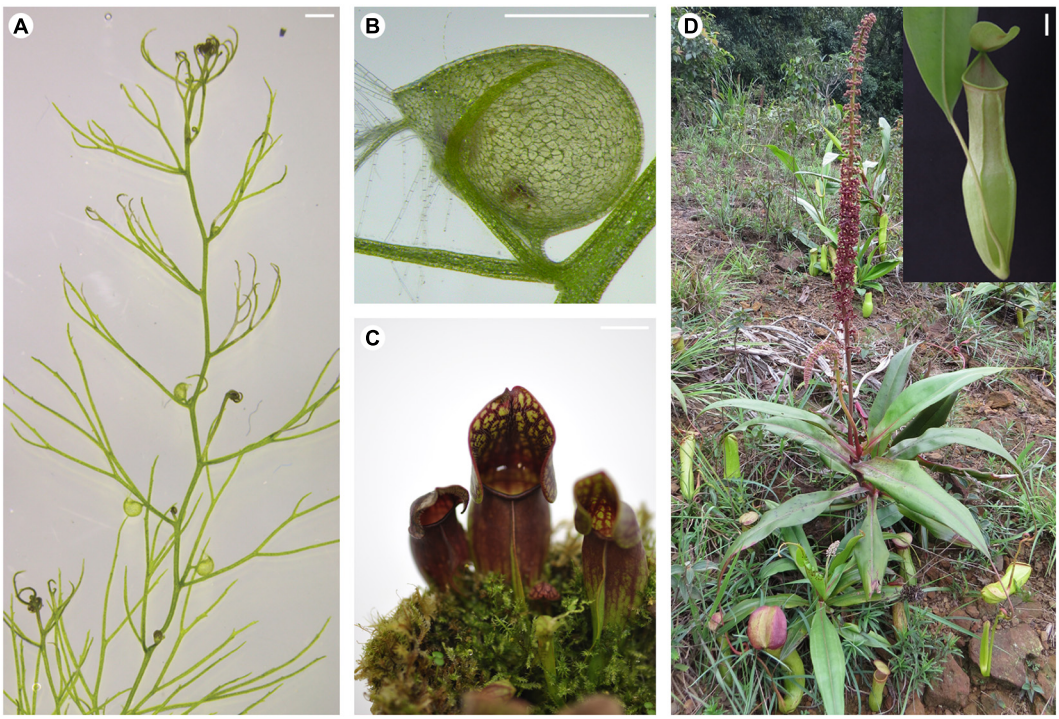


FIGURE 1 | Innovative leaf morphologies of selected carnivorous plants. **(A)** Stolon of a *U. gibba* plant showing leaves with or without traps. **(B)** Close-up view of a *U. gibba* trap. **(C)** Pitcher-shaped leaves of *S. purpurea*. **(D)** A *N. khasiana* plant in the wild bearing pitcher-shaped leaves (inset shows close-up view of the *N. khasiana* pitcher). **(A,B)** Bar = 1 mm; **(C,D)** bar = 10 mm.

1-Naphthaleneacetic acid (NAA) nor 1-N-Naphthylphthalamic acid (NPA), an auxin transport inhibitor, prevented the formation of the ridge and the hollow regions in the leaf primordia of *S. purpurea*. In *N. khasiana*, addition of NPA at 5, 10, 20, and 40 μM concentrations to the growing 1/4 strength MS medium increases shoot branching, as a result of axillary bud growth (**Figures 3A,B**). At higher NPA concentration, pitchers of *N. khasiana* failed to develop, although a tiny structure at the tip of the leaf can still be seen (**Figure 3A**). Altogether, these results suggest that auxin might have an indirect role in the formation (or the expansion) of the *N. khasiana* pitcher.

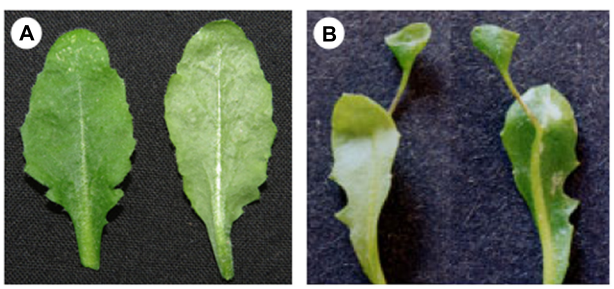


FIGURE 2 | Leaf phenotypes of wild-type and mutant *Arabidopsis* plants. **(A)** Flat-shaped leaves of wild-type *Arabidopsis* plant. **(B)** Trumpet-shaped leaves of *rev* gain-of-function mutant *Arabidopsis* plant. Note the resemblance of the mutant leaf with those of *Nepenthes* (inset in **Figure 1D**). Images on the left and right of **(A,B)** correspond to the adaxial and abaxial sides, respectively. **(B)** Is reproduced from Zhong and Ye (2004) with permission from Oxford University Press, United Kingdom.

TEMPERATURE AND NUTRIENT AVAILABILITY AFFECT PITCHER FORMATION

Leaf shape can vary to change in environmental conditions such as temperature and light (Dkhar and Pareek, 2014). Interestingly, carnivorous plants also respond to change in temperature and nutrient availability. When the Australian pitcher plant *Cephalotus follicularis* was grown at 15°C under continuous light conditions, more flat leaves developed than pitcher-shaped leaves (Fukushima et al., 2017). At 25°C, however, the reverse was observed. Could this be a result of the plant's response to higher temperatures? We know that *Nepenthes*, including *N. khasiana*, grow in nutrient-deficient soil. So, upon adding sufficient nutrients to the substratum, prey-deprived *N. talangensis* plants respond to this change in condition by developing leaves that lack pitchers (Pavlovič et al., 2010). We replicated this experiment *in vitro* by growing dissected shoot apices of 3-months old *N. khasiana* seedlings in MS medium of varying strengths (Dkhar et al., unpublished data). Most shoot apices of *N. khasiana* grown in 2 MS (twice the strength of full MS i.e., 1 MS) failed to grow, but if they survive, the plantlets develop narrower leaves

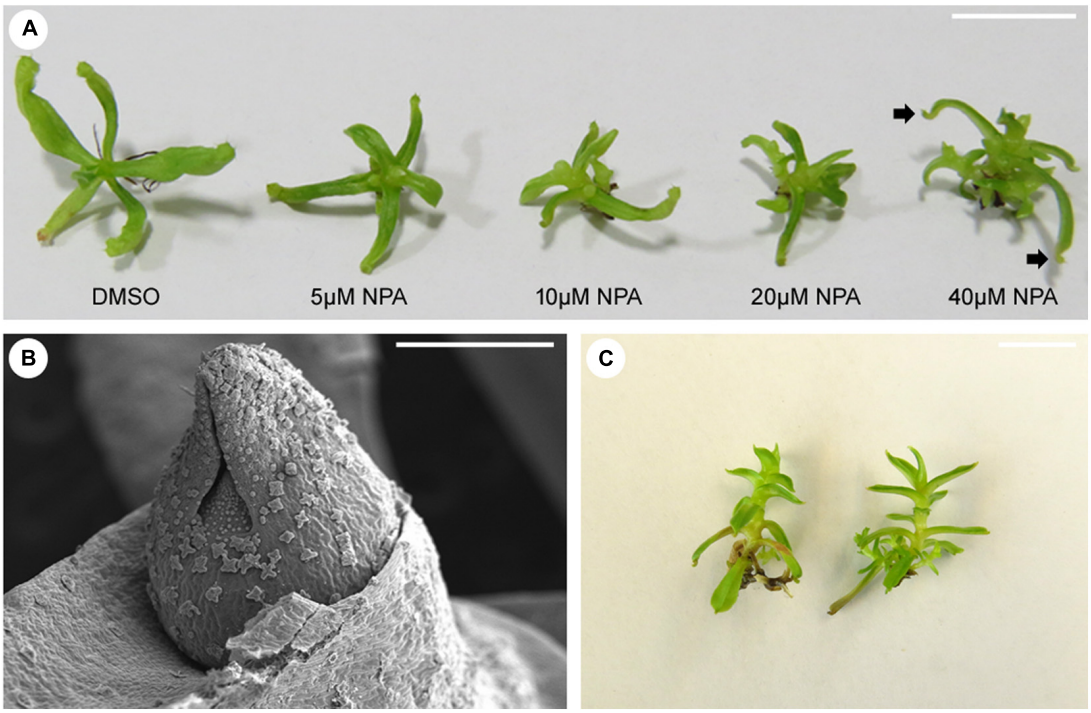


FIGURE 3 | Effects of auxin transport inhibition and nutrient availability on pitcher formation in *N. khasiana*. **(A)** Morphological phenotypes of *N. khasiana* plantlets after 1 month of treatment with varying concentrations of NPA (arrow points to tiny structures at the tip of the leaf). **(B)** A scanning electron micrograph of an axillary bud of *N. khasiana*, arising as a result of NPA treatment. **(C)** *N. khasiana* plantlets grown in 2 MS (twice the strength of full MS i.e., 1 MS). Note the dramatic changes in leaf morphology. **(A,C)** Bar = 10 mm; **(B)** bar = 0.5 mm.

completely devoid of pitchers (**Figure 3C**). By modifying nutrient content in the growing medium, we are now presented with *N. khasiana* plants lacking pitchers. It would be interesting to know how nutrient availability signals the plant to develop or not to develop pitchers.

CONCLUSION

Carnivorous plants evolved pitcher- or cup-shaped leaves independently in four angiosperm lineages viz. Cephalotaceae (*Cephalotus*), Lentibulariaceae (*Utricularia*), Nepenthaceae (*Nepenthes*) and Sarraceniaceae (*Sarracenia*) (Whitewoods and Coen, 2017). Although these leaves may look complex in shape and form, research suggests that their formation may be a result of simple modification – through changes in gene expression patterns and growth modulation – of existing leaf developmental programs at work in simpler leaf forms. Studies on the model plant *Arabidopsis* have led to a deeper understanding of the underlying mechanisms controlling leaf development, which now form the basis of ongoing research in the study of the evolution and development of innovative leaf morphologies seen across angiosperms, including carnivorous plants. The tremendous progress seen in *Arabidopsis* is, in part, attributed to the ease at which this plant can be genetically transformed, allowing various genes for leaf development to be functionally validated. Recent progress in this direction, such as the establishment of *Agrobacterium*-mediated transformation protocols in *N. mirabilis* (Miguel et al., 2020) and *U. gibba* (Oropeza-Aburto et al., 2020), offers much-needed hope for a complete genetic dissection of these complex leaf shapes.

AUTHOR CONTRIBUTIONS

JD conceived the review. AA wrote the initial manuscript, which was then corrected and finalized by JD. AP did the final editing. All authors read and agreed with the final content of the manuscript.

FUNDING

Research in JD lab is supported by the Council of Scientific and Industrial Research (CSIR), Government of India (MLP 202) and Science and Engineering Research Board (SERB), Government of India [(SRG/2020/001139) (GAP 273)].

ACKNOWLEDGMENTS

The authors acknowledge the Editor and a Reviewer for critical reading and suggestions to improve the manuscript. JD acknowledges Sanjay Kumar, Director, CSIR-IHBT, Palampur, for providing the necessary facilities needed during the preparation

of this review article, Kamlesh Kant Nutan (Pareek Lab) for the *Arabidopsis* image in **Figure 2A**, and the permission granted by Oxford University Press, United Kingdom to use the image in **Figure 2B**, which is reproduced from Zhong and Ye (2004). This review would not have been complete without the help of Enrico Coen, Department of Cell and Developmental Biology, John Innes Centre, United Kingdom and Kenji Fukushima, Department for Molecular Plant-Physiologie and Biophysics, University of Wuerzburg, Germany. Special thanks to Chris Whitewoods and Karen Lee (Coen Lab) for the *U. gibba* images used in **Figures 1A,B** and Matthias Freund (Fukushima Lab) for the *S. purpurea* image used in **Figure 1C**. This mini review represents CSIR-IHBT communication number 4969.

REFERENCES

Bowman, J. L., and Floyd, S. K. (2008). Patterning and polarity in seed plant shoots. *Ann. Rev. Plant Biol.* 59, 67–88. doi: 10.1146/annurev.arplant.57.032905.105356

Darwin, C. (1875). *Insectivorous Plants*. New York: D. Appleton and Company. doi: 10.5962/bhl.title.99933

Dkhar, J., and Pareek, A. (2014). What determines a leaf's shape? *EvoDevo* 5:47. doi: 10.1186/2041-9139-5-47

Dkhar, J., and Pareek, A. (2019). ASYMMETRIC LEAVES1 and REVOLUTA are the key regulatory genes associated with pitcher development in Nepenthes khasiana. *Sci. Rep.* 9:6318. doi: 10.1038/s41598-019-42779-6

Ellison, A. M., and Gotelli, N. J. (2009). Energetics and the evolution of carnivorous plants—Darwin's 'most wonderful plants in the world'. *J. Exp. Bot.* 60, 19–42. doi: 10.1093/jxb/ern179

Emery, J. F., Floyd, S. K., Alvarez, J., Eshed, Y., Hawker, N. P., Izhaki, A., et al. (2003). Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. *Curr. Biol.?* 13, 1768–1774. doi: 10.1016/j.cub.2003.09.035

Fukushima, K., Fang, X., Alvarez-Ponce, D., Cai, H., Carretero-Paulet, L., Chen, C., et al. (2017). Genome of the pitcher plant Cephalotus reveals genetic changes associated with carnivory. *Nat. Ecol. Evol.* 1:59. doi: 10.1038/s41559-016-0059

Fukushima, K., Fujita, H., Yamaguchi, T., Kawaguchi, M., Tsukaya, H., and Hasebe, M. (2015). Oriented cell division shapes carnivorous pitcher leaves of *Sarracenia purpurea*. *Nat. Commun.* 6:6450. doi: 10.1038/ncomms7450

Lee, K. J. I., Bushell, C., Koide, Y., Fozard, J. A., Piao, C., Yu, M., et al. (2019). Shaping of a three-dimensional carnivorous trap through modulation of a planar growth mechanism. *PLoS Biol.* 17:e3000427. doi: 10.1371/journal.pbio.3000427

Matsumoto, N., and Okada, K. (2001). A homeobox gene, PRESSED FLOWER, regulates lateral axis-dependent development of Arabidopsis flowers. *Genes Dev.* 15, 3355–3364. doi: 10.1101/gad.931001

McConnell, J. R., and Barton, M. K. (1998). Leaf polarity and meristem formation in Arabidopsis. *Development* 125, 2935–2942. doi: 10.1242/dev.125.15.2935

McConnell, J. R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M. K. (2001). Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. *Nature?* 411, 709–713. doi: 10.1038/35079635

Miguel, S., Michel, C., Biteau, F., Hehn, A., and Bourgaud, F. (2020). In vitro plant regeneration and Agrobacterium-mediated genetic transformation of a carnivorous plant, *Nepenthes mirabilis*. *Sci. Rep.* 10:17482. doi: 10.1038/s41598-020-74108-7

Oropeza-Aburto, A., Cervantes-Pérez, S. A., Albert, V. A., and Herrera-Estrella, L. (2020). *Agrobacterium tumefaciens* mediated transformation of the aquatic carnivorous plant *Utricularia gibba*. *Plant Methods* 16:50. doi: 10.1186/s13007-020-00592-7

Owen, T. P., and Lennon, K. A. (1999). Structure and development of the pitchers from the carnivorous plant *Nepenthes alata* (Nepenthaceae). *Am. J. Bot.* 86, 1382–1390. doi: 10.2307/2656921

Pavlovič, A., Masarovičová, E., and Hudák, J. (2007). Carnivorous syndrome in Asian pitcher plants of the genus Nepenthes. *Ann. Bot.* 100, 527–536. doi: 10.1093/aob/mcm145

Pavlovič, A., Singerová, L., Demko, V., Šantrůček, J., and Hudák, J. (2010). Root nutrient uptake enhances photosynthetic assimilation in prey-deprived carnivorous pitcher plant *Nepenthes talangensis*. *Photosynthetica* 48, 227–233. doi: 10.1007/s11099-010-0028-1

Qi, J., Wang, Y., Yu, T., Cunha, A., Wu, B., Vernoux, T., et al. (2014). Auxin depletion from leaf primordia contributes to organ patterning. *Proc. Natl. Acad. Sci. U. S. A.* 111, 18769–18774. doi: 10.1073/pnas.1421878112

Reinhardt, D., Pesce, E. R., Stieger, P., Mandel, T., Baltensperger, K., Bennett, M., et al. (2003). Regulation of phyllotaxis by polar auxin transport. *Nature* 426, 255–260. doi: 10.1038/nature02081

Tsukaya, H. (2013). Leaf development. *Arabidopsis Book* 11:e0163. doi: 10.1199/tab.0163

Whitewoods, C. D., and Coen, E. (2017). Growth and development of three-dimensional plant form. *Curr. Biol.* 27, R910–R918. doi: 10.1016/j.cub.2017.05.079

Whitewoods, C. D., Gonçalves, B., Cheng, J., Cui, M., Kennaway, R., Lee, K., et al. (2020). Evolution of carnivorous traps from planar leaves through simple shifts in gene expression. *Science* 367, 91–96. doi: 10.1126/science.aay5433

Zhong, R., and Ye, Z. H. (2004). amphivasal vascular bundle 1, a gain-of-function mutation of the IFL1/REV gene, is associated with alterations in the polarity of leaves, stems and carpels. *Plant Cell Physiol.* 45, 369–385. doi: 10.1093/pcp/pch051

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 06 June 2022
doi: 10.3389/fpls.2022.887635

Characterization and Comparison of Convergence Among *Cephalotus follicularis* Pitcher Plant-Associated Communities With Those of *Nepenthes* and *Sarracenia* Found Worldwide

Leonora S. Bittleston[1,2], Elizabeth L. Benson[2], Jessica R. Bernardin[1] and Naomi E. Pierce[2]*

[1]*Department of Biological Sciences, Boise State University, Boise, ID, United States,* [2]*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, United States*

OPEN ACCESS

Edited by:
Kenji Fukushima, Julius Maximilian University of Würzburg, Germany

Reviewed by:
Catalina Cuellar-Gempeler, Humboldt State University, United States
Alexandra Stoll, Universidad Católica del Norte, Chile

***Correspondence:**
Leonora S. Bittleston
leonorabittleston@boisestate.edu

Specialty section:
This article was submitted to Functional Plant Ecology, a section of the journal Frontiers in Plant Science

Received: *01 March 2022*
Accepted: *13 May 2022*
Published: *06 June 2022*

Citation:
Bittleston LS, Benson EL, Bernardin JR and Pierce NE (2022) Characterization and Comparison of Convergence Among Cephalotus follicularis Pitcher Plant-Associated Communities With Those of Nepenthes and Sarracenia Found Worldwide. Front. Plant Sci. 13:887635. doi: 10.3389/fpls.2022.887635

The Albany pitcher plant, *Cephalotus follicularis*, has evolved cup-shaped leaves and a carnivorous habit completely independently from other lineages of pitcher plants. It is the only species in the family Cephalotaceae and is restricted to a small region of Western Australia. Here, we used metabarcoding to characterize the bacterial and eukaryotic communities living in *C. follicularis* pitchers at two different sites. Bacterial and eukaryotic communities were correlated in both richness and composition; however, the factors associated with richness were not the same across bacteria and eukaryotes, with bacterial richness differing with fluid color, and eukaryotic richness differing with the concentration of DNA extracted from the fluid, a measure roughly related to biomass. For turnover in composition, the variation in both bacterial and eukaryotic communities primarily differed with fluid acidity, fluid color, and sampling site. We compared *C. follicularis*-associated community diversity with that of Australian *Nepenthes mirabilis*, as well as a global comparison of Southeast Asian *Nepenthes* and North American *Sarracenia*. Our results showed similarity in richness with communities from other pitcher plants, and specific bacterial taxa shared among all three independent lineages of pitcher plants. Overall, we saw convergence in richness and particular clades colonizing pitcher plants around the world, suggesting that these highly specialized habitats select for certain numbers and types of inhabitants.

Keywords: carnivorous plant, bacteria, eukaryote, convergent evolution, microbe, microbiome

INTRODUCTION

Carnivorous pitcher plants are a clear example of convergent evolution, where the same form and function has evolved independently across three different lineages (Albert et al., 1992). These "true" pitcher plants have pitcher-shaped, single leaves containing small ecosystems with complex communities of arthropods and microbes (Adlassnig et al., 2011). In recent years,

there has been a growing recognition of the importance of plant-associated communities beyond just the organisms that are clearly symbionts or pathogens (Bell et al., 2019; Leveau, 2019). Pitcher plants provide an opportunity to examine how hosts with similar morphology and life-history traits that have evolved independently in dramatically different geographic locations, can nevertheless attract and cultivate similar communities within their structures. Pitcher plant-associated communities from the genera *Sarracenia* (family Sarraceniaceae) and *Nepenthes* (family Nepenthaceae) have now been characterized across numerous studies (Peterson et al., 2008; Koopman and Carstens, 2011; Gray, 2012; Gray et al., 2012; Kanokratana et al., 2016; Bittleston et al., 2018; Grothjan and Young, 2019; Gilbert et al., 2020a,b). Many of the same bacterial and eukaryotic taxonomic groups are common in both *Sarracenia* and *Nepenthes* communities and their metagenomes show enrichment in specific degradation genes (Bittleston et al., 2018), suggesting that convergent interactions (Bittleston et al., 2016b) may exist among these convergently evolved plants and the organisms living in their pitchers. However, there is very little known about the communities associated with natural populations of the third pitcher plant lineage in the genus *Cephalotus*, and its bacterial microbiome has never been characterized.

The Albany pitcher plant, *C. follicularis* (**Figure 1A**), is the only genus and species in the family Cephalotaceae (Conran, 2004). It is endemic to the coastal region of South-Western Australia near the town of Albany, and few studies have examined the plant and its associated organisms in its natural habitat (Parkes and Hallam, 1984; Clarke, 1988; Yeates, 1992; Pavlovic, 2011; Lymbery et al., 2016; Just et al., 2019). Despite their small native range, the unusual *C. follicularis* plants are prized by collectors and have been cultivated and sold around the world (Conran, 2004). The plants grow in wet, peat-soil swamps often in seepage areas (Conran, 2004) and have been shown to get about 26% of their nitrogen from captured prey (Schulze et al., 1997). Recently, the genome of *C. follicularis* was sequenced, and transcriptomic analysis of carnivorous vs. non-carnivorous leaves identified genetic changes associated with carnivory (Fukushima et al., 2017). This diminutive and charismatic plant is listed as Vulnerable by the IUCN, due to population reduction from habitat loss and overcollection, highlighting the necessity for a better understanding of its ecology.

Carnivorous pitcher plants trap and digest insects as a strategy for gaining nitrogen, phosphorus and other nutrients that are lacking in the generally water-logged soils where they grow. Despite their consumption of insect prey, a living community of arthropods, protozoa, algae, fungi, and bacteria thrive in the aquatic pitcher pools (Kitching, 2000). Pitcher microbiomes likely assist in degradation of prey and support plant nutrient access (Butler et al., 2008; Luciano and Newell, 2017; Bittleston et al., 2018). In *Nepenthes* and *Sarracenia* species, pitcher-associated communities have been found to vary by host species and pitcher fluid pH and volume, among other characteristics (Kanokratana et al., 2016; Bittleston et al., 2018; Gilbert et al., 2020b). Pitcher plants, despite being unusual within the plant world, are an emerging model system for community ecology because their pitchers contain small, naturally replicated microcosms that assemble anew each time a pitcher opens (Srivastava et al., 2004; Peterson et al., 2008; Miller et al., 2018; Ellison and Gotelli, 2021).

The extreme geographic isolation of *C. follicularis* from other pitcher plants may lead to it hosting fewer metazoan inhabitants (inquilines). A study of aquatic organisms, ranging from bacteria to fish, found higher dispersal limitation in larger organisms (De Bie et al., 2012). Beaver (1985) and later Clarke (1988) hypothesized that metazoan food web complexity in *Nepenthes* pitcher plants had geographical underpinnings, related to factors such as landmass size, degree of spatial and temporal isolation, the size of the local species pool capable of colonizing pitchers and the number of *Nepenthes* species present. One of the few known pitcher inhabitants of *C. follicularis*, and the only insect, is the larvae of the fly *Badisis ambulans*. This inquiline has never been found outside of *C. follicularis* pitchers and its interaction with the plant may be mutualistic or parasitic (Yeates, 1992; Lymbery et al., 2016). The thesis of Sally A. Clarke examined *C. follicularis*-associated protists in depth, but did not list many metazoans.

To the best of our knowledge, this is the first metabarcoding study characterizing the entire bacterial and eukaryotic communities of *C. follicularis* pitcher pools and comparing them with those of other pitcher plant species. Here, we aim to: (1) characterize the bacteria and eukaryotes associated with *C. follicularis* pitchers using DNA metabarcoding, and to assess how they are affected by site and host characteristics and (2) investigate possible convergence in community richness and shared taxa due to the very distinctive and similar life form among pitcher plants. We ask: What organisms live in *C. follicularis* pitchers? How do growing site and characteristics of the pitcher fluid impact community composition and structure? Do different factors affect bacterial and eukaryotic communities? Are *C. follicularis* pitcher communities similar to those found in Australian *Nepenthes*? Do we see convergent interactions such as those found in *Sarracenia* and *Nepenthes*?

MATERIALS AND METHODS

Sampling in Australia

In January 2016, we collected fluid samples from 40 healthy *C. follicularis* pitchers, split evenly between two sites roughly 20 km apart near Albany, Western Australia designated Gull Rock Road (GRR) and Two Peoples Bay (TPB; GPS coordinates are not listed here to protect the vulnerable plants but are available from the authors upon request). We haphazardly selected 20 healthy pitchers on different plants at each site that had no visible damage. Although it is difficult to gauge pitcher time-since-opening in the field without marking unopened pitchers and returning over time, we aimed to select pitchers of different ages to get a full representation of the existing diversity. The entire pitcher fluid from each pitcher was collected using a sterile pipette for each sample and was transferred to sterile tubes. Fluid color was noted qualitatively by a single researcher and total volume was recorded before removing an aliquot of fluid for pH measurement

and sequencing. We measured the pH of the samples with Macherey-Nagel pH-Fix 4.5–10.0 and BDH VWR Analytical pH-Test 0.0–6.0 indicator strips. We preserved samples with a 1:1 volume of cetyl trimethylammonium bromide (CTAB) buffer at room temperature before transporting them to the laboratory and freezing them until DNA extraction. In addition to these samples of *C. follicularis*, 21 *N. mirabilis* pitchers were sampled near Bramston Beach, Queensland, Australia in May 2016 following the same protocol as the *C. follicularis* sample collection.

DNA Extraction and 16S- and 18S-Amplicon Next-Generation Sequencing

We extracted DNA from sample fluids using a phenol-chloroform bead-beating extraction method (Sambrook et al., 2001; Bittleston et al., 2018). DNA extracts were quantified with a Qubit fluorometer. DNA amplification and Illumina MiSeq next-generation sequencing were performed at Argonne National Laboratory's Environmental Sample Preparation and Sequencing Facility (ESPSF). We used the Earth Microbiome Project's original barcoded 16S (515F-806R, V4 region) and 18S (1391f-1510r, V9 region) primers that were adapted for the Illumina MiSeq (Amaral-Zettler et al., 2009; Stoeck et al., 2010; Caporaso et al., 2012). Each 25 µl PCR reaction contained 9.5 µl of MO BIO PCR Water (Certified DNA-Free), 12.5 µl of QuantaBio's AccuStart II PCR ToughMix (2x concentration, 1x final), 1 µl Golay barcode tagged Forward Primer (5 µM concentration, 200 pM final), 1 µl Reverse Primer (5 µM concentration, 200 pM final), and 1 µl of template DNA. The conditions for PCR were: 94°C for 3 min, with 35 cycles at 94°C for 45 s, 50°C for 60 s, and 72°C for 90 s; with a final extension of 10 min at 72°C. Amplicons were quantified using PicoGreen (Invitrogen) and a plate reader (Infinite 200 PRO, Tecan) before being pooled in equimolar amounts. This pool was then cleaned with AMPure XP Beads (Beckman Coulter), and quantified using a fluorometer (Qubit, Invitrogen). After quantification, the molarity of the pool was determined and it was diluted down to 2 nM, denatured, and then diluted to a final concentration of 6.75 pM with a 10% PhiX spike for sequencing on the Illumina MiSeq. The 16S and 18S amplicons were sequenced on separate $151 \times 12 \times 151$ bp MiSeq runs.

Amplicon sequence data for the Australian samples (*C. follicularis* and *N. mirabilis*) have been deposited in the Sequence Read Archive NCBI BioProject PRJNA810039.

Inclusion of Data From *Nepenthes*, *Sarracenia*, and Surrounding Habitats

In order to compare our samples from *C. follicularis* and *N. mirabilis* pitchers with other pitcher plant and environmental communities, we used existing data from Bittleston et al., 2018 (sequence data are available from BioProject PRJNA448553). The comparison samples are described in detail in Bittleston et al., 2018 and were collected, transported, extracted and sequenced in very similar ways.

The environmental samples represent the habitats were the *Nepenthes* and *Sarracenia* plants were growing, e.g., the bog water or soil directly next to the plants. Unfortunately, due to the restrictions of our Australian permits, we were not able to collect soil or bog water samples at our Australian sites. Our comparison with soil and bog water is limited to environmental samples collected from the direct surroundings of *Nepenthes* and *Sarracenia* in Southeast Asia and North America.

Sequencing Quality Control and Statistical Analyses

To be as fully comparable as possible, all samples (those collected in Australia in 2016 as well as those that had been previously collected elsewhere and sequenced) were analyzed together in the same way. We used QIIME2 versions 2018.4 and 2019.10 (Bolyen et al. 2019) on the Boise State computing cluster for quality control, Amplicon Sequence Variant (ASV) generation, taxonomic assignment, and phylogenetic trees. We denoised sequences and generated ASVs with the DADA2 (Callahan et al., 2016) plugin, and 16S taxonomy assignment with the classify-sklearn Naive Bayes classifier (Pedregosa et al., 2011), and a pre-trained classifier made with the Greengenes database, version 13_8 (McDonald et al., 2012; Bokulich et al., 2018). For 18S taxonomy assignment, we used the classify-consensus-vsearch method and the SILVA 128 database. For both loci, we used the QIIME2 SEPP plugin (Janssen et al., 2018) to build phylogenies. All downstream analyses were done in R (version 4.0.2, R Core Team, 2020).

We removed all sequences assigned as chloroplasts from the 16S data, and bacteria and archaea from the 18S data. To minimize misassigned sequences and to focus on organisms most likely to be associated with pitchers, we removed all observations with fewer than 10 sequences and samples with fewer than 1,000 sequences. We focused on *C. follicularis* communities first, creating a subset of just these samples and rarefying at 4,039 sequences for 16S and 2,704 for 18S. We ran generalized linear models (GLMs) using a Gamma distribution (log link function) to examine the association of bacterial and eukaryotic richness (measured as Effective Number of Species, a Hill number that is the exponential of the Shannon index; Jost, 2007; Chao et al., 2014) with site, DNA concentration, and pitcher fluid pH, volume, and color. We chose to use GLMs in order to measure all of our variables together to control for their effects on each other and to avoid transforming non-normally distributed data. We made sure that our data met the assumptions for a Gamma GLM (see R code in the Dataverse repository). For beta diversity analyses, we used unweighted UniFrac, a phylogenetic distance metric, and NMDS plots to visualize in ordination space. We ran PERMANOVAs with the function *adonis* from the vegan package (Oksanen et al., 2020) to measure the effects of categorical variables and their interactions, and mantel tests for continuous variables. The R package, metacoder (Foster et al., 2017) was used to parse and plot significant ASVs by fluid color based on taxonomy and relative abundance. Some of the Australian *N. mirabilis* samples had lower sequence numbers, so in our comparison of the *N. mirabilis* with *C. follicularis* samples, we rarefied at 1,306 for bacteria and 1,556 for eukaryotes.

For our phylogenetic tree figures, we focused on ASVs that were represented by at least 100 sequences, and that were present in at least 10% of each category for our *Cephalotus*, *Nepenthes*, or *Sarracenia* samples. This is the same criteria used previously (Bittleston et al., 2018) with the goal of avoiding spurious taxa that only randomly or occasionally appeared in pitchers, in order to focus on taxa more likely to be pitcher-associated. These were compared with ASVs from surrounding habitats, including bog water, soil, water held in fallen leaves, and water collected in sterile glass tubes. We used the online interactive tree of life, iToL, software to plot and annotate the trees. The annotations around the outside of the circular trees show the log of the cumulative relative abundance of each ASV, normalized by the number of samples in that category.

Our R code and all the necessary input files and metadata have been deposited in the Harvard Dataverse and are accessible at: https://doi.org/10.7910/DVN/WG1TI6.

RESULTS

Characterizing *Cephalotus follicularis*-Associated Communities

Cephalotus follicularis Pitcher Fluid Volume Varies Across Sites

Despite the proximity of our two sampling sites (roughly 20 km), the volume of pitcher fluid in *C. follicularis* pitchers significantly differed between GRR and TPB samples (Mann–Whitney U test, $Z = -2.2568$, $p = 0.02402$; **Supplementary Figure 1A**), with higher volumes on average at TPB.

Although there was a trend toward more brown fluid in pitchers at TPB, pitcher fluid color did not differ significantly between sites (Fisher's exact test of independence, $p = 0.06006$). Pitcher fluid pH and the total extracted DNA concentration from each sample also did not differ by site (Mann–Whitney U tests, $p > 0.05$), and fluid pH also (marginally) did not significantly differ by fluid color (Kruskal-Wallis test, $p = 0.06763$). Fluid pH and DNA concentration were not correlated with volume (linear models, $p > 0.05$); however, fluid pH and DNA concentration were correlated with each other (linear model, t value = 2.88, $p = 0.0065$), with more acidic pitchers having lower DNA concentrations (**Supplementary Figure 1B**). A similar effect of pH on DNA extraction success has been noted in a previous study in *Nepenthes* pitcher plant microbial communities (Gilbert et al., 2020b). A PCA biplot of pitcher pH, volume, and DNA concentration showed separation of the different factors among our samples (**Supplementary Figure 1C**).

Top Organisms in *Cephalotus follicularis*-Associated Communities

We measured community composition of *C. follicularis*-associated bacteria and eukaryotes by analyzing data from 16S and 18S rRNA amplicon sequencing. In the *C. follicularis* 16S sequence data, post-quality control we had 334,718 sequences across 36 samples with a mean of 9,298 sequences per sample and 709 distinct amplicon sequence variants (ASVs). Rarefaction curves showed that all samples reached a plateau in diversity before our rarefaction cutoff (**Supplementary Figure 2A**). The most abundant phyla were Proteobacteria, Firmicutes, and Bacteroidetes. The taxonomy, relative abundances, and NCBI BLAST results for the top 20 ASVs are reported in **Supplementary Dataset 1**, accounting for 49.0% of the sequences, and the most abundant families were Burkholderiaceae and Rhodanobacteriaceae, representing 39.7% and 28.0% of the top 20 ASV reads, respectively.

The top bacterial ASV, in the family Burkholderiaceae with the best BLASTn match to *Paraburkholderia aromaticivorans* (**Supplementary Dataset 1**), had the highest relative abundance of all ASVs at both sampling sites. Interestingly, the second most abundant ASV at each of our sites was in the genus *Rhodanobacter* but was represented by a different ASV at each site. In general, the *C. follicularis* microbiome was dominated by bacteria from the Gammaproteobacteria (**Figure 1B**).

In the 18S *C. follicularis* sequence data, post-quality control we had 668,485 sequences across 40 samples with a mean of 16,712 sequences per sample and 526 distinct ASVs. Rarefaction curves showed that most samples reached a plateau in diversity before our rarefaction cutoff (**Supplementary Figure 2B**). The most abundant groups were Metazoa and Fungi, which represented 41.2% and 40.4% of the top 20 ASV reads, respectively (**Figure 1B**). The taxonomy, relative abundances, and NCBI BLAST results for the top 20 ASVs are reported in **Supplementary Dataset 1**, accounting for 59.0% of the sequences.

Within Metazoa some of the most commonly identified taxa included: Chromadorea (roundworms), Diptera (flies), Araneae (spiders), and Adinetida (rotifers). The two most common fungal families found were Aspergillaceae and Mortierellaceae.

Bacterial and Eukaryotic Richness Are Correlated but Affected by Different Factors

The effective number of species is a measure of alpha diversity or species richness that is calculated as the exponential of the Shannon diversity of species or, as in this case, ASVs. This measure of richness also incorporates the evenness of the community. For bacteria, the effective number of species per sample ranged from 2.5 to 56.7, with a median of 11.1 and a mean of 13.7 (**Figure 1C**). Eukaryotic communities generally had lower richness with a range of 1.7 to 17, a median of 6 and a mean of 7.1 (**Figure 1C**). Our results indicate that bacterial and eukaryotic richness are correlated with each other (**Table 1**; **Figure 1D**); however, surprisingly, the significant effect is only seen in the bacterial GLM but not the eukaryotic one. This result suggests that the diversity of bacterial taxa increases in response to increased diversity of eukaryotes, although it could also be a response to unmeasured environmental or host factors.

Variation in richness in each of the two broad taxonomic groups is associated with different factors. When controlling for other variables, bacterial but not eukaryotic richness significantly varies with color of the pitcher fluid (specifically clear vs. brown and green) but not site, volume or DNA concentration, while the 18S GLM indicates that eukaryotic richness significantly varies only with DNA concentration (**Table 1**). Effect plots for the bacterial and eukaryotic GLMs are shown in **Supplementary Figures 3**, **4**, respectively.

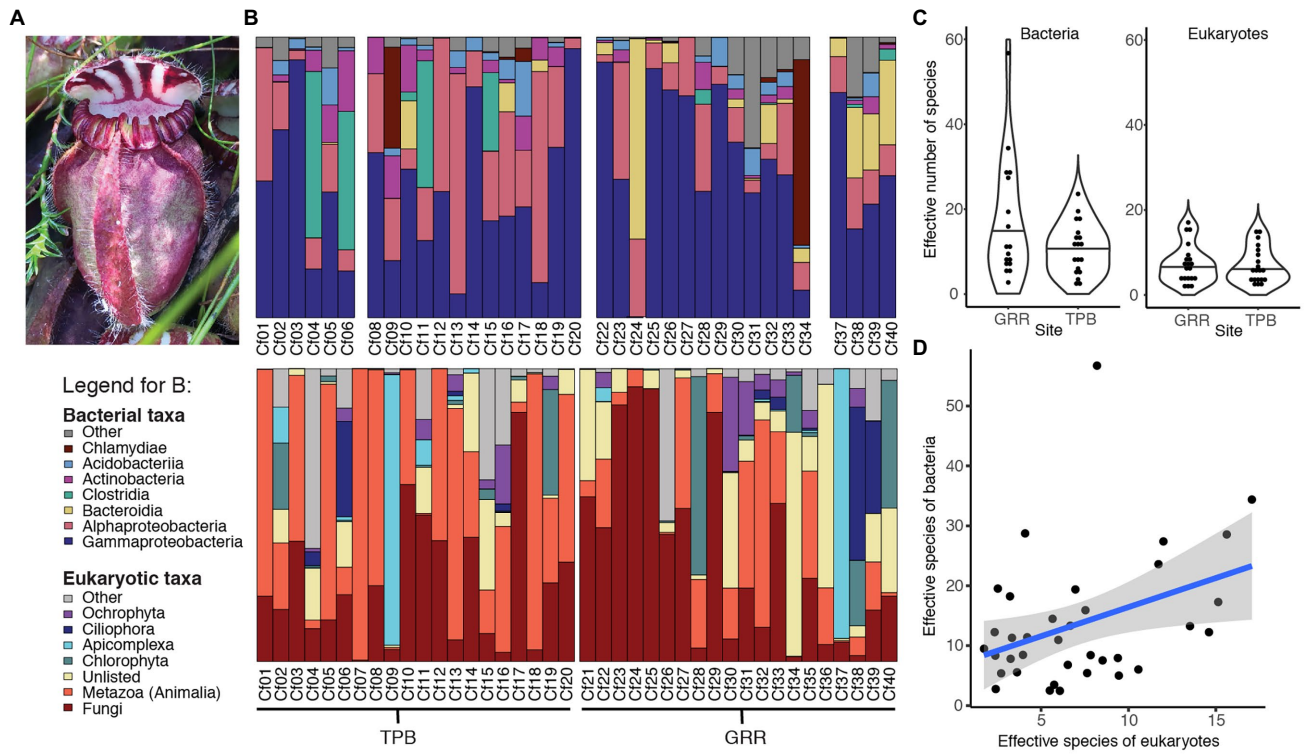


FIGURE 1 | Characterizing *Cephalotus follicularis*-associated communities. **(A)** Image of *C. follicularis* pitcher in its natural habitat (Photo credit: L.S. Bittleston). **(B)** Relative abundances of the most common bacterial (top) and eukaryotic (bottom) taxa across our samples. The first 20 were from site Two Peoples Bay (TPB) and the second set from Gull Rock Road (GRR). Some bacterial samples did not sequence well (Cf07, 21, 35, and 36); those columns are left blank. **(C)** Richness, measured as effective number of species [the exponential of Shannon diversity of the amplicon sequence variants (ASVs)] by site for bacteria (left) and eukaryotes (right). **(D)** Bacterial and eukaryotic richness (effective numbers of species) are correlated.

TABLE 1 | Factors affecting bacterial and eukaryotic community richness [generalized linear model (GLM) results].

Factor	16S—Bacteria		18S—Eukaryotes	
	t Value	Pr(>|*t*|)	*t* Value	Pr(>|*t*|)
(Intercept)	2.845	**0.00854**	1.937	0.0637
pH	−0.522	0.60632	0.353	0.727
Volume	−0.981	0.33561	1.377	0.1802
Site TPB	−1.839	0.07731	−1.08	0.2899
Fluid color, brown	2.14	**0.04194**	1.311	0.2014
Fluid color, cloudy	1.501	0.14531	0.196	0.8464
Fluid color, green	2.772	**0.01016**	0.769	0.449
Fluid color, yellow	1.108	0.27796	−0.646	0.5237
DNA concentration	0.681	0.50173	−2.315	**0.0288**
18S or 16S Effective number of species	2.503	**0.01894**	1.319	0.1985

Bolded values are significant at alpha = 0.05.

Turnover in Community Composition Varies Significantly by Pitcher Fluid Characteristics, by Site, and Between Bacterial and Eukaryotic Communities

Despite our two sites being less than 20 km apart, both pitcher bacterial and eukaryotic communities show significant differences in compositional turnover (beta diversity) by site, which explains ~4–5% of the existing variation (PERMANOVA, Bacteria $R^2 = 0.04215$, $p = 0.010$; Eukaryotes $R^2 = 0.04755$, $p = 0.001$; **Table 2**). Our NMDS plots indicate separation by site diagonally along both axes 1 and 2 for bacterial and eukaryotic communities (**Figures 2A,B**). For both bacterial and eukaryotic community compositions, a large portion of the variation is explained by fluid pH (mantel tests: Bacteria $r = 0.2837$, $p = 0.001$; Eukaryotes $r = 0.3903$, $p = 0.001$; **Table 2**; **Figures 2C,D**). In the NMDS plots, bacterial communities have more distinct clustering by site, and a more linear gradient of community composition dissimilarity by fluid acidity, than eukaryotic communities. Although exact reasons for the color change of pitcher fluids are not fully known, fluid color is also correlated with specific pitcher communities, and explains more of the variation in community composition than site does for both bacteria and eukaryotes (Bacteria $R^2 = 0.16485$, $p = 0.001$; Eukaryotes $R^2 = 0.12727$, $p = 0.001$). In eukaryotes, there is a significant interaction between site and fluid color ($R^2 = 0.09836$, $p = 0.007$), while in bacteria, the interaction shows a trend that is not significant ($R^2 = 0.08946$, $p = 0.076$). Community compositions of bacteria and eukaryotes are both not significantly correlated with pitcher fluid volume, and, similar to species richness (alpha diversity), only eukaryotic community composition (beta diversity) varies significantly with DNA concentration (**Table 2**). The strongest

beta diversity correlation is seen between the weighted UniFrac dissimilarities of bacteria and eukaryotes themselves (mantel test, $r=0.4026$, $p=0.001$; **Table 2**).

Specific Taxa Are Present in *Cephalotus follicularis* Pitcher Fluids of Different Colors

Our taxonomic heat tree analysis found no bacterial taxa that significantly differed by site, however there were bacterial taxa with significant differences based on pitcher fluid color: taxa in the phylum Firmicutes were enriched in green fluids, taxa in the order Bacillales were enriched in cloudy fluids and taxa in the family Simkaniaceae (order Chlamydiales) were enriched in yellow fluids (**Figure 2E**).

Analysis of the eukaryotic data found one significantly differing taxon by site, a fungus classified as *Mortierella* (phylum Mucoromycota) enriched at the TPB site. As with bacteria, eukaryotic taxa—specifically different groups of green algae—significantly differed with the color of the pitcher fluid: Chlorophyta were generally enriched in green fluid, Chlorophyta plus a *Cocomyxa* species in brown fluid and *Cocomyxa dispar* in yellow fluid. In addition to the green algae, Vannellida (an order of freshwater and marine amoebae) was enriched in both green and brown fluids, and Araeolaimida (freshwater nematodes) were enriched in green fluids (**Figure 2F**).

TABLE 2 | Factors affecting bacterial and eukaryotic turnover in community composition.

Factor	Bacteria—16S		Eukaryotes—18S	
PERMANOVA	R^2	p	R^2	p
Fluid color	**0.16485**	**0.001**	**0.12727**	**0.006**
Site	**0.04215**	**0.010**	**0.04755**	**0.001**
Fluid color:site	0.08946	0.076	**0.09836**	**0.007**
Mantel tests	r	p	r	p
pH	**0.2837**	**0.001**	**0.3903**	**0.001**
Volume	0.03139	0.315	−0.07258	0.891
DNA concentration	0.08681	0.125	**0.2853**	**0.001**
18S	**0.4026**	**0.001**		

Bolded values are significant at alpha = 0.05.

Comparing Communities Across Independently Evolved Hosts

In a Global Comparison, *Cephalotus follicularis* Communities Converge With Those of *Nepenthes* and *Sarracenia*

The alpha diversity, or richness as measured by effective number of species, of *C. follicularis* pitcher communities is similar to that of *Nepenthes* pitcher communities in Australia and Southeast Asia and *Sarracenia* pitcher communities in North America. All pitcher communities are relatively low compared with communities of water-filled fallen leaves and other environmental samples collected from surrounding habitats (**Figures 3A,B**). This is true for both bacteria and eukaryotes, with a few slight

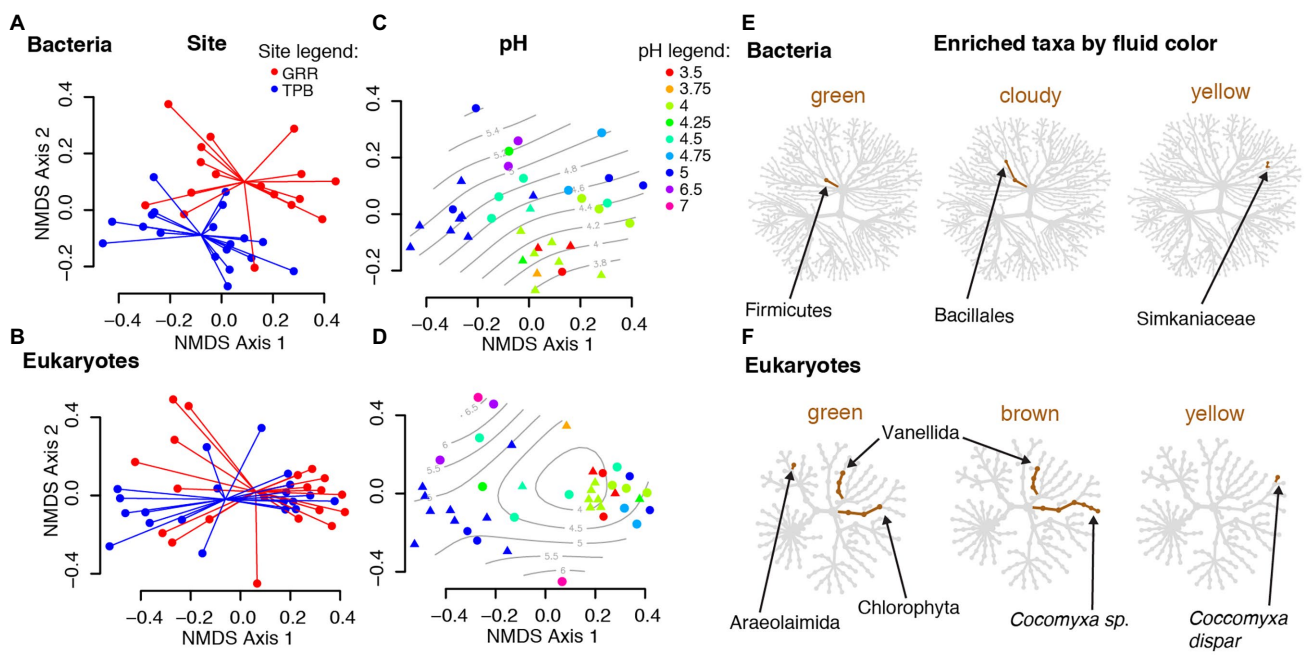


FIGURE 2 | Differences in *Cephalotus follicularis* microbial community composition. **(A–D)** Beta diversity nonmetric multidimensional scaling (NMDS) plots using unweighted UniFrac distances for bacterial and eukaryotic community composition by collecting site and fluid pH. Points in plots **(A)** and **(B)** are colored by site and connected at the mean for each site, while points in plots **(C)** and **(D)** are colored by pH and overlaid with a smooth surface representing pH. **(E,F)** Taxonomic heat trees for bacteria and eukaryotes showing taxa significantly enriched in pitcher fluids of different colors when compared with clear fluid. Taxa shown are enriched by a $\log_2$ ratio median proportion of 3.

differences. *Cephalotus follicularis* bacterial richness is not significantly different when compared with Australian *N. mirabilis* or North American *Sarracenia* bacterial richness but is significantly lower than non-Australian *Nepenthes*, and much lower than all environmental samples (**Figure 3A**). *Cephalotus follicularis* eukaryotic richness is not significantly different when compared with Australian *N. mirabilis* or other *Nepenthes* but is significantly higher than *Sarracenia*, and much lower than all environmental samples (**Figure 3B**).

The bacteria in *C. follicularis* pitchers tend to come from the same phylogenetic clades as those from *Nepenthes* and *Sarracenia* pitchers (**Figure 3C**). Clades with shared taxa among all three pitcher plant genera include: Burkhoderiaceae, Enterobacteriaceae, Rhodanobacteraceae, Acetobacteraceae, Chitinophagaceae, Actinobacteria, and Acidobacteriaceae. In contrast, many fewer eukaryotic taxa are shared with *Nepenthes* and *Sarracenia* with one exception: flies in the order Diptera (flies, mosquitos, midges, etc.) show a clear, strong signal of high relative abundance across all three pitcher plant genera (**Figure 3D**) and another shared group maps to Formicidae (ants).

Similarities and Differences of *Cephalotus follicularis* Communities vs. Australian *Nepenthes mirabilis*

Our rationale for selecting and presenting a comparison sample is largely heuristic. The closest pitcher plants to *C. follicularis* are on the opposite side of the continent. Since detailed and

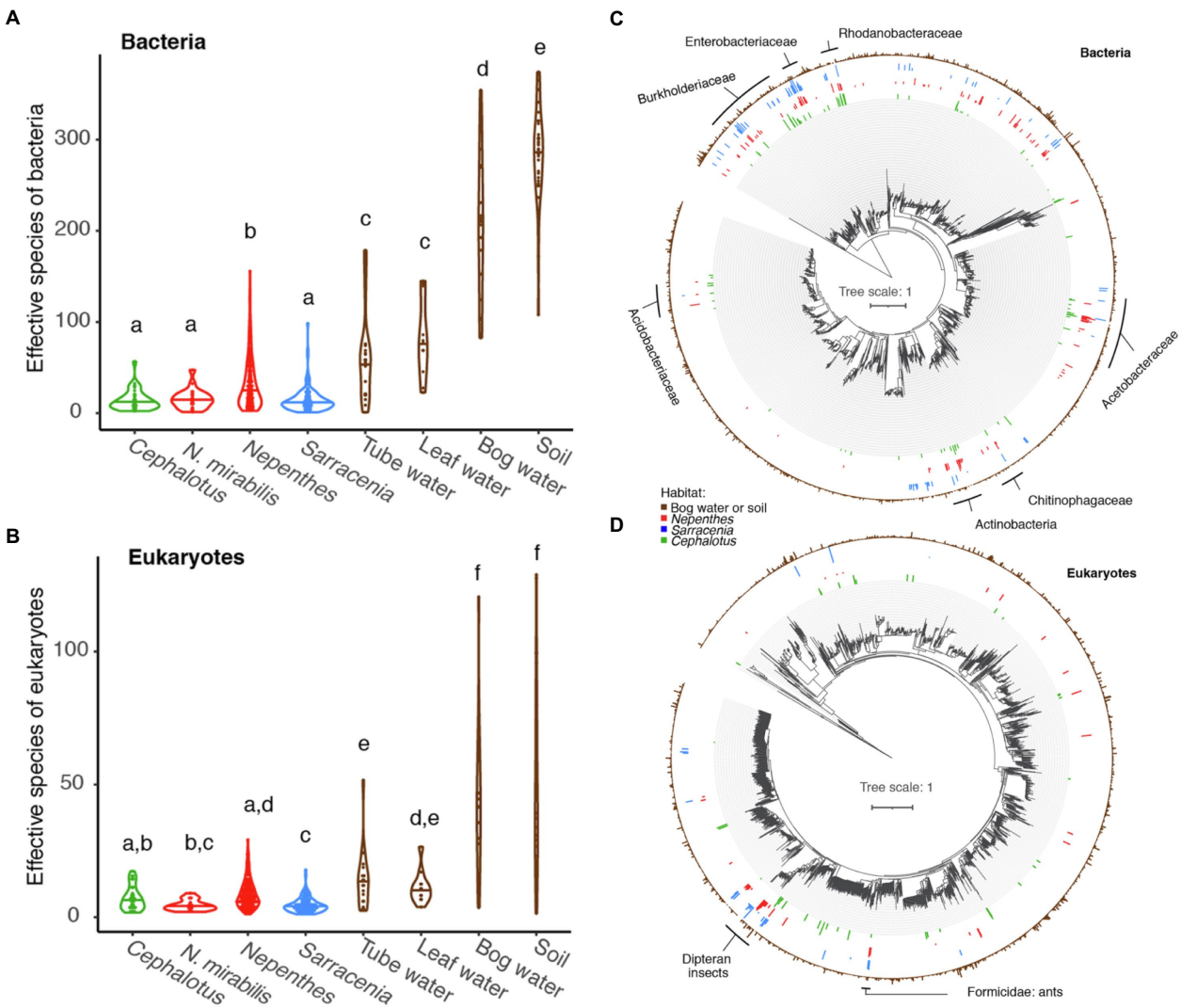


FIGURE 3 | Global comparison of pitcher plant-associated communities. **(A,B)** Richness measured as effective number of species for *Cephalotus follicularis* bacteria **(A)** and eukaryotes **(B)** in comparison with Australian *Nepenthes mirabilis*, other *Nepenthes*, *Sarracenia*, and environmental communities. Different letters above the violin plots represent significant differences as measured by a pairwise Wilcoxon test, with Benjamini and Hochberg adjustment for multiple comparisons. **(C,D)** Phylogenetic trees of bacteria **(C)** and eukaryotes **(D)**, comparing organisms from environmental samples (brown, outer ring) with those found in at least 10% of our *Cephalotus follicularis* (green, innermost ring), *Nepenthes* (red, second innermost ring) and *Sarracenia* (blue, second outermost ring) samples. Names of selected shared clades are listed.

repeated sampling over time was not possible, for this initial, descriptive comparison of the pitcher microcosms from *C. follicularis* with those of other pitcher plant species, we identified a similarly sized, naturally occurring population of *N. mirabilis* pitcher plants in Northeast Australia, and sampled them at a time where pitchers were actively being produced in a manner similar to those observed in the *C. follicularis* population. Although the calendar dates for comparison sampling both fell within 2016, they were separated by 5 months (January for *C. follicularis*, roughly midsummer at a temperate, less humid latitude, and May for *N. mirabilis*, the start of winter months at a humid, tropical latitude on the opposite side of the country, more than 3,400 km away).

The 21 samples collected from *N. mirabilis* at one site near Bramston Beach hosted communities similar to those of *C. follicularis* with regard to bacterial taxa, with Gammaproteobacteria and Alphaproteobacteria at similar relative abundances and the highest representation. However, there were relatively more taxa in the class Bacteroidia in *N. mirabilis*, and fewer in the classes Actinobacteria and Acidobacteriia (**Supplementary Figure 5**). The two Australian pitcher plant species were more different with regard to their associated eukaryotic taxa; in particular *N. mirabilis* had relatively more Metazoa (mostly represented by insects) and fewer Fungi (**Supplementary Figure 5**). In terms of species richness, the Australian *N. mirabilis*-associated had fewer eukaryotic ASVs, on average, than the Southeast Asian *Nepenthes* communities (**Figure 3B**).

DISCUSSION

The pitchers of *C. follicularis* represent a unique habitat that is vulnerable to extinction along with its charismatic host plant (Conran, 2004). Here, for the first time, we used metabarcoding to characterize the full eukaryotic and bacterial communities living within *C. follicularis* pitchers and compared them with pitcher communities from across the globe. We found differences in the factors affecting richness of *C. follicularis* bacterial and eukaryotic communities (**Table 1**), while factors affecting community composition were generally more similar (**Table 2**; **Figure 2**). We also saw convergence in the richness and composition of bacterial taxa of *C. follicularis* when comparing with the convergently evolved *Nepenthes* and *Sarracenia* species from other parts of the world (**Figure 3**). Eukaryotic taxa showed less striking convergence, with strong similarities in richness but primarily only dipteran insects as shared inquilines. In our samples, *C. follicularis* pitchers tended to have fewer arthropod inquilines than have been reported from *Nepenthes* and *Sarracenia* pitchers. The consistently small *C. follicularis* pitcher size may be partly responsible for the differences in eukaryotic taxa, but it also seems possible that their highly isolated and remote location in Western Australia—almost an island within an island—may have limited the kinds and numbers of invertebrate colonists the pitchers have received (Beaver, 1985; Clarke, 1998).

Previous studies have noted differences in factors affecting the diversity of bacterial and eukaryotic communities, as well as correlations between these different taxa (Ragon et al., 2012; Ramirez et al., 2014; Li et al., 2017). For example, a study of *Nepenthes*-associated communities along an elevational gradient in the Philippines found clear differences in the environmental and host factors affecting bacterial and eukaryotic communities, with elevation having a stronger effect on eukaryotes while pH had a stronger effect on bacteria (Gilbert et al., 2020a). Another study across the communities of eight *Nepenthes* species found a strong correlation between bacterial and eukaryotic community composition, and noted tight clustering of protist and bacterial taxa in a co-occurrence network (Bittleston, 2018). Correlations in the diversity of eukaryotic and bacterial communities could be driven by species interactions; for example, a more diverse suite of eukaryotic algae may lead to a more diverse suite of bacteria able to metabolize a variety of algal compounds.

When exploring the richness of *C. follicularis* pitcher-associated communities, we found that bacterial and eukaryotic richness were correlated, but potentially driven by different factors. Bacterial richness differed by the color of the pitcher fluid, while eukaryotic richness differed by DNA concentration. It is unclear why eukaryotic richness would decrease with increasing DNA concentration, since DNA concentration can be a proxy for biomass within the sample. However, it could indicate a bloom of a particular organism: if one species multiplied quickly, it would increase biomass of the focal species while decreasing the overall effective number of species. We also noted a weak correlation between DNA concentration and fluid pH, which either could suggest that lower pH environments house fewer organisms and thus have lower overall biomass, or could suggest that DNA is more difficult to extract from lower pH fluids. In soils, lower pH is associated with decreased microbial growth rates but not biomass (Malik et al., 2018) and can cause adsorption of extracellular DNA to adsorption onto the soil matrix (Young et al., 2014). It is unclear if these processes would apply to pitcher fluid.

In contrast to our richness results, the turnover in community composition of bacterial and eukaryotic communities was generally affected by the same factors, with few exceptions. Both pH and fluid color explained relatively large portions of the variation in bacterial and eukaryotic communities, and the site was also significant although with smaller effect sizes, while neither taxonomic group was significantly explained by pitcher fluid volume. As with richness, only eukaryotic community composition had a significant relationship with DNA concentration. The largest effect size from our mantel tests was for the correlation between bacterial and eukaryotic phylogenetic dissimilarities themselves, suggesting that bacterial and eukaryotic compositions change across samples in similar ways, either due to species interactions or to similar responses to habitat differences.

Our analysis using taxonomic heat trees revealed enrichment of certain taxa in different colors of pitcher fluid and one between sites. The taxa in differently colored fluids likely either caused the color changes or were affected by a shift in chemistry

or nutrients related to a color change. Green, brown, and yellow fluids were enriched with algae in the Chlorophyta group, while cloudy fluids were enriched with bacteria in the order Bacillales. The algae likely contributed to the green, brown, and yellow coloration, while the bacteria may have contributed to the cloudiness. Only one taxon differed significantly by site, the fungal genus *Mortierella*. Species of *Mortierella* are generally decomposers, and some have the ability to degrade the chitin in insect exoskeletons (Okafor, 1966; Edgington et al., 2014) or to enhance plant growth when part of the rhizosphere (Li et al., 2020). Potentially the *Mortierella* that were enriched at the TPB site were involved in breaking down insect prey captured by *C. follicularis* pitchers, and a possible symbiosis between the fungus and pitcher plant would be interesting to explore further.

Pitcher communities change over successional time, from pitcher opening to senescence. This successional change has been well-documented only in *Sarracenia purpurea* (e.g., Gray, 2012; Miller and terHorst, 2012; Luciano and Newell, 2017; Boynton et al., 2019), but occurs in *N. mirabilis* (Benson, Bittleston, and Pierce, unpublished results), and likely also in *C. follicularis*. Limitations of our current study prevented us from sampling the same pitchers at different successional stages; however, we aimed to select healthy pitchers of different ages to get a full representation of the existing diversity. Future studies in *C. follicularis* could sample across different seasons and successional stages.

All three genera of pitcher plants host specialized insects (Adlassnig et al., 2011), and host specificity of inquilines could lead to co-evolution of the plant and certain associated organisms over time. For example, in the *Sarracenia alata* system, pitcher communities—and arthropods in particular—have co-diversified with their host plants across a biogeographic barrier (Koopman and Carstens, 2011; Satler and Carstens, 2016, 2019). The main eukaryotic taxonomic group showing convergence in our study is Diptera, or flies. These highly-represented ASVs likely come from DNA of specialized inquilines and not from prey, since previous pitcher plant studies have demonstrated corroboration between amplicon sequencing-derived taxa and barcoded DNA extracted directly from inquilines (Bittleston et al., 2016a). The most abundant dipteran ASVs from *C. follicularis* map most closely to *Drosophila*; however, it is likely that they are actually the inquiline fly *Badisis ambulans*, whose larvae we saw swimming in our pitcher samples. *Badisis ambulans* is not represented in the NCBI database, in fact, currently only one taxon in the whole Micropezidae family is represented by a section of the 18S rRNA gene sequence. The dipteran ASVs from *Nepenthes* and *Sarracenia* are generally assigned to the mosquitoes, midges, and other flies that are known inquilines of these plants (Bittleston et al., 2016a, 2018). Overall, we saw lower eukaryotic species richness in the Australian *N. mirabilis* pitchers when compared with other *Nepenthes*, potentially because Australian *N. mirabilis* is geographically isolated on the edge of the *Nepenthes* distribution without other co-occurring *Nepenthes* species (Beaver, 1985; Clarke, 1998). Similarly, *C. follicularis* had fewer insect ASVs than *N. mirabilis* and the other pitcher plant species, perhaps due to a combination of geographic isolation and small size. The other eukaryotic group showing convergence across pitcher plant genera was Formicidae, or ants, and these ASVs likely come from the primary prey of pitcher plants, as all three groups commonly capture and digest ants (Ellison and Gotelli, 2009).

Beyond arthropods, many bacterial groups were present across the three genera of pitcher plants. It is possible that taxa in these shared groups (such as the Acetobacteraceae, Rhodanobacteraceae, Chitinophagaceae, Burkholderiaceae, among others) could be specialists in the pitcher habitats; however, host specificity in bacteria is rare and difficult to test. It is more likely that they are present due to environmental filtering as a result of host convergence with regard to the pitcher habitat. Pitchers are not just passive receptacles, they affect the habitat inside *via* excretion of digestive enzymes, oxygenation, and sometimes active pH adjustment (Bradshaw and Creelman, 1984; An et al., 2001; Adlassnig et al., 2011; Sirota et al., 2013; Ravee et al., 2018; Gilbert et al., 2020b). A recent experiment found *S. purpurea* regulates the structure of its inquiline food web, and direct contact with the plant tissue was important (Ellison et al., 2021). Similarities in the chemistry and conditions of other, non-host associated habitats can also lead to convergence in bacterial communities; for example, similarities in pH, salinity and other soil characteristics drive similarity in soil microbial communities (Ramirez et al., 2014; Rath et al., 2019). However, non-host associated systems are less likely to represent convergent interactions (Bittleston et al., 2016b), where a specific functional interaction—such as bacteria assisting pitcher plants in the degradation of insect prey—emerges independently in multiple instances. Due to the combination of convergent evolution at the host level and likely chemical similarities at the pitcher habitat level, we consider this convergence as both evolutionary and ecological.

The similarities in richness and bacterial taxa in *C. follicularis*, *Nepenthes* and *Sarracenia* pitcher communities in very different parts of the world—Australia, Southeast Asia and North America—suggests that pitchers in general cultivate a particular habitat that selects for a similar microbiome, regardless of location or evolutionary history (Bittleston et al., 2018). This is not to say that communities hosted by different genera of pitcher plants do not show variation. In fact, each particular plant can host unique combinations of taxa that, at the fine scale, vary between genus, species, site, and even within a population due to unique characteristics like pitcher fluid acidity. However, at a broad scale, pitcher microbial communities on opposite sides of the world are more similar to each other than to the water-filled leaf, bog water, or soil communities in the surrounding habitats.

DATA AVAILABILITY STATEMENT

The names of the repositories and accession numbers are: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA810039; https://doi.org/10.7910/DVN/WG1TI6.

AUTHOR CONTRIBUTIONS

LB and NP conceived of the study. LB and EB collected and processed samples. LB, EB, and JB analyzed the data. All authors helped to write and edit the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The collection of *Cephalotus follicularis* samples was funded by a Putnam Expedition Grant from the Harvard Museum of Comparative Zoology (MCZ) to LB. The collection of *Nepenthes mirabilis* samples was funded by a Bolitho Scholarship from the Harvard University Committee on Australian Studies. Partial funding for the DNA extraction and amplification and initial analyses was provided by The Museum of Comparative Zoology, Harvard University Herbaria, and Arnold Arboretum Grant-in-aid of Undergraduate Research and Harvard College Research Program Undergraduate Summer Funding to EB. LB and JB are currently funded by an NSF award 2025250, and LB is also supported by the NSF Idaho EPSCoR Program and award number OIA-1757324. Published by a grant from the Wetmore Colles Fund.

ACKNOWLEDGMENTS

We appreciate the helpful comments and suggestions from two reviewers. Samples from *Cephalotus follicularis* were collected and exported under permits from the Government of Western Australia Department of Parks and Wildlife, numbers SF010649 and OS002547. Samples from Nepenthes mirabilis were exported under permit number PWS2016-AU-000982. We are grateful to Stephen Hopper and Samuel Lymbery for assistance with locating and visiting *Cephalotus follicularis* populations, and to Charles Clarke for assistance with *Nepenthes mirabilis* sampling. We appreciate the help of Dan Sternof Beyer with *Cephalotus follicularis* collections.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.887635/full#supplementary-material

REFERENCES

Adlassnig, W., Peroutka, M., and Lendl, T. (2011). Traps of carnivorous pitcher plants as a habitat: composition of the fluid, biodiversity and mutualistic activities. *Ann. Bot.* 107, 181–194. doi: 10.1093/aob/mcq238

Albert, V. A., Williams, S. E., and Chase, M. W. (1992). Carnivorous plants: phylogeny and structural evolution. *Science* 257, 1491–1495. doi: 10.1126/science.1523408

Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W., and Huse, S. M. (2009). A method for studying Protistan diversity using massively parallel sequencing of V9 Hypervariable regions of small-subunit ribosomal RNA genes. *PLoS One* 4:e6372. doi: 10.1371/journal.pone.0006372

An, C.-I., Fukusaki, E.-I., and Kobayashi, A. (2001). Plasma-membrane H + -ATPases are expressed in pitchers of the carnivorous plant *Nepenthes alata* Blanco. *Planta* 212, 547–555. doi: 10.1007/s004250000455

Beaver, R. A. (1985). Geographical variation in food web structure in *Nepenthes* pitcher plants. *Ecol. Entomol.* 10, 241–248. doi: 10.1111/j.1365-2311.1985.tb00720.x

Bell, T. H., Hockett, K. L., Alcalá-Briseño, R. I., Barbercheck, M., Beattie, G. A., Bruns, M. A., et al. (2019). Manipulating wild and tamed phytobiomes: challenges and opportunities. *Phytobiomes J.* 3, 3–21. doi: 10.1094/PBIOMES-01-19-0006-W

Bittleston, L. S. (2018). "Commensals of Nepenthes Pitchers," in *Carnivorous Plants: Physiology, Ecology, and Evolution*. eds. A. M. Ellison and L. Adamec (Oxford, England: Oxford University Press).

Bittleston, L. S., Baker, C. C. M., Strominger, L. B., Pringle, A., and Pierce, N. E. (2016a). Metabarcoding as a tool for investigating arthropod diversity in *Nepenthes* pitcher plants. *Austral Ecol.* 41, 120–132. doi: 10.1111/aec.12271

Bittleston, L. S., Pierce, N. E., Ellison, A. M., and Pringle, A. (2016b). Convergence in multispecies interactions. *Trends Ecol. Evol.* 31, 269–280. doi: 10.1016/j.tree.2016.01.006

Bittleston, L. S., Wolock, C. J., Yahya, B. E., Chan, X. Y., Chan, K. G., Pierce, N. E., et al. (2018). Convergence between the microcosms of southeast Asian and north American pitcher plants. *eLife* 7:e36741. doi: 10.7554/eLife.36741

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., et al. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 6:90. doi: 10.1186/s40168-018-0470-z

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9

Boynton, P. J., Peterson, C. N., and Pringle, A. (2019). Superior dispersal ability can Lead to persistent ecological dominance throughout succession. *Appl. Environ. Microbiol.* 85, e02421–e024218. doi: 10.1128/AEM.02421-18

Bradshaw, W. E., and Creelman, R. A. (1984). Mutualism between the carnivorous purple pitcher plant and its inhabitants. *Am. Midl. Nat.* 112, 294–304. doi: 10.2307/2425436

Butler, J. L., Gotelli, N. J., and Ellison, A. M. (2008). Linking the Brown and Green: nutrient transformation and fate in the *Sarracenia* microecosystem. *Ecology* 89, 898–904. doi: 10.1890/07-1314.1

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/ismej.2012.8

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., et al. (2014). Rarefaction and extrapolation with hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* 84, 45–67. doi: 10.1890/13-0133.1

Clarke, S. A. (1988). Seasonal growth and mortality of the pitchers of the Albany pitcher plant, *Cephalotus follicularis* Labill. *Aust. J. Bot.* 36, 643–653. doi: 10.1071/BT9880643

Clarke, C. (1998). A re-examination of geographical variation in *Nepenthes* food webs. *Ecography* 21, 430–436. doi: 10.1111/j.1600-0587.1998.tb00408.x

Conran, J. G. (2004). "Cephalotaceae," in *Flowering Plants. Dicotyledons: Celastrales, Oxalidales, Rosales, Cornales, Ericales*. ed. K. Kubitzki (Berlin, Heidelberg: Springer), 65–68.

De Bie, T., De Meester, L., Brendonck, L., Martens, K., Goddeeris, B., Ercken, D., et al. (2012). Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. *Ecol. Lett.* 15, 740–747. doi: 10.1111/j.1461-0248.2012.01794.x

Edgington, S., Thompson, E., Moore, D., Hughes, K. A., and Bridge, P. (2014). Investigating the insecticidal potential of Geomyces (Myxotrichaceae: Helotiales)

and Mortierella (Mortierellacea: Mortierellales) isolated from Antarctica. *Springerplus* 3:289. doi: 10.1186/2193-1801-3-289

Ellison, A. M., and Gotelli, N. J. (2009). Energetics and the evolution of carnivorous plants—Darwin's 'most wonderful plants in the world'. *J. Exp. Bot.* 60, 19–42. doi: 10.1093/jxb/ern179

Ellison, A. M., and Gotelli, N. J. (2021). *Scaling in Ecology with a Model System*. Princeton, NJ: Princeton University Press.

Ellison, A. M., Gotelli, N. J., Błędzki, L. A., and Butler, J. L. (2021). Regulation by the pitcher plant *Sarracenia purpurea* of the structure of its inquiline food web. *Am. Midl. Nat.* 186, 1–15. doi: 10.1674/0003-0031-186.1.1

Foster, Z. S. L., Sharpton, T. J., and Grünwald, N. J. (2017). Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLoS Comput. Biol.* 13:e1005404. doi: 10.1371/journal.pcbi.1005404

Fukushima, K., Fang, X., Alvarez-Ponce, D., Cai, H., Carretero-Paulet, L., Chen, C., et al. (2017). Genome of the pitcher plant *Cephalotus* reveals genetic changes associated with carnivory. *Nat. Ecol. Evol.* 1, 1–9. doi: 10.1038/s41559-016-0059

Gilbert, K. J., Bittleston, L. S., Naive, M. A. K., Kiszewski, A. E., Buenavente, P. A. C., Lohman, D. J., et al. (2020a). Investigation of an elevational gradient reveals strong differences between bacterial and eukaryotic communities Coinhabiting Nepenthes Phytotelmata. *Microb. Ecol.* 80, 334–349. doi: 10.1007/s00248-020-01503-y

Gilbert, K. J., Bittleston, L. S., Tong, W., and Pierce, N. E. (2020b). Tropical pitcher plants (*Nepenthes*) act as ecological filters by altering properties of their fluid microenvironments. *Sci. Rep.* 10:4431. doi: 10.1038/s41598-020-61193-x

Gray, S. M. (2012). Succession in the aquatic *Sarracenia purpurea* community: deterministic or driven by contingency? *Aquat. Ecol.* 46, 487–499. doi: 10.1007/s10452-012-9417-9

Gray, S. M., Akob, D. M., Green, S. J., and Kostka, J. E. (2012). The bacterial composition within the *Sarracenia purpurea* model system: local scale differences and the relationship with the other members of the food web. *PLoS One* 7:e50969. doi: 10.1371/journal.pone.0050969

Grothjan, J. J., and Young, E. B. (2019). Diverse microbial communities hosted by the model carnivorous pitcher plant *Sarracenia purpurea*: analysis of both bacterial and eukaryotic composition across distinct host plant populations. *PeerJ* 7:e6392. doi: 10.7717/peerj.6392

Janssen, S., McDonald, D., Gonzalez, A., Navas-Molina, J. A., Jiang, L., Xu, Z. Z., et al. (2018). Phylogenetic placement of exact amplicon sequences improves associations with clinical information. *mSystems* 3, e00021–e000218. doi: 10.1128/mSystems.00021-18

Jost, L. (2007). Partitioning diversity into independent alpha and Beta components. *Ecology* 88, 2427–2439. doi: 10.1890/06-1736.1

Just, M. P., Merritt, D. J., Turner, S. R., Conran, J. G., and Cross, A. T. (2019). Seed germination biology of the Albany pitcher plant, *Cephalotus follicularis*. *Aust. J. Bot.* 67, 480–489. doi: 10.1071/BT19053

Kanokratana, P., Mhuanthong, W., Laothanachareon, T., Tangphatsornruang, S., Eurwilaichitr, L., Kruetreepradit, T., et al. (2016). Comparative study of bacterial communities in *Nepenthes* pitchers and their correlation to species and fluid acidity. *Microb. Ecol.* 72, 381–393. doi: 10.1007/s00248-016-0798-5

Kitching, R. L. (2000). *Food Webs and Container Habitats: The Natural History and Ecology of Phytotelmata*. Cambridge, England: Cambridge University Press.

Koopman, M. M., and Carstens, B. C. (2011). The microbial phyllogeography of the carnivorous plant *Sarracenia alata*. *Microb. Ecol.* 61, 750–758. doi: 10.1007/s00248-011-9832-9

Leveau, J. H. (2019). A brief from the leaf: latest research to inform our understanding of the phyllosphere microbiome. *Curr. Opin. Microbiol.* 49, 41–49. doi: 10.1016/j.mib.2019.10.002

Li, Y., Adams, J., Shi, Y., Wang, H., He, J.-S., and Chu, H. (2017). Distinct soil microbial communities in habitats of differing soil water balance on the Tibetan plateau. *Sci. Rep.* 7:46407. doi: 10.1038/srep46407

Li, F., Zhang, S., Wang, Y., Li, Y., Li, P., Chen, L., et al. (2020). Rare fungus, Mortierella capitata, promotes crop growth by stimulating primary metabolisms related genes and reshaping rhizosphere bacterial community. *Soil Biol. Biochem.* 151:108017. doi: 10.1016/j.soilbio.2020.108017

Luciano, C. S., and Newell, S. J. (2017). Effects of prey, pitcher age, and microbes on acid phosphatase activity in fluid from pitchers of *Sarracenia purpurea* (Sarraceniaceae). *PLoS One* 12:e0181252. doi: 10.1371/journal.pone.0181252

Lymbery, S. J., Didham, R. K., Hopper, S. D., and Simmons, L. W. (2016). Mutualists or parasites? Context-dependent influence of symbiotic fly larvae on carnivorous investment in the Albany pitcher plant. *R. Soc. Open Sci.* 3:160690. doi: 10.1098/rsos.160690

Malik, A. A., Puissant, J., Buckeridge, K. M., Goodall, T., Jehmlich, N., Chowdhury, S., et al. (2018). Land use driven change in soil pH affects microbial carbon cycling processes. *Nat. Commun.* 9:3591. doi: 10.1038/s41467-018-05980-1

McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., et al. (2012). An improved greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6, 610–618. doi: 10.1038/ismej.2011.139

Miller, T. E., Bradshaw, W. E., and Holzapfel, C. M. (2018). "Pitcher-Plant Communities as Model Systems for Addressing Fundamental Questions in Ecology and Evolution," in *Carnivorous Plants: Physiology, Ecology, and Evolution*. eds. A. M. Ellison and L. Adamec (Oxford, England: Oxford University Press).

Miller, T. E., and terHorst, C. P. (2012). Testing successional hypotheses of stability, heterogeneity, and diversity in pitcher-plant inquiline communities. *Oecologia* 170, 243–251. doi: 10.1007/s00442-012-2292-1

Okafor, N. (1966). The ecology of micro-organisms on, and the decomposition of, insect wings in the soil. *Plant Soil* 25, 211–237. doi: 10.1007/BF01347820

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2020). Vegan: Community Ecology Package. R package version 2.5-7. Available at: https://CRAN.R-project.org/package=vegan

Parkes, D. M., and Hallam, N. D. (1984). Adaptation for Carnivory in the west Australian pitcher plant *Cephalotus follicularis* Labill. *Aust. J. Bot.* 32, 595–604. doi: 10.1071/BT9840595

Pavlovic, A. (2011). Photosynthetic characterization of Australian pitcher plant *Cephalotus follicularis*. *Photosynthetica* 49, 253–258. doi: 10.1007/s11099-011-0032-0

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

Peterson, C. N., Day, S., Wolfe, B. E., Ellison, A. M., Kolter, R., and Pringle, A. (2008). A keystone predator controls bacterial diversity in the pitcher-plant (*Sarracenia purpurea*) microecosystem. *Environ. Microbiol.* 10, 2257–2266. doi: 10.1111/j.1462-2920.2008.01648.x

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ragon, M., Fontaine, M. C., Moreira, D., and López-García, P. (2012). Different biogeographic patterns of prokaryotes and microbial eukaryotes in epilithic biofilms. *Mol. Ecol.* 21, 3852–3868. doi: 10.1111/j.1365-294X.2012.05659.x

Ramirez, K. S., Leff, J. W., Barberán, A., Bates, S. T., Betley, J., Crowther, T. W., et al. (2014). Biogeographic patterns in below-ground diversity in new York City's Central Park are similar to those observed globally. *Proc. R. Soc. B Biol. Sci.* 281:20141988. doi: 10.1098/rspb.2014.1988

Rath, K. M., Fierer, N., Murphy, D. V., and Rousk, J. (2019). Linking bacterial community composition to soil salinity along environmental gradients. *ISME J.* 13, 836–846. doi: 10.1038/s41396-018-0313-8

Ravee, R., Salleh, F. I. M., and Goh, H.-H. (2018). Discovery of digestive enzymes in carnivorous plants with focus on proteases. *PeerJ* 6:e4914. doi: 10.7717/peerj.4914

Sambrook, J., Russell, D. W., and Laboratory, C. S. H.. (2001). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Cold Spring Harbor, N.Y.

Satler, J. D., and Carstens, B. C. (2016). Phylogeographic concordance factors quantify phylogeographic congruence among co-distributed species in the *Sarracenia alata* pitcher plant system. *Evolution* 70, 1105–1119. doi: 10.1111/evo.12924

Satler, J. D., and Carstens, B. C. (2019). The *Sarracenia alata* pitcher plant system and obligate arthropod inquilines should be considered an evolutionary community. *J. Biogeogr.* 46, 485–496. doi: 10.1111/jbi.13498

Schulze, W., Schulze, E. D., Pate, J. S., and Gillison, A. N. (1997). The nitrogen supply from soils and insects during growth of the pitcher plants *Nepenthes mirabilis*, *Cephalotus follicularis* and *Darlingtonia californica*. *Oecologia* 112, 464–471. doi: 10.1007/s004420050333

Sirota, J., Baiser, B., Gotelli, N. J., and Ellison, A. M. (2013). Organic-matter loading determines regime shifts and alternative states in an aquatic ecosystem. *Proc. Natl. Acad. Sci.* 110, 7742–7747. doi: 10.1073/pnas.1221037110

Srivastava, D. S., Kolasa, J., Bengtsson, J., Gonzalez, A., Lawler, S. P., Miller, T. E., et al. (2004). Are natural microcosms useful model systems for ecology? *Trends Ecol. Evol.* 19, 379–384. doi: 10.1016/j.tree.2004.04.010

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H.-W., et al. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* 19, 21–31. doi: 10.1111/j.1365-294X.2009.04480.x

Yeates, D. (1992). Immature stages of the apterous fly Badisis ambulans McAlpine (Diptera: Micropezidae). *J. Nat. Hist.* 26, 417–424. doi: 10.1080/00222939200770241

Young, J. M., Rawlence, N. J., Weyrich, L. S., and Cooper, A. (2014). Limitations and recommendations for successful DNA extraction from forensic soil samples: A review. *Sci. Justice* 54, 238–244. doi: 10.1016/j.scijus.2014.02.006

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

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 02 February 2023
DOI 10.3389/fpls.2023.1111968

OPEN ACCESS

EDITED BY
Magdy S. Alabady,
University of Georgia, United States

REVIEWED BY
Diaga Diouf,
Cheikh Anta Diop University, Senegal
Muhammad Aamir Manzoor,
Anhui Agricultural University, China

*CORRESPONDENCE
Yong-Min Kim
ymkim@kribb.re.kr

†These authors have contributed equally to this work

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

RECEIVED 30 November 2022
ACCEPTED 19 January 2023
PUBLISHED 02 February 2023

CITATION
Koo H, Shin A-Y, Hong S and Kim Y-M (2023) The complete chloroplast genome of *Hibiscus syriacus* using long-read sequencing: Comparative analysis to examine the evolution of the tribe Hibisceae. *Front. Plant Sci.* 14:1111968. doi: 10.3389/fpls.2023.1111968

The complete chloroplast genome of *Hibiscus syriacus* using long-read sequencing: Comparative analysis to examine the evolution of the tribe Hibisceae

Hyunjin Koo[1†], Ah-Young Shin[1,2†], Seongmin Hong[1] and Yong-Min Kim[1,2,3*]

[1]Plant Systems Engineering Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, Republic of Korea, [2]Department of Bioinformatics, Korea Research Institute of Bioscience and Biotechnology (KRIBB) School of Bioscience, Korea University of Science and Technology (UST), Daejeon, Republic of Korea, [3]Digital Biotech Innovation Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, Republic of Korea

Hibiscus syriacus, a member of the tribe Hibisceae, is considered an important ornamental and medicinal plant in east Asian countries. Here, we sequenced and assembled the complete chloroplast genome of *H. syriacus* var. *Baekdansim* using the PacBio long-read sequencing platform. A quadripartite structure with 161,026 base pairs was obtained, consisting of a pair of inverted repeats (IRA and IRB) with 25,745 base pairs, separated by a large single-copy region of 89,705 base pairs and a short single-copy region of 19,831 base pairs. This chloroplast genome had 79 protein-coding genes, 30 transfer RNA genes, 4 ribosomal RNA genes, and 109 simple sequence repeat regions. Among them, *ndhD* and *rpoC1*, containing traces of RNA-editing events associated with adaptive evolution, were identified by analysis of putative RNA-editing sites. Codon usage analysis revealed a preference for A/U-terminated codons. Furthermore, the codon usage pattern had a clustering tendency similar to that of the phylogenetic analysis of the tribe Hibisceae. This study provides clues for understanding the relationships and refining the taxonomy of the tribe Hibisceae.

KEYWORDS

long-read sequencing platform, complete chloroplast genome assembly, *Hibiscus syriacus*, comparative analysis, Hibisceae

1 Introduction

Hibiscus is one of the most diverse and widespread genera in the Malvaceae tribe Hibisceae (Rizk and Soliman, 2014). The members of the tribe Hibisceae are widely distributed from tropical to temperate regions worldwide (Akpan, 2000). Several species of the tribe Hibisceae are regarded as valuable research crops since they are economically

important for food and medicines and can be utilized as biofuels due to their high biomass content and photosynthetic efficiency (Akpan et al., 2000; Saba et al., 2015; Wu et al., 2017; Cheng et al., 2020). *Hibiscus syriacus*, a member of the tribe Hibisceae, is a flowering shrub that originated in the Korean peninsula and southern China. It is one of the most widely planted ornamental species in temperate zones and is a fast-growing species with attractive white, red, pink, purple, and lavender flowers (Paoletti et al., 2009; Kim et al., 2017). Along with ornamental value, the dried flowers and root bark of *H. syriacus* have been used as a traditional remedy in Oriental countries (Yoo et al., 1998). Particularly, three naphthalene chemicals (syriacusins A–C) and novel pentacyclic triterpene esters identified from the plant's root bark have been used as anthelmintic, antipyretic, and antifungal agents (Yoo et al., 1998; Yun et al., 1999).

Chloroplasts are multifunctional organelles that carry their own genetic sources responsible for photosynthesis, various types of metabolism, and carbon fixation (Li et al., 2019). Chloroplast genomes typically have a quadripartite structure with two copies of inverted repeat (IR) regions separating the large and small single-copy (LSC and SSC, respectively) regions. Most chloroplast genomes range from 120 to 160 kb. Generally, chloroplast genomes of angiosperms contain approximately 120 genes including protein-coding genes, transfer RNA (tRNA), and ribosomal RNA (rRNA) (Daniell et al., 2016). Several mutational events, including mutations, duplications, rearrangements, and gene deletions, occur in chloroplast genomes (Lee et al., 2007). Nevertheless, compared to the nuclear or mitochondrial genome, the chloroplast genome is structurally conserved; hence, it is commonly employed to elucidate the genome evolution and phylogenetic relationships of land plants (Huang et al., 2014; Wu et al., 2017). With the emergence of high-throughput sequencing, the chloroplast genome assemblies of various species of the tribe Hibisceae have been completed (Cheng et al., 2020; Li et al., 2020; Mehmood et al., 2020). Although the phylogenetic relationship among several species of the family Malvaceae was estimated in previous studies, it is insufficient that a comprehensive comparative analysis of the chloroplast genomes in the tribe Hibisceae.

Here, we report the whole chloroplast genome of *H. syriacus* var. *Baekdansim* (hereafter referred to as Baekdansim) using PacBio long-read sequencing data for the first time. Further comparative genome analyses were carried out using the complete chloroplast genomes of other species belonging to the tribe Hibisceae that were obtained from the NCBI database. The findings of this study will be helpful for the development of genetic markers to resolve taxonomic discrepancies or to infer phylogenetic and evolutionary relationships within the tribe Hibisceae.

2 Materials and methods

2.1 Plant material and chloroplast DNA extraction

To isolate high-purity Baekdansim chloroplast DNA from cells, chloroplasts and mitochondria were the first separated from other components, especially nucleus DNA. This step was achieved by homogenizing 5–10 g (fresh weight) of young leaf tissue followed by a nuclei isolation step according to previous protocols (Zerpa-Catanho et al., 2021). For chloroplast DNA extraction, nuclei removed extract was transferred to 10 mL lysis buffer (50 mM Tris-HCl pH 7.5, 1.4 M NaCl, 20 mM EDTA, pH 8.0, 0.5% SDS) and incubated for 1 h in a 65°C water bath with gentle inversion every 20 min. The supernatant was separated by centrifugation at 3000 rpm for 10 min and transferred to a new tube. RNase A (10 mg/mL) was then added, and the mixture was incubated for 30 min at room temperature. Next, an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the supernatant, and the sample was mixed by gentle inversion for 5 min before centrifugation at 3000 rpm for 10 min. After the aqueous phase was transferred to a new tube, an equal volume of chloroform was added and mixed. The mixture was separated by centrifugation at 3000 rpm for 10 min. The upper, DNA-containing phase was transferred to a new tube, and an equal volume of isopropanol was added to precipitate the DNA, followed by centrifugation at 3000 rpm for 5 min. The DNA pellet was washed with 70% ethanol and resuspended in 100 µL of TE buffer (pH 8.0). Solubilized DNA was stored at 4°C until library preparation.

2.2 Library construction and sequencing

Purified genomic DNA (gDNA) was used for library construction with the SMRTbell Express Template Prep Kit (Pacific Biosciences, Cat. No. 101-357-000). In brief, gDNA was mechanically sheared to an average size of 20 kb using a Covaris g-TUBE device (Part No. 520079). In total, 5 µg of sheared gDNA was damage-repaired and end-repaired using polishing enzymes. Blunt-end adapter ligation was used to create the SMRTbell template. Adapter dimers and contaminants were removed using the AMPure XP bead purification system (Beckman Coulter, Cat. No. A63882). A BluePippin size selection system (Sage Science, Cat. No. BLU0001) was used to size select the SMRTbell template and enrich for fragments > 15 kb. Sequencing primer v4 was annealed to the SMRTbell template, and a DNA polymerase/template complex was created using the Sequel Binding Kit 2.1 (Pacific Biosciences, Cat. No. 101-365-900). An additional AMPure XP purification step was performed to remove excess primer and polymerase prior to sequencing. The library was sequenced on a Sequel instrument using SMRT Cell 1M v2 (Pacific Biosciences), taking one movie of 10 hours per cell with the Sequel Sequencing Kit 2.0 (Pacific Biosciences).

2.3 Genome assembly and annotation

Reads from chloroplasts were extracted by alignment of all reads onto the five chloroplast complete genome assemblies of *Hibiscus* species (*H. syriacus*: NC_026909.1, *H. cannabinus*: NC_045873.1, *H. trionum*: NC_060636.1, *H. rosa-sinensis*: NC_042239.1, and *H. taiwanensis*: NC_045873.1) deposited in the NCBI database (https://www.ncbi.nlm.nih.gov/nucleotide/). Each chloroplast genome was duplicated and concatenated to facilitate the alignment of reads on

the circularized region as suggested by Wang et al. (Wang et al., 2018). Long reads were mapped to chloroplast genomes using minimap2 version 2.24 (Li, 2018). Then, the short reads were mapped using bwa version 0.7.17 (Li and Durbin, 2009). A data set of extracted chloroplast reads was constructed using Unicycler v0.5.0 with the hybrid assembly strategy (Wick et al., 2017). Genome annotation was performed on the GeSeq platform using the complete chloroplast genome (Tillich et al., 2017). Coding sequence (CDS) and rRNAs were predicted by BLAT (Kent, 2002) and HMMER (Finn et al., 2011) search. In addition, the tRNAs were further verified by tRNAscan-SE v2.0.7 (Lowe and Chan, 2016) and ARAGORN v1.2.38 (Laslett and Canback, 2004) with default option. Then, a circular chloroplast map was constructed according to the genome annotation using the online program OGDRAW v1.3.1 (Greiner et al., 2019). The final Baekdansim plastome was deposited in GenBank with accession number OP874596.1. The corresponding circular genome map is shown in Figure 1.

2.4 Repeat sequence identification

Two programs were used to detect repeat motifs. Regarding microsatellites, MISA software (Beier et al., 2017) was used to examine the locations and motifs of simple sequence repeats (SSRs). SSRs were detected using thresholds of 10, 5, 4, 3, 3 and 3 repeat units for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides, respectively. To identify long repeat motifs, forward, reverse, complementary, and palindromic sequences were determined using REPuter v1.0, with a minimum repetition size of 30 bp and 90% identity (Kurtz et al., 2001).

2.5 Genetic divergence and chloroplast genome comparison

The nucleotide divergence (π) among the 13 species of the tribe Hibisceae was determined using DnaSP v6.0 based on sliding window analysis (Supplementary Table 1) (Rozas et al., 2017). The window length was set to 600 bp with a 100-bp step size. Comprehensive alignments of the complete chloroplast genomes of the tribe Hibisceae were examined using the mVISTA program (Frazer et al., 2004) to reveal interspecific variations. Furthermore, expansion and contraction between the LSC/IRB/SSC/IRA regions at junction sites were identified using IRscope (Amiryousefi et al., 2018). Genes in the chloroplast genomes of 13 species were investigated to determine the presence of introns. Alterations of genes containing intron regions were identified using in-house Python code.

2.6 Codon usage and RNA-editing sites

Relative synonymous codon usage (RSCU) analysis of coding sequences was conducted using MEGA v11.0 (Kumar et al., 1994), and an RSCU value greater than one was regarded as a high codon frequency. The putative RNA-editing sites of the start and stop codons of the coding sequences from species of the tribe Hibisceae were predicted using in-house Python code.

2.7 Phylogenetic analysis

The complete chloroplast genome sequence of Baekdansim, together with those of the other 12 species of the tribe Hibisceae

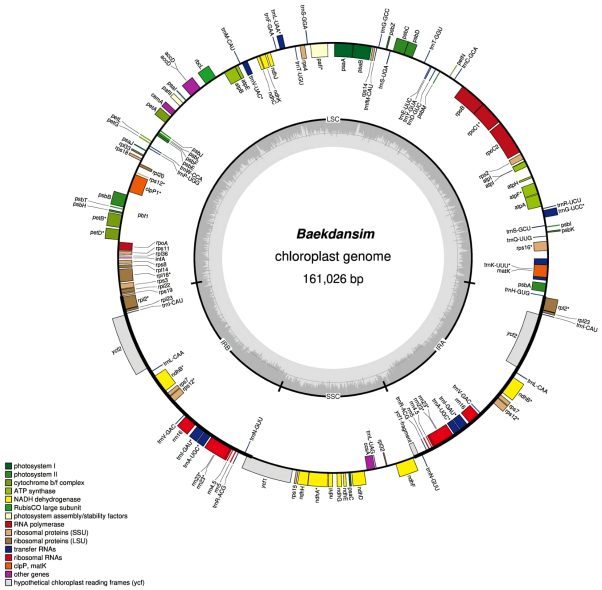


FIGURE 1
Circular chloroplast genome map of *H. syriacus* var. *Baekdansim*. The inner grey circle indicates the proportion of GC in each region. The genes illustrated in the inner circle are transcribed clockwise. Genes corresponding to distinct functional groups are denoted using distinct colors.

available in the NCBI database, were used for comparative and phylogenetic analyses (Supplementary Table 1). The chloroplast sequence of *Gossypium hirsutum* (NC_007944) was also included as an outgroup (Supplementary Table 1). All chloroplast sequences were aligned in MAFFT using default parameters. The best-fit model (K3Pu+F+R4) was estimated using ModelFinder (Kalyaanamoorthy et al., 2017) with Bayesian Information Criterion (BIC) implemented in IQ-TREE v2.0.3 (Minh et al., 2020). Based on the best-fit model (K3Pu+F+R4), we inferred a maximum likelihood tree with 1,000 bootstrap replicates using IQ-TREE. The tree was rooted at midpoint and visualized using FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

3 Results and discussion

3.1 Chloroplast genome assembly

The complete chloroplast genome of Baekdansim was constructed using the PacBio long-read sequencing platform (Figure 1). Due to the high-quality sequence data provided by the PacBio long-read sequencing technique and its capacity to assemble long reads, a single contig and, ultimately, the whole chloroplast genome of *H. syriacus* could be extracted (Chin et al., 2013). The complete chloroplast genome was 161,026 bp and had a quadripartite structure, including a pair of IR regions (IRA and IRB) separated by an LSC (89,705 bp) region and an SSC (19,831 bp) region. The SSC region in the genome assemblies of species of the family Malvaceae in the NCBI database, as determined using short reads, are usually bidirectional. Therefore, the direction of the SSC region was a focus of comparative genome analysis. The genome assembly of Baekdansim consisted of a single contig and was used as a resource to investigate the direction of the SSC region. In a previous study, the primary hypothesis was that the direction of the SSC may have been due to a recombination event between the two IR regions. The alternative hypothesis was that the direction of the SSC region depended on the assembly method; the precise direction of the SSC region was unknown because a short-read-based sequencing platform was used (Cheng et al., 2020). This study showed that the whole chloroplast sequence, which was obtained as a single contig, spanned the whole LSC-IR-SSC area and that the gene order of close species of the family Malvaceae with an inverted SSC structure was exactly reversed. Based on these results, it could be concluded that the SSC direction was changed because of a misassembly induced by constraints of the short read-based sequencing platform. To perform an accurate comparative analysis of species of the tribe Hibisceae, the mis-assembled section was corrected based on the SSC strand derived from long-read sequencing using an in-house Python script.

3.2 Genome structure and gene content

The complete chloroplast genome of Baekdansim contains 113 genes, including 79 protein-coding genes, 4 rRNAs, and 30 tRNAs. Multiple genes were duplicated in the IR regions, including 5 protein-coding genes (*rpl2, rps7, rpl23, ndhB*, and *ycf2*), 7 tRNAs (*trnA-UGC, trnI-GAU, trnN-GUU, trnV-GAC, trnL-CAA, trnR-ACG*, and *trnI-CAU*), and 4 rRNAs (*rrn5, rrn4.5, rrn23*, and *rrn16*). As generally observed in other angiosperms, 18 intron-containing genes were also detected in the Baekdansim chloroplast genome (Redwan et al., 2015; Mo et al., 2020). Eighteen genes contained one or more introns; of which, 11 encoded proteins (*atpF, clpP1, ndhA, ndhB, pafI, petB, petD, rpoC1, rps16, rpl2*, and *rpl16*), 6 encoded tRNAs *(trnA-UGC, trnG-UCC, trnI-GAU, trnK-UUU, trnL-UAA*, and *trnV-UAC*), and 1 encoded an rRNA (*rrn23*). The *rps12* gene exhibited a trans-spliced form with its 5′ terminal present in the LSC region, and its 3′ end had a single copy present in each of the two IR regions, similar to the patterns observed in other terrestrial plants (Hildebrand et al., 1988; Lee et al., 2019) (Supplementary Table 2). A single copy of *ycf1* was present due to its position in the SSC region instead of in the IR regions. This finding is consistent with a previous study of *Distemonanthus benthamianus*, in which IR contraction was observed (Bai et al., 2021). According to previous reports, the plastome length varies according to the IR length, suggesting that the chloroplast length of *H. syriacus* is also affected by this IR length variation (Zhu et al., 2016; Liang et al., 2022; Lu et al., 2022).

3.3 Repeat analysis

Repeat motifs, which are widely distributed in chloroplast genomes, play an important role in genome evolution (Powell et al., 1995; Yang et al., 2011; Xue et al., 2012; Liu et al., 2018). The number of SSR motifs in the Baekdansim plastome was investigated using MISA software. We identified 109 SSRs (microsatellites) among which 81 (74.3%) consisted exclusively of A/T. Similar to a previous report, the majority of mononucleotide repeats were A/T, and most SSRs consisted of mononucleotide repeats (George et al., 2015). We found 82 (75.2%) mono-, 10 (9.3%) di-, 7 (6.4%) tri-, 6 (5.5%) tetra-, and 4 (3.7%) penta-nucleotides (Supplementary Table 3).

Long-repeat elements are crucial for not only structural variation in chloroplast genomes but also intermolecular recombination, leading to genome diversity (Park et al., 2017; Kong et al., 2021). Complex repeats in Baekdansim were discovered using the REPuter program. The repeat length ranged from 30 bp to 78 bp, which corresponds to the typical range of other plastomes (Greiner et al., 2008; Li et al., 2017). The most abundant repeats were forward repeats, followed by palindromic repeats and reverse repeats (Supplementary Figure 1). These identified repeats and SSRs will be useful for developing molecular markers for genetic diversity and evolution studies.

3.4 Comparison of chloroplast genome structure and nucleotide diversity

Only *Abelmoschus* species contained *rps3b, rps19b*, and *rpl22b*, but other genes were detected in all 13 species of the tribe Hibisceae (Supplementary Figure 2). IRs were more conserved than LSC and SSC sections, while non-coding regions were more divergent than coding regions (Supplementary Figure 3). These results were

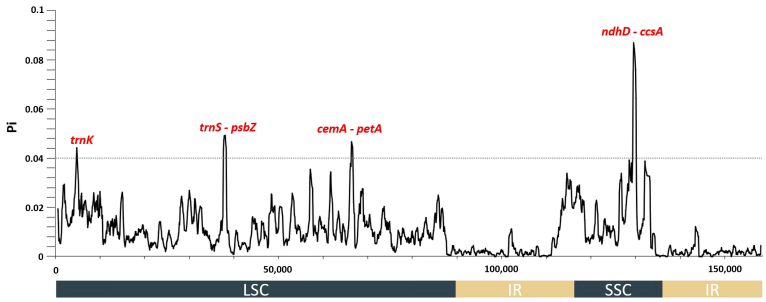


FIGURE 2
Nucleotide diversity values among 13 species of the tribe Hibisceae were calculated using whole plastomes. Mutational hotspots (Pi > 0.04) are denoted above the corresponding gene position.

congruent with findings in other land plant species (Jian et al., 2018; Kim et al., 2020). In addition, the intergenic spacer regions among several gene pairs varied remarkably in chloroplast genomes of the tribe Hibisceae. For instance, these regions differed markedly among *trnH-GUG-psbA*, *trnK-UUU-rps16*, *trnF-GAA-ndhJ*, *atpB-rbcL*, *rps12-trnV-GAC*, *ndhl-ndhG*, and *ndhD-ccsA*. The highest level of nucleotide diversity was identified in a few of these intergenic spacer regions. Collectively, these results suggested these regions might indicate the rapid evolution of the tribe Hibisceae (Figure 2). The nucleotide diversity among the 13 chloroplast genomes of the tribe Hibisceae was calculated. The results indicated four highly divergent hotspots, *trnK*, *trnS-psbZ*, *cemA-petA*, and *ndhD-ccsA*, with a threshold of 0.04. All of these hotspots were found in single-copy (LSC and SSC) regions. The most variation was observed in the *ndhD-ccsA* region (0.08703) (Figure 2). It will be important to determine whether these regions could be employed as DNA barcodes to clarify close relationships within the tribe Hibisceae.

Contraction or expansion of the single-copy and IR regions commonly occurs in various angiosperms (Jian et al., 2018; Li and Zheng, 2018; Henriquez et al., 2020; Guo et al., 2021; Liu et al., 2021). These alterations are considered a major mechanism that causes size variation of the chloroplast genome and evolutionary events (Zhu et al., 2016; Liang et al., 2022; Lu et al., 2022). Four junctions between the two single-copy regions and the two IR regions of 12 representative species of the tribe Hibisceae were thoroughly compared to examine chloroplast genome variation in the tribe Hibisceae (Figure 3; Supplementary Table 4). IR regions are relatively conserved in the *Hibiscus* genus; nonetheless, considerable contraction and expansion occur in the IR/SSC regions. The *ycf1* gene was displaced from the IRB to the SSC region at the IRB/SSC boundary in the chloroplast genomes of *H. syriacus* and *H. rosa-sinensis* by 608 bp and 113 bp, respectively.

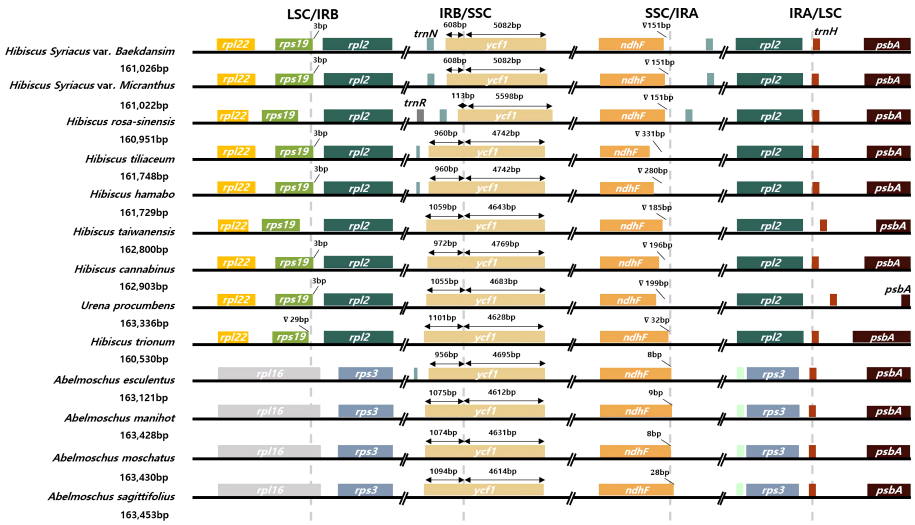


FIGURE 3
Comparison of the junction sites among 13 species of the tribe Hibisceae.

This movement indicates IR contraction in the chloroplast genomes of these species. The *ndhF* gene was shifted from the SSC region to the IRA region, and *rpl16* was shifted from the LSC region to the IRB region in *Abelmoschus* species, according to comparisons between *Hibiscus* and *Abelmoschus* species. The longer chloroplast genome in *Abelmoschus* than in the *Hibiscus* species could be attributed to this IR expansion. In previous reports, shifting of genes to the IR regions or SSC region led to size variation of IR regions in the family Malvaceae. The current study showed that the overall length of the plastome was affected by this size variation shown in previous reports (Dugas et al., 2015; Wang et al., 2016; Guo et al., 2021).

3.5 Putative RNA-editing sites

RNA editing is a post-transcriptional regulation mechanism that can result in the alteration of ribonucleotides at specific sites (Maier et al., 1996). According to previous research, C-to-U conversion is the primary factor responsible for RNA editing (Tsudzuki et al., 2001). Two possible RNA-editing sites were predicted in the start codon of *ndhD* and the stop codon of *rpoC1* in the chloroplast genomes of the tribe Hibisceae (Supplementary Figure 4). In particular, *ndhD* was edited at a high level in *Galium* species. In addition, the start codon (ACG) of *ndhD* in nine species of the family Rubiaceae was affected by an RNA-editing event, which is consistent with the pattern of RNA editing in species of the tribe Hibisceae (Zhang et al., 2019). RNA editing regulates gene expression and has a substantial impact on translation (Maier et al., 1996). RNA editing in the protein-coding region results in codon alterations that lead to amino acid substitution, which may affect the stability of the tertiary structure of proteins (Gommans et al., 2009). Furthermore, these alterations have been related to the generation of genetic diversity, which is a factor in adaptive evolution (Gommans et al., 2009). In tobacco, frequent editing occurs in *ndh*, which encodes the subunits of a plastid NAD(P)H dehydrogenase (Hirose and Sugiura, 1997; Fiebig et al., 2004). Expression of *ndhD* in the tobacco chloroplast, as determined by RNA editing, to create the start codon was greatest in young and photosynthetically active leaves (Hirose et al., 1997). In addition, although *ndh* gene products are dispensable under normal growth conditions, editing is likely essential for the appropriate function of the Ndh protein complex and cyclic electron flow under stress conditions. Fixation of a mutation in a non-essential gene allows plasticity and sufficient time for the evolution of a mutation-compensating editing capacity under moderate selective pressure (Fiebig et al., 2004). Therefore, the occurrence of RNA editing in *ndhD* at the same site in all species of the tribe Hibisceae could be regarded as a result of environmental adaptation. In general, species of the tribe Hibisceae are tolerant of abiotic stresses such as cold, drought, and salt stresses (Zhan et al., 2019; An et al., 2020; Eo et al., 2020; Mahougnon et al., 2021; Chen et al., 2022). Thus, the fixation of RNA editing might have occurred *via* long periods of adaptation to environmental changes during evolution.

3.6 Codon usage pattern and phylogenetic analysis

Many terrestrial plants exhibit codon usage bias, which is considered to play a substantial role in regulating translation dynamics (Du et al., 2020). Recent studies have demonstrated that codon preferences significantly influence the evolution of the chloroplast genome by balancing natural selection and mutational biases (Akashi and Eyre-Walker, 1998; Raubeson et al., 2007; Hershberg and Petrov, 2008). In this study, the RSCU of protein-coding genes in the chloroplast genome of the tribe Hibisceae was investigated and identified (Figure 4). Among the protein-coding codons, the most frequently encoded was leucine, followed by those encoding arginine and alanine; the GAC codon, which encodes aspartic acid, had the lowest usage frequency. If neutral mutations occur at the third codon position, GC and AT would equally present among the codon groups within a chloroplast genome (Zhang et al., 2007). However, most codons showed a bias toward an A/U ending, and these findings are consistent with those observed in other chloroplast genomes (Yan et al., 2019; Du et al., 2020). Previous studies revealed that this unequal usage of nucleotides derived from mutation selection and natural selection was the primary driver of codon bias in angiosperms (Nie et al., 2014; Wang et al., 2020; Zhang et al., 2021). These findings indicate that the high proportion of A/U-ending codons in the chloroplast genome, along with the selective pressure of the chloroplast genome of the tribe Hibisceae, may have driven several degenerate codon biases.

Moreover, species within the tribe Hibisceae were largely clustered into five groups by the RSCU pattern (Group I: *H. syriacus*, Group II: *H. taiwanensis*, *H. tiliaceum*, and *H. hamabo*, Group III: *Urena procumbens* and *H. cannabinus*, Group IV: *H. rosa-sinensis*, Group V: *Abelmoschus sagittifolius*, *A. moschatus*, *A. manihot*, *A. esculentus*, and *H. trionum*) (Figure 4). Phylogenetic analysis was performed on an alignment of the whole chloroplast genome sequences of 13 species of the tribe Hibisceae (Figure 5). It is noteworthy that *H. trionum* formed a clade with *Abelmoschus* species. This tendency was linked to the codon usage pattern of protein-coding genes in the tribe Hibisceae (Figure 4). Despite belonging to a distinct genus, it was assumed that the similarity of codons with other genera may have affected clade formation among other genera. The association between codon usage patterns and the phylogenic topology inferred from the whole chloroplast genome provides strong support for the hypothesis that nucleotide bias induces codon bias.

4 Conclusion

In this study, the complete chloroplast genome of Baekdansim was constructed *via* a long-read sequencing platform for the first time. Through comparisons among species of the tribe Hibisceae, we found that four mutational hot spots could be used to develop DNA barcodes. Furthermore, we identified fixation of candidate RNA-editing sites, a preference for A/U-terminated codons, and a notable codon usage pattern related to phylogenetic relationships.

10.3389/fpls.2023.1111968

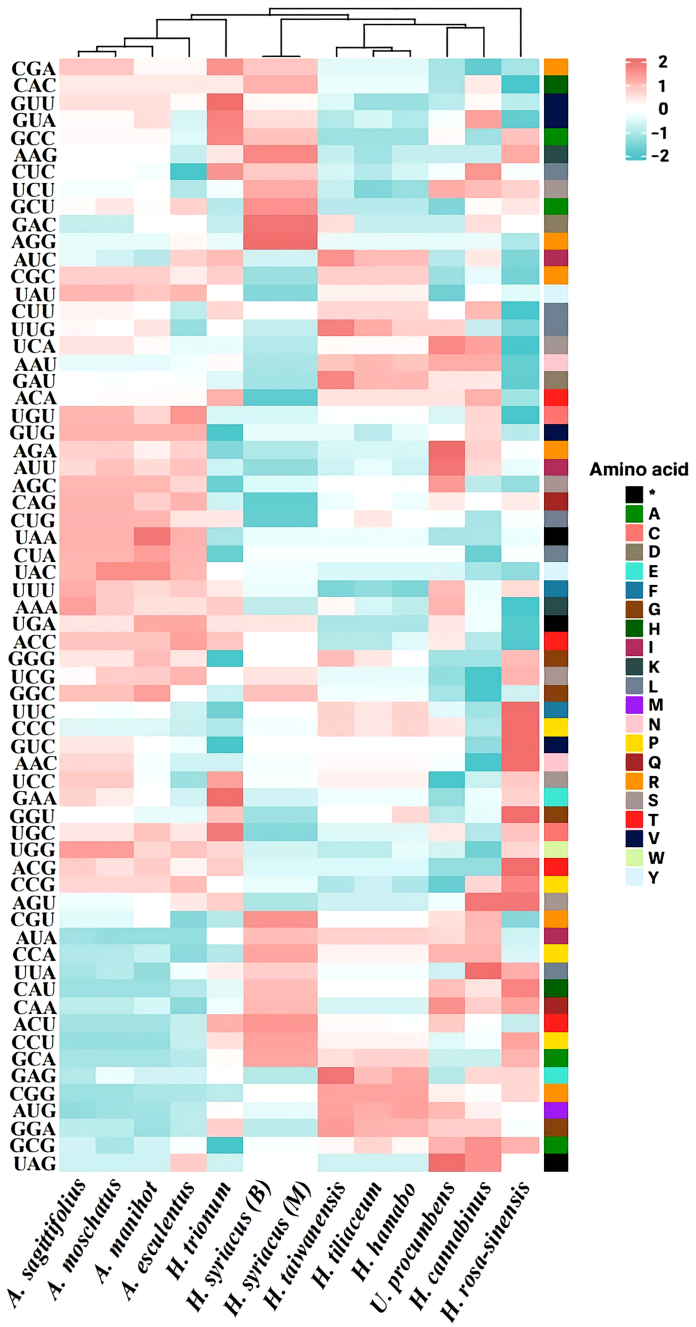


FIGURE 4
Relative synonymous codon usage (RSCU) pattern of chloroplast genes among 13 species of the tribe Hibisceae.

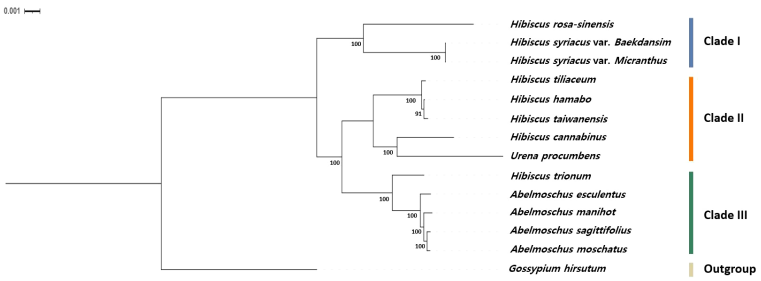


FIGURE 5
Phylogenetic relationships based on the whole chloroplast genomes of 13 species of the tribe Hibisceae. The bootstrap values were based on 1000 replicates and are denoted next to the branches.

Comparison analysis of whole chloroplast genomes of the tribe Hibisceae offers a valuable genomic resource for understanding the evolution and adaptation of this tribe and its relatives.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

HK and Y-MK conceived and designed this study; HK, A-YS, and SH analyzed the data; HK and Y-MK wrote the manuscript; HK, A-YS, and Y-MK revised the manuscript; Y-MK supervised this study. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1I1A2044678), and the Korea Forest Service of the Korean government through the R&D Program for Forestry Technology (Project No. 2014071H10-2122-AA04).

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1111968/full#supplementary-material

References

Akashi, H., and Eyre-Walker, A. (1998). Translational selection and molecular evolution. *Curr. Opin. Genet. Dev.* 8 (6), 688–693. doi: 10.1016/S0959-437X(98)80038-5

Akpan, G. (2000). Cytogenetic characteristics and the breeding system in six *Hibiscus* species. *Theor. Appl. Genet.* 100 (2), 315–318. doi: 10.1007/s001220050041

Amiryousefi, A., Hyvönen, J., and Poczai, P. (2018). IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34 (17), 3030–3031. doi: 10.1093/bioinformatics/bty220

An, X., Jin, G., Luo, X., Chen, C., Li, W., and Zhu, G. (2020). Transcriptome analysis and transcription factors responsive to drought stress in *Hibiscus cannabinus*. *PeerJ* 8, e8470. doi: 10.7717/peerj.8470

Bai, H.-R., Oyebanji, O., Zhang, R., and Yi, T.-S. (2021). Plastid phylogenomic insights into the evolution of subfamily dialioideae (Leguminosae). *Plant Diversity* 43 (1), 27–34. doi: 10.1016/j.pld.2020.06.008

Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33 (16), 2583–2585. doi: 10.1093/bioinformatics/btx198

Cheng, Y., Zhang, L., Qi, J., and Zhang, L. (2020). Complete chloroplast genome sequence of *Hibiscus cannabinus* and comparative analysis of the malvaceae family. *Front. Genet.* 11, 227. doi: 10.3389/fgene.2020.00227

Chen, M., She, Z., Aslam, M., Liu, T., Wang, Z., Qi, J., et al. (2022). Genomic insights of the *WRKY* genes in kenaf (*Hibiscus cannabinus* l.) reveal that *HcWRKY44* improves the plant's tolerance to the salinity stress. *Front. Plant Sci.* 13, 984233. doi: 10.3389/fpls.2022.984233

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10 (6), 563–569. doi: 10.1038/nmeth.2474

Daniell, H., Lin, C.-S., Yu, M., and Chang, W.-J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17 (1), 1–29. doi: 10.1186/s13059-016-1004-2

Dugas, D. V., Hernandez, D., Koenen, E. J., Schwarz, E., Straub, S., Hughes, C. E., et al. (2015). Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions and accelerated rate of evolution in *clpP*. *Sci. Rep.* 5 (1), 1–13. doi: 10.1038/srep16958

Du, X., Zeng, T., Feng, Q., Hu, L., Luo, X., Weng, Q., et al. (2020). The complete chloroplast genome sequence of yellow mustard (*Sinapis alba* l.) and its phylogenetic relationship to other brassicaceae species. *Gene* 731, 144340. doi: 10.1016/j.gene.2020.144340

Eo, H. J., Kwon, H. Y., Da Kim, S., Kang, Y., Park, Y., and Park, G. H. (2020). GC/MS analysis and anti-inflammatory effect of leaf extracts from *Hibiscus syriacus* through inhibition of NF-κB and MAPKs signaling in LPS-stimulated RAW264.7 macrophages. *Plant Biotechnol. Rep.* 14 (5), 539–546. doi: 10.1007/s11816-020-00628-3

Fiebig, A., Stegemann, S., and Bock, R. (2004). Rapid evolution of RNA editing sites in a small non-essential plastid gene. *Nucleic Acids Res.* 32 (12), 3615–3622. doi: 10.1093/nar/gkh695

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39 (suppl_2), W29–W37. doi: 10.1093/nar/gkr367

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32 (suppl_2), W273–W279. doi: 10.1093/nar/gkh458

George, B., Bhatt, B. S., Awasthi, M., George, B., and Singh, A. K. (2015). Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. *Curr. Genet.* 61 (4), 665–677. doi: 10.1007/s00294-015-0495-9

Gommans, W. M., Mullen, S. P., and Maas, S. (2009). RNA Editing: a driving force for adaptive evolution? *Bioessays* 31 (10), 1137–1145. doi: 10.1002/bies.200900045

Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47 (W1), W59–W64. doi: 10.1093/nar/gkz238

Greiner, S., Wang, X., Rauwolf, U., Silber, M. V., Mayer, K., Meurer, J., et al. (2008). The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. sequence evaluation and plastome evolution. *Nucleic Acids Res.* 36 (7), 2366–2378. doi: 10.1093/nar/gkn081

Guo, Y.-Y., Yang, J.-X., Bai, M.-Z., Zhang, G.-Q., and Liu, Z.-J. (2021). The chloroplast genome evolution of Venus slipper (*Paphiopedilum*): IR expansion, SSC contraction, and highly rearranged SSC regions. *BMC Plant Biol.* 21 (1), 1–14. doi: 10.1186/s12870-021-03053-y

Henriquez, C. L., Ahmed, I., Carlsen, M. M., Zuluaga, A., Croat, T. B., and McKain, M. R. (2020). Evolutionary dynamics of chloroplast genomes in subfamily aroideae (Araceae). *Genomics* 112 (3), 2349–2360. doi: 10.1016/j.ygeno.2020.01.006

Hershberg, R., and Petrov, D. A. (2008). Selection on codon bias. *Annu. Rev. Genet.* 42, 287–299. doi: 10.1146/annurev.genet.42.110807.091442

Hildebrand, M., Hallick, R. B., Passavant, C. W., and Bourque, D. P. (1988). *Trans*-splicing in chloroplasts: the *rps 12* loci of nicotiana tabacum. *Proc. Natl. Acad. Sci.* 85 (2), 372–376. doi: 10.1073/pnas.85.2.372

Hirose, T., and Sugiura, M. (1997). Both RNA editing and RNA cleavage are required for translation of tobacco chloroplast *ndhD* mRNA: a possible regulatory mechanism for the expression of a chloroplast operon consisting of functionally unrelated genes. *EMBO J.* 16 (22), 6804–6811. doi: 10.1093/emboj/16.22.6804

Huang, H., Shi, C., Liu, Y., Mao, S.-Y., and Gao, L.-Z. (2014). Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evolutionary Biol.* 14 (1), 1–17. doi: 10.1186/1471-2148-14-151

Jian, H.-Y., Zhang, Y.-H., Yan, H.-J., Qiu, X.-Q., Wang, Q.-G., Li, S.-B., et al. (2018). The complete chloroplast genome of a key ancestor of modern roses, *Rosa chinensis* var. *spontanea*, and a comparison with congeneric species. *Molecules* 23 (2), 389. doi: 10.3390/molecules23020389

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14 (6), 587–589. doi: 10.1038/nmeth.4285

Kent, W. J. (2002). BLAT–the BLAST-like alignment tool. *Genome Res.* 12 (4), 656–664. doi: 10.1101/gr.229202

Kim, Y.-M., Kim, S., Koo, N., Shin, A.-Y., Yeom, S.-I., Seo, E., et al. (2017). Genome analysis of *Hibiscus syriacus* provides insights of polyploidization and indeterminate flowering in woody plants. *DNA Res.* 24 (1), 71–80. doi: 10.1093/dnares/dsw049

Kim, Y., Shin, J., Oh, D.-R., Kim, A.-Y., and Choi, C. (2020). Comparative analysis of complete chloroplast genome sequences and insertion-deletion (Indel) polymorphisms to distinguish five *vaccinium* species. *Forests* 11 (9), 927. doi: 10.3390/f11090927

Kong, B. L.-H., Park, H.-S., Lau, T.-W. D., Lin, Z., Yang, T.-J., and Shaw, P.-C. (2021). Comparative analysis and phylogenetic investigation of Hong Kong *Ilex* chloroplast genomes. *Sci. Rep.* 11 (1), 1–13. doi: 10.1038/s41598-021-84705-9

Kumar, S., Tamura, K., and Nei, M. (1994). MEGA: molecular evolutionary genetics analysis software for microcomputers. *Bioinformatics* 10 (2), 189–191. doi: 10.1093/bioinformatics/10.2.189

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29 (22), 4633–4642. doi: 10.1093/nar/29.22.4633

Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32 (1), 11–16. doi: 10.1093/nar/gkh152

Lee, H.-L., Jansen, R. K., Chumley, T. W., and Kim, K.-J. (2007). Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Mol. Biol. Evol.* 24 (5), 1161–1180. doi: 10.1093/molbev/msm036

Lee, K., Park, S. J., Colas des Francs-Small, C., Whitby, M., Small, I., and Kang, H. (2019). The coordinated action of PPR 4 and EMB 2654 on each intron half mediates *trans*-splicing of *rps12* transcripts in plant chloroplasts. *Plant J.* 100 (6), 1193–1207. doi: 10.1111/tpj.14509

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34 (18), 3094–3100. doi: 10.1093/bioinformatics/bty191

Liang, D., Wang, H., Zhang, J., Zhao, Y., and Wu, F. (2022). Complete chloroplast genome sequence of *Fagus longipetiolata* seemen (Fagaceae): Genome structure, adaptive evolution, and phylogenetic relationships. *Life* 12 (1), 92. doi: 10.3390/life12010092

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, B., Lin, F., Huang, P., Guo, W., and Zheng, Y. (2017). Complete chloroplast genome sequence of *Decaisnea insignis*: Genome organization, genomic resources and comparative analysis. *Sci. Rep.* 7 (1), 1–10. doi: 10.1038/s41598-017-10409-8

Liu, J., Jiang, M., Chen, H., Liu, Y., Liu, C., and Wu, W. (2021). Comparative genome analysis revealed gene inversions, boundary expansions and contractions, and gene loss in the *Stemona sessilifolia* (Miq.) miq. chloroplast genome. *PloS One* 16 (6), e0247736. doi: 10.1371/journal.pone.0247736

Liu, H.-Y., Yu, Y., Deng, Y.-Q., Li, J., Huang, Z.-X., and Zhou, S.-D. (2018). The chloroplast genome of *Lilium henrici*: genome structure and comparative analysis. *Molecules* 23 (6), 1276. doi: 10.3390/molecules23061276

Li, J., Ye, G.-y., Liu, H.-l., and Wang, Z.-h. (2020). Complete chloroplast genomes of three important species, *Abelmoschus moschatus, a. manihot* and *A. sagittifolius*: Genome structures, mutational hotspots, comparative and phylogenetic analysis in malvaceae. *PloS One* 15 (11), e0242591. doi: 10.1371/journal.pone.0242591

Li, W., Zhang, C., Guo, X., Liu, Q., and Wang, K. (2019). Complete chloroplast genome of *Camellia japonica* genome structures, comparative and phylogenetic analysis. *PloS One* 14 (5), e0216645. doi: 10.1371/journal.pone.0216645

Li, B., and Zheng, Y. (2018). Dynamic evolution and phylogenomic analysis of the chloroplast genome in schisandraceae. *Sci. Rep.* 8 (1), 1–11. doi: 10.1038/s41598-018-27453-7

Lowe, T. M., and Chan, P. P. (2016). tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44 (W1), W54–W57. doi: 10.1093/nar/gkw413

Lu, Q.-X., Chang, X., Gao, J., Wu, X., Wu, J., Qi, Z.-C., et al. (2022). Evolutionary comparison of the complete chloroplast genomes in *Convallaria* species and phylogenetic study of asparagaceae. *Genes* 13 (10), 1724. doi: 10.3390/genes13101724

Mahougnon, B. G. G., Julien, K. K., Armel, C. G. M., and Christophe, B. G. (2021). Salinity resistance strategy of okra (*Abelmoschus esculentus* l. moench) cultivars produced in Benin republic. *Int. J. Plant Physiol. Biochem.* 13 (1), 19–29. doi: 10.5897/IJPPB2021.0308

Maier, R. M., Zeitz, P., Kössel, H., Bonnard, G., Gualberto, J. M., and Grienenberger, J. M. (1996). RNA Editing in plant mitochondria and chloroplasts. *Post-transcriptional control Gene Expression Plants* 32 (1-2), 343–365. doi: 10.1007/978-94-009-0353-1_16

Mehmood, F., Shahzadi, I., Waseem, S., Mirza, B., Ahmed, I., and Waheed, M. T. (2020). Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): comparative analyses and identification of mutational hotspots. *Genomics* 112 (1), 581–591. doi: 10.1016/j.ygeno.2019.04.010

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37 (5), 1530–1534. doi: 10.1093/molbev/msaa015

Mo, Z., Lou, W., Chen, Y., Jia, X., Zhai, M., Guo, Z., et al. (2020). The chloroplast genome of *Carya illinoinensis*: genome structure, adaptive evolution, and phylogenetic analysis. *Forests* 11 (2), 207. doi: 10.3390/f11020207

Nie, X., Deng, P., Feng, K., Liu, P., Du, X., You, F. M., et al. (2014). Comparative analysis of codon usage patterns in chloroplast genomes of the asteraceae family. *Plant Mol. Biol. Rep.* 32 (4), 828–840. doi: 10.1007/s11105-013-0691-z

Paoletti, E., Ferrara, A. M., Calatayud, V., Cerveró, J., Giannetti, F., Sanz, M. J., et al. (2009). Deciduous shrubs for ozone bioindication: *Hibiscus syriacus* as an example. *Environ. pollut.* 157 (3), 865–870. doi: 10.1016/j.envpol.2008.11.009

Park, I., Yang, S., Choi, G., Kim, W. J., and Moon, B. C. (2017). The complete chloroplast genome sequences of *Aconitum pseudolaeve* and *Aconitum longecassidatum*, and development of molecular markers for distinguishing species in the *Aconitum* subgenus *Lycoctonum*. *Molecules* 22 (11), 2012. doi: 10.3390/molecules22112012

Powell, W., Morgante, M., McDevitt, R., Vendramin, G., and Rafalski, J. (1995). Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proc. Natl. Acad. Sci.* 92 (17), 7759–7763. doi: 10.1073/pnas.92.17.7759

Raubeson, L. A., Peery, R., Chumley, T. W., Dziubek, C., Fourcade, H. M., Boore, J. L., et al. (2007). Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 8 (1), 1–27. doi: 10.1186/1471-2164-8-174

Redwan, R., Saidin, A., and Kumar, S. (2015). Complete chloroplast genome sequence of MD-2 pineapple and its comparative analysis among nine other plants from the subclass commelinidae. *BMC Plant Biol.* 15 (1), 1–20. doi: 10.1186/s12870-015-0587-1

Rizk, R. M., and Soliman, M. I. (2014). Biochemical and molecular genetic characterization of some species of family malvaceae, Egypt. *Egyptian J. Basic Appl. Sci.* 1 (3-4), 167–176. doi: 10.1016/j.ejbas.2014.06.002

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34 (12), 3299–3302. doi: 10.1093/molbev/msx248

Saba, N., Jawaid, M., Hakeem, K. R., Paridah, M. T., Khalina, A., and Alothman, O. Y. (2015). Potential of bioenergy production from industrial kenaf (*Hibiscus cannabinus* l.) based on Malaysian perspective. *Renewable Sustain. Energy Rev.* 42, 446–459. doi: 10.1016/j.rser.2014.10.029

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq–versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45 (W1), W6–W11. doi: 10.1093/nar/gkx391

Tsudzuki, T., Wakasugi, T., and Sugiura, M. (2001). Comparative analysis of RNA editing sites in higher plant chloroplasts. *J. Mol. Evol.* 53 (4), 327–332. doi: 10.1007/s002390010222

Wang, L., Du, H., Wang, D., and Cao, D. (2016). Complete chloroplast genome sequences of *Eucommia ulmoides*: genome structure and evolution. *Tree Genet. Genomes* 12 (1), 1–15. doi: 10.1007/s11295-016-0970-6

Wang, W., Schalamun, M., Morales-Suarez, A., Kainer, D., Schwessinger, B., and Lanfear, R. (2018). Assembly of chloroplast genomes with long-and short-read data: a comparison of approaches using *Eucalyptus pauciflora* as a test case. *BMC Genomics* 19 (1), 1–15. doi: 10.1186/s12864-018-5348-8

Wang, Z., Xu, B., Li, B., Zhou, Q., Wang, G., Jiang, X., et al. (2020). Comparative analysis of codon usage patterns in chloroplast genomes of six *Euphorbiaceae* species. *PeerJ* 8, e8251. doi: 10.7717/peerj.8251

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PloS Comput. Biol.* 13 (6), e1005595. doi: 10.1371/journal.pcbi.1005595

Wu, M., Li, Q., Hu, Z., Li, X., and Chen, S. (2017). The complete amomum kravanh chloroplast genome sequence and phylogenetic analysis of the commelinids. *Molecules* 22 (11), 1875. doi: 10.3390/molecules22111875

Xue, J., Wang, S., and Zhou, S. L. (2012). Polymorphic chloroplast microsatellite loci in *Nelumbo* (Nelumbonaceae). *Am. J. Bot.* 99 (6), e240–e244. doi: 10.3732/ajb.1100547

Yan, C., Du, J., Gao, L., Li, Y., and Hou, X. (2019). The complete chloroplast genome sequence of watercress (*Nasturtium officinale* r. br.): Genome organization, adaptive evolution and phylogenetic relationships in cardamineae. *Gene* 699, 24–36. doi: 10.1016/j.gene.2019.02.075

Yang, A. H., Zhang, J. J., Yao, X. H., and Huang, H. W. (2011). Chloroplast microsatellite markers in *Liriodendron tulipifera* (Magnoliaceae) and cross-species amplification in l. chinense. *Am. J. Bot.* 98 (5), e123–e126. doi: 10.3732/ajb.1000532

Yoo, I.-D., Yun, B.-S., Lee, I.-K., Ryoo, I.-J., Choung, D.-H., and Han, K.-H. (1998). Three naphthalenes from root bark of *Hibiscus syriacus*. *Phytochemistry* 47 (5), 799–802. doi: 10.1016/S0031-9422(97)00674-2

Yun, B.-S., Ryoo, I.-J., Lee, I.-K., Park, K.-H., Choung, D.-H., Han, K.-H., et al. (1999). Two bioactive pentacyclic triterpene esters from the root bark of *Hibiscus syriacus*. *J. Natural products* 62 (5), 764–766. doi: 10.1021/np9804637

Zerpa-Catanho, D., Zhang, X., Song, J., Hernandez, A. G., and Ming, R. (2021). Ultra-long DNA molecule isolation from plant nuclei for ultra-long read genome sequencing. *STAR protocols* 2 (1), 100343.

Zhang, P., Xu, W., Lu, X., and Wang, L. (2021). Analysis of codon usage bias of chloroplast genomes in *Gynostemma* species. *Physiol. Mol. Biol. Plants* 27 (12), 2727–2737. doi: 10.1007/s12298-021-01105-z

Zhang, Y., Zhang, J.-W., Yang, Y., and Li, X.-N. (2019). Structural and comparative analysis of the complete chloroplast genome of a mangrove plant: *Scyphiphora hydrophyllacea* gaertn. f. and related rubiaceae species. *Forests* 10 (11), 1000. doi: 10.3390/f10111000

Zhang, W. J., Zhou, J., Li, Z. F., Wang, L., Gu, X., and Zhong, Y. (2007). Comparative analysis of codon usage patterns among mitochondrion, chloroplast and nuclear genes in *Triticum aestivum* l. *J. Integr. Plant Biol.* 49 (2), 246–254. doi: 10.1111/j.1744-7909.2007.00404.x

Zhan, Y., Wu, Q., Chen, Y., Tang, M., Sun, C., Sun, J., et al. (2019). Comparative proteomic analysis of okra (*Abelmoschus esculentus* l.) seedlings under salt stress. *BMC Genomics* 20 (1), 1–12. doi: 10.1186/s12864-019-5737-7

Zhu, A., Guo, W., Gupta, S., Fan, W., and Mower, J. P. (2016). Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol.* 209 (4), 1747–1756. doi: 10.1111/nph.13743

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 August 2023
DOI 10.3389/fpls.2023.1237749

OPEN ACCESS

EDITED BY
Jeremie Benjamin Fant,
Chicago Botanic Garden, United States

REVIEWED BY
Michael J. Moore,
Oberlin College, United States
Josué Barrera-Redondo,
Max Planck Society, Germany

*CORRESPONDENCE
Ethan Baldwin
ethan.baldwin@uga.edu

RECEIVED 09 June 2023
ACCEPTED 20 July 2023
PUBLISHED 14 August 2023

CITATION
Baldwin E, McNair M and Leebens-Mack J (2023) Rampant chloroplast capture in Sarracenia revealed by plastome phylogeny. *Front. Plant Sci.* 14:1237749. doi: 10.3389/fpls.2023.1237749

Rampant chloroplast capture in Sarracenia revealed by plastome phylogeny

Ethan Baldwin[1]*, Mason McNair[2] and Jim Leebens-Mack[1]

[1]Department of Plant Biology, University of Georgia, Athens, GA, United States, [2]Department of Plant & Environmental Science, Clemson University, Florence, SC, United States

Introgression can produce novel genetic variation in organisms that hybridize. Sympatric species pairs in the carnivorous plant genus *Sarracenia* L. frequently hybridize, and all known hybrids are fertile. Despite being a desirable system for studying the evolutionary consequences of hybridization, the extent to which introgression occurs in the genus is limited to a few species in only two field sites. Previous phylogenomic analysis of *Sarracenia* estimated a highly resolved species tree from 199 nuclear genes, but revealed a plastid genome that is highly discordant with the species tree. Such cytonuclear discordance could be caused by chloroplast introgression (i.e. chloroplast capture) or incomplete lineage sorting (ILS). To better understand the extent to which introgression is occurring in *Sarracenia*, the chloroplast capture and ILS hypotheses were formally evaluated. Plastomes were assembled *de-novo* from sequencing reads generated from 17 individuals in addition to reads obtained from the previous study. Assemblies of 14 whole plastomes were generated and annotated, and the remaining fragmented assemblies were scaffolded to these whole-plastome assemblies. Coding sequence from 79 homologous genes were aligned and concatenated for maximum-likelihood phylogeny estimation. The plastome tree is extremely discordant with the published species tree. Plastome trees were simulated under the coalescent and tree distance from the species tree was calculated to generate a null distribution of discordance that is expected under ILS alone. A t-test rejected the null hypothesis that ILS could cause the level of discordance seen in the plastome tree, suggesting that chloroplast capture must be invoked to explain the discordance. Due to the extreme level of discordance in the plastome tree, it is likely that chloroplast capture has been common in the evolutionary history of *Sarracenia*.

KEYWORDS

hybridization, chloroplast capture, gene flow, carnivorous plant, *Sarracenia*, phylogenomics, plastome

1 Introduction

Evolutionary biologists have long been interested in hybridization as a process that generates biodiversity. Hybridization leading to introgression introduces genetic information to a species, which increases genetic variation for selection to act on and provides opportunity for adaptive evolution (Pease et al., 2016; Grant and Grant, 2019; Meier et al., 2019). Organisms that readily hybridize may be subject to these evolutionary forces. However, the formation of hybrids does not imply that introgression (transfer of genome segments between hybridizing species) is occurring, as hybrids must reproduce with the parental population and introgressed alleles must survive in the face of natural selection and genetic drift. Identifying the extent to which hybridizing taxa are exchanging genetic material sheds light on the processes that generate and maintain variation within them.

Sarracenia L. is a genus of 8-11 species of carnivorous plants native to North America. It is one of the three extant genera in the family Sarraceniaceae, with species forming tube shaped traps adapted to catch and digest insects. Due to this unique adaptation they are commonly called pitcher plants, although pitcher-shaped carnivorous leaves have evolved convergently in at least two other lineages (*Nepenthes* L., *Cephalotus* Labill.) (Albert et al., 1992). Most *Sarracenia* species occur sympatrically with at least one other species, and all species pairs can produce fertile hybrids (Bell, 1952). Hybrids between sympatric species are frequently observed in nature (Bell, 1952), and population genetics studies using a few microsatellite loci have shown evidence of gene-flow between species at some sites and not at others (Furches et al., 2013; Rentsch and Holland, 2020). The forces maintaining species boundaries are not well known, but it is possible that outbreeding depression is contributing to species coherence in the face of hybridization. *Sarracenia* hybrids exhibit intermediate pitcher morphology which may decrease prey capture efficacy. Another possible factor contributing to the maintenance of species boundaries is asynchronous flowering phenology (Bell, 1952).

Sarracenia diverged from the rest of Sarraceniaceae an estimated 23 MYA, with most of the diversification within *Sarracenia* occurring between 1-3 MYA (Ellison et al., 2012). Given the rapid speciation, significant gene tree discordance is expected due to incomplete lineage sorting (ILS) (Degnan and Rosenberg, 2009). Despite this, Stephens et al. (2015) estimated a multi-species coalescent phylogeny using 199 nuclear genes that resolved most of the species relationships with high support. This study also presented a plastome tree that was highly discordant with the nuclear tree; no species was reciprocally monophyletic. Cytonuclear discordance such as this can be the result of ILS or introgression of the plastid genome, otherwise referred to as chloroplast capture.

Although the plastome phylogeny estimated in Stephens et al. (2015) is relatively well supported, the analysis was limited by the recovery of only 42kbp of plastome sequence limited to the long single copy and short single copy regions of the plastome. To confirm that the extreme cytonuclear discordance observed in the Stephens et al. (2015) phylogenies was not an artifact of a lack of data, we reassembled plastomes from those sequencing reads using an alternative assembly pipeline to recover more sequence. Seventeen additional accessions are added to this analysis. The cause of cytonuclear discordance is formally assessed using a coalescent based simulation approach to distinguish between ILS and chloroplast capture. Additionally, whole plastomes are assembled and gene content evolution is assessed within the context of carnivory.

2 Materials and methods

2.1 Sequence data

Leaf tissue was obtained from 17 individuals in total: 11 accessions were obtained from the Atlanta Botanical Garden's living conservation collection (*S. oreophila*, *S. jonesii*, *S. alata*, *S. alabamensis* and *S. rubra*) and six accessions were obtained from two field sites (*S. rubra subsp. rubra* and *S. rubra subsp. viatorum*). DNA was extracted from silica dried samples using the Qiagen DNeasy Plant Mini Kit. Library prep was performed using the Kapa Biosystems HyperPlus Kit using iTru adapters (Glenn et al., 2019). Libraries were pooled at equal concentrations and enriched for putative single-copy orthologs enrichment using the Angiosperms353 bait set (Johnson et al., 2019). The enriched pool was sequenced on an Illumina NextSeq 500 at the Georgia Genomics and Bioinformatics Core using a High Output 300 cycle flow cell generating 150bp paired-end reads.

In addition, sequencing reads from Stephens et al. (2015) were downloaded from NCBI Short Read Archive. The Stephens et al. data set includes 71 accessions of *Sarracenia* and 4 accessions of outgroups in Sarraceniaceae (*Heliamphora minor* and *Darlingtonia californica*).

2.2 Plastome assembly

All raw reads were trimmed using Trimmomatic (v. 0.39) (Bolger et al., 2014). Both the new data set and the data set obtained from Stephens et al. were sequenced from libraries enriched for targeted nuclear loci. However, the majority of the reads from both data sets are off-target. Stephens et al. reported an average of 1.6% of reads on target, and analysis of the new data set revealed that less than 1% of the reads were on target. The large proportion of off-target reads enable the assembly of the plastome.

Initial *de-novo* plastome assembly was attempted with GetOrganelle (v. 1.7.5.2) (Jin et al., 2020). GetOrganelle often produced two assembly versions differing only in the orientation of the short single copy regions (SSC). SSC orientation was determined by aligning assemblies to the reference plastome (*Clethra* L. *delavayi*, Genbank accession NC_041129) using MUMmer (v. 4.0.0) (Kurtz et al., 2004), and only the assemblies with concordant SSC orientation were retained.

GetOrganelle did not generate complete *de-novo* plastome assemblies from every sample. In these cases, the following reference-based pipeline was used. Reads were aligned to one of the complete *Sarracenia* plastome assemblies using BWA (v. 0.7.17)

(Li et al., 2009). The aligned reads were then extracted and assembled *de-novo* using SPAdes (Bankevich et al., 2012). Afin (https://github.com/afinit/afin) was used to extend the resulting contigs and fuse any contigs with significant overlap. At this stage, assemblies were either mostly complete (1-3 contigs consisting of the large single copy region (LSC), short single copy regions (SSC), and one IR), or they were more fragmented. The mostly complete assemblies were manually pasted together. The IR boundaries were verified by mapping reads to the assemblies and identifying the coordinate where half of the reads spanned the IR and LSC and the other half spanned the IR and SSC.

2.3 Plastome annotation

Complete plastome assemblies were annotated using PGA (Qu et al., 2019). Fragmented assemblies were aligned to one of the complete, PGA annotated plastomes using the Minimap2 (v. 2.17) (Li, 2018) plugin in Geneious. The "transfer annotation" function was used before generating a consensus sequence.

2.4 Alignment and phylogeny estimation

Coding sequences (CDS) from 79 plastid genes were extracted from the annotated assemblies and aligned with MAFFT (v. 7.470) (Katoh and Standley, 2013). All resulting gene alignments were concatenated. Regions of the concatenated alignment that were poorly aligned or had gaps in 50% or more of the samples were filtered out of the gene alignments using Gblocks (v. 0.91b) (Castresana, 2000). A maximum-likelihood phylogeny was estimated from the concatenated gene alignments using IQ-Tree (v. 2.0.6) (Nguyen et al., 2015). 1000 bootstrap replicates were performed using UFBoot (Minh et al., 2013). The GTR + F + R4 substitution model was used.

2.5 Plastome tree simulations

To differentiate between incomplete lineage sorting (ILS) and chloroplast capture, a tree simulation approach similar to Folk et al., 2017 (Folk et al., 2016) was used. Plastome trees under ILS were simulated using the dendropy python package (v. 4.5.2) (Sukumaran and Holder, 2010) with the species tree from Stephens et al. (2015) as a guide tree. Since plastomes are effectively haploid and inherited uniparentally, plastomes have one quarter of the effective population size of diploid nuclear loci. Since the guide tree used for these simulations was estimated exclusively using nuclear loci, its branch lengths were scaled by four to account for the effective population size differential between plastomes and nuclear loci. A distribution of tree discordance under the null hypothesis of ILS was generated by calculating a tree distance metric [information-based generalized Robinson-Foulds distance (Smith, 2020)] between 1000 simulated trees and the species tree. Then the distance between the empirical plastome tree from this study and the species tree was calculated and compared to the null distribution. Since the empirical plastome tree has samples that are not in the Stephens et al. (2015) species tree, those tips were dropped from the plastome tree to enable calculating distance.

3 Results

3.1 Plastome assemblies

Fourteen complete, circularized plastomes have been assembled and annotated including the following *Sarracenia* species: *S. jonesii*, *S. alabamensis*, *S. oreophila*, *S. rubra subsp. gulfensis, S. rubra subsp. rubra*, and *S. rubra subsp. viatorum*. Average assembly statistics for the all assemblies are shown in Table 1. The assembly pipeline for fragmented assemblies recovered an average of 114kbp of plastome sequence, almost tripling the 42kbp recovered in Stephens et al. (2015). The use of different references is one potential factor explaining this difference; this study used a complete *Sarracenia* plastome (Ericales) as a reference whereas Stephens et al. (2015) used a plastome from *Vitis vinifera* (Vitales). Eighty protein-coding genes were extracted from assemblies, and sequences were aligned for all samples, and alignments were concatenated for the phylogeny estimation.

3.2 Pseudogenization of plastome encoded genes

All complete *Sarracenia* plastomes include some pseudogenized plastome-encoded genes. With the exception of *ndhB* and *ndhE*, all *ndh* genes either have been pseudogenized due to premature stop codons or large deletions (Figure 1). Similarly, all samples contain a premature stop codon within the *rps12* gene.

3.3 Plastid phylogeny

Consistent with Stephens et al. (2015), no species were found to exhibit monophyly of their plastomes, and the plastid tree is highly incongruent with the published species tree (Figure 2). Support values across the backbone of the tree are all greater than 70, and most internal nodes are highly supported as well (Figure 2). Branch lengths within *Sarracenia* are generally very short in comparison to the outgroups. An exception is the split at the base of the *Sarracenia* clade. This branch splits *Sarracenia* into two distinct plastid lineages. These main lineages are arbitrarily termed clade A and clade B (Figure 2). Clade B contains all sampled individuals of *minor, oreophila, jonesii*, and *purpurea* var. *montana*, and clade A contains all sampled individuals of *alata* and *purpurea* (excluding var. *montana*). All other species are split across these two main lineages (*flava, psittacina, rubra*, and *leucophylla*).

3.3.1 Southern Appalachian species

S. purpurea var. *montana* and *S. jonesii* form a clade. Both taxa have distributions restricted to a small area in the southern Appalachian Mountains (Figure 3) and hybridize at sympatric

10.3389/fpls.2023.1237749

TABLE 1 Accession information and assembly statistics for all samples used in this study.

Taxon	Sample ID	NCBI Biosample	Collector	Herbarium ID	Total contigs	Total length
Darlingtonia californica						
	DarlingtoniaOR_j028	SAMN03354578	J. D. Stephens	UGA66	36	121605
	DarlingtoniaUN1_j029	SAMN03354579	J. D. Stephens	N/A	65	116346
	DarlingtoniaUN2_j030	SAMN03354580	J. D. Stephens	UGA54	60	112573
Heliamphora minor						
	HeliamphoraVE_j031	SAMN03354581	J. D. Stephens	UGA55	45	130627
S. alabamensis						
	AlabamensisAL_j018	SAMN03354582	J. D. Stephens	UGA19	76	110275
	Alabamensis_m004	SAMN31020169	E. Baldwin	1004	1	154984
S. alata						
	AlataMS1_j033	SAMN03354583	J. D. Stephens	UGA21	51	118063
	AlataMS2_j034	SAMN03354584	J. D. Stephens	N/A	62	117941
	AlataLA1_j035	SAMN03354585	J. D. Stephens	UGA67	25	125730
	AlataTX_j036	SAMN03354586	J. D. Stephens	TAES253951	39	123698
	AlataLA2_j037	SAMN03354587	J. D. Stephens	UGA60	38	126597
	Alata_m003	SAMN36416359	E. Baldwin	1003	29	149266
S. flava						
	FlavaGA_j039	SAMN03354588	J. D. Stephens	UGA15	57	115895
	FlavaFL_j042	SAMN03354589	J. D. Stephens	UGA65	1	153868
	FlavaNC1_j045	SAMN03354590	J. D. Stephens	UGA48	34	125019
	FlavaSC_j046	SAMN03354591	J. D. Stephens	UGA45	18	129423
	FlavaNC2_j047	SAMN03354592	J. D. Stephens	UGA50	22	129654
	FlavaVA_j048	SAMN03354593	J. D. Stephens	UGA64	23	128338
S. flava var. rubricorpora						
	FlavaRubricorpaFL1_j041	SAMN03354594	J. D. Stephens	UGA18	14	132800
	FlavaRubricorpaFL2_j043	SAMN03354595	J. D. Stephens	UGA18	15	128594
S. flava var. rugelii						
	FlavaRugeliiGA1_j038	SAMN03354597	J. D. Stephens	UGA26	52	117830
	FlavaRugeliiGA2_j040	SAMN03354598	J. D. Stephens	UGA44	26	126308
	FlavaRugeliiAL_j044	SAMN03354596	J. D. Stephens	UGA51	8	130660
S. jonesii						
	JonesiiSC1_j023	SAMN03354599	J. D. Stephens	UGA32	24	125346
	JonesiiNC1_j024	SAMN03354600	J. D. Stephens	UGA31	9	127465
	JonesiiNC2_j025	SAMN03354601	J. D. Stephens	UGA33	66	116650
	JonesiiSC2_j026	SAMN03354602	J. D. Stephens	UGA30	53	118729
	JonesiiNC1_m007	SAMN31020170	E. Baldwin	1007	1	151409
	JonesiiSC1_m008	SAMN31020171	E. Baldwin	1008	1	151385

(Continued)

10.3389/fpls.2023.1237749

TABLE 1 Continued

Taxon	Sample ID	NCBI Biosample	Collector	Herbarium ID	Total contigs	Total length
S. minor						
	MinorGA1_j056	SAMN03354609	J. D. Stephens	N/A	23	126435
	MinorGA2_j058	SAMN03354610	J. D. Stephens	UGA8	82	111289
	MinorGA3_j059	SAMN03354611	J. D. Stephens	UGA39	57	117630
	MinorSC1_j060	SAMN03354612	J. D. Stephens	UGA46	20	127646
	MinorSC2_j062	SAMN03354613	J. D. Stephens	UGA13	32	124716
S. minor var. okefenokeensis						
	MinorOkefenokeensisGA_j055	SAMN03354614	J. D. Stephens	UGA23	22	125047
S. oreophila						
	OreophilaAL1_j063	SAMN03354615	J. D. Stephens	UGA2	9	128216
	OreophilaAL2_j064	SAMN03354616	J. D. Stephens	UGA28	32	126491
	OreophilaAL3_j065	SAMN03354617	J. D. Stephens	UGA27	66	95604
	OreophilaNC_j066	SAMN03354618	J. D. Stephens	UGA20	29	124312
	OreophilaAL4_j067	SAMN03354619	J. D. Stephens	UGA24	40	120414
	OreophilaGA_j068	SAMN03354620	J. D. Stephens	UGA22	43	118692
	Oreophila_m002	SAMN31020172	E. Baldwin	1002	1	156118
S. psittacina						
	PsittacinaGA1_j070	SAMN03354621	J. D. Stephens	UGA43	40	121580
	PsittacinaAL1_j072	SAMN03354623	J. D. Stephens	UGA11	38	128169
	PsittacinaGA3_j073	SAMN03354624	J. D. Stephens	UGA10	23	129214
	PsittacinaAL2_j074	SAMN03354625	J. D. Stephens	UGA1	43	122414
	PsittacinaFL_j075	SAMN03354626	J. D. Stephens	UGA35	36	119271
	PsittacinaAL3_j076	SAMN03354627	J. D. Stephens	UGA53	18	128493
	PsittacinaLA_j077	SAMN03354628	J. D. Stephens	UGA59	29	126403
S. purpurea ssp. purpurea						
	PurpureaPurpureaNS_j006	SAMN03354629	J. D. Stephens	UGA61	37	124790
	PurpureaPurpureaWI1_j007	SAMN03354630	J. D. Stephens	UGA47	39	119984
	PurpureaPurpureaWI2_j008	SAMN03354631	J. D. Stephens	UGA47	56	125476
S. purpurea ssp. venosa						
	PurpureaVenosaGA_j001	SAMN03354463	J. D. Stephens	UGA12	28	126463
	PurpureaVenosaNC_j003	SAMN03354632	J. D. Stephens	UGA49	33	124039
	PurpureaVenosaMD_j004	SAMN03354633	J. D. Stephens	UGA62	61	121892
	PurpureaVenosaVA_j005	SAMN03354634	J. D. Stephens	UGA63	48	123570
S. purpurea ssp. venosa var. montana						
	PurpureaMontanaGA_j078	SAMN03354636	J. D. Stephens	UGA41	31	126220
	PurpureaMontanaNC_j079	SAMN03354635	J. D. Stephens	UGA34	39	118268
S. rosea (S. purpurea ssp. venosa var. burkii)						
	RoseaFL2_j002	SAMN03354640	J. D. Stephens	UGA5	54	121220

(Continued)

10.3389/fpls.2023.1237749

TABLE 1 Continued

Taxon	Sample ID	NCBI Biosample	Collector	Herbarium ID	Total contigs	Total length
	RoseaFL1_j009	SAMN03354637	J. D. Stephens	UGA16	83	105680
	RoseaAL_j010	SAMN03354638	J. D. Stephens	UGA4	51	122458
	RoseaMS_j080	SAMN03354639	J. D. Stephens	UGA7	28	127190
S. rubra						
	RubraGA1_j011	SAMN03354641	J. D. Stephens	UGA42	35	123685
	RubraGA2_j012	SAMN03354642	J. D. Stephens	UGA58	17	130198
	RubraGA3_j013	SAMN03354643	J. D. Stephens	UGA37	34	126171
	RubraGA4_j014	SAMN03354644	J. D. Stephens	UGA36	22	128963
	RubraGA5_j015	SAMN03354645	J. D. Stephens	UGA36	37	124041
	RubraGA6_j016	SAMN03354646	J. D. Stephens	UGA14	45	114813
	RubraSC_j017	SAMN03354661	J. D. Stephens	N/A	1	154655
	RubraSC_m001	SAMN31020178	E. Baldwin	1001	1	155181
	Rubra1_m005	SAMN31020173	E. Baldwin	1005	1	155212
	Rubra2_m006	SAMN36416360	E. Baldwin	1006	8	128073
S. rubra ssp. gulfensis						
	RubraGulfensisFL1_j020	SAMN03354647	J. D. Stephens	UGA3	65	115074
	RubraGulfensisFL2_j021	SAMN03354648	J. D. Stephens	UGA29	59	109430
	RubraGulfensisFL3_j022	SAMN03354649	J. D. Stephens	UGA25	23	125074
	RubraGulfensisFL1_m009	SAMN31020174	E. Baldwin	1009	1	154989
	RubraGulfensisFL2_m010	SAMN36416361	E. Baldwin	1010	9	127752
	RubraGulfensisFL3_m011	SAMN31020175	E. Baldwin	1011	1	154974
S. rubra ssp. Rubra						
	RubraRubraNC1_m012	SAMN31020176	E. Baldwin	1012	1	155283
	RubraRubraNC2_m013	SAMN31020177	E. Baldwin	1013	1	155302
S. rubra ssp. viatorum						
	RubraViatorumGA1_m014	SAMN36416362	E. Baldwin	1014	4	128950
	RubraViatorumGA2_m015	SAMN36416363	E. Baldwin	1015	13	132536
	RubraViatorumGA3_m016	SAMN31020179	E. Baldwin	1016	1	155157
	RubraViatorumGA4_m017	SAMN31020180	E. Baldwin	1017	1	155185
S. rubra ssp. wherryi						
	RubraWherryiAL_j027	SAMN03354650	J. D. Stephens	UGA38	38	123432
S.leucophylla						
	LeucophyllaFL1_j049	SAMN03354603	J. D. Stephens	UGA57	11	129896
	LeucophyllaAL1_j050	SAMN03354604	J. D. Stephens	UGA40	19	127213
	LeucophyllaGA_j051	SAMN03354605	J. D. Stephens	UGA17	19	129788
	LeucophyllaFL2_j052	SAMN03354606	J. D. Stephens	UGA56	24	132845
	LeucophyllaAL2_j053	SAMN03354607	J. D. Stephens	UGA52	12	132508
	LeucophyllaFL3_j054	SAMN03354608	J. D. Stephens	UGA6	20	126675

sites. The only other species found in the southern Appalachians is *S. oreophila*, although it is not sympatric with *S. jonesii* or *S. purpurea* var. *montana*, but may have been historically (McPherson and Schnell, 2011). Two *S. oreophila* accessions from Alabama are sister to the Appalachian clade, and the other *S. oreophila* accessions are placed in a clade sister to this.

3.3.2 *Sarracenia flava*, *S. minor*, and *S. psittacina*

S. flava, *S. minor*, and *S. psittacina* form a clade sister to *S. purpurea* on the species tree, however the placement of these species on the plastid tree is not congruent. All *S. minor* accessions are placed within clade B sister to the clade containing *S. oreophila*, *S. jonesii*, and *S. purpurea* var. *montana*. Some *S. flava* and *S. psittacina* accessions from the Gulf coastal plain are also placed in the *S. minor* clade, despite all *S. minor* accessions in this study originating from the Atlantic coastal plain. This could indicate either ancient introgression or retention of plastome diversity from the ancestor of these three species. *S. flava* and *S. psittacina* are scattered across the chloroplast phylogeny; both species have accessions found in clades A and B. In *S. flava*, all Gulf coastal plain accessions are found in clade B and all Atlantic coastal plain accessions are found in clade A.

3.3.3 *Sarracenia purpurea* complex

With the exception of *S. purpurea* var. *montana*, all *S. purpurea* accessions (including *S. rosea*) are placed in clade A. There is no discernible pattern to their placement within this lineage. This is surprising given the vast geographic range represented by these taxa; the individuals sampled for this study originate from throughout their distribution from Mississippi to Nova Scotia. Only *S. purpurea subsp. purpurea* is found north of Maryland, so the relatedness of plastomes between this taxon and other species are unlikely to be the result of recent introgression.

3.4 Plastome phylogeny simulations

The tree distance metric that was used ranges from 0 (an identical tree) to 1 (the most distal tree). The plastome trees simulated under the pure coalescent model have distances from the species tree ranging from 0.29 to 0.56, while the distance from the empirical plastome tree is 0.73 (Figure 4). A T-test using the distribution of simulated plastome tree distances as the null distribution gives a p-value of >2.2e-16, rejecting the null hypothesis of ILS causing the discordance alone.

4 Discussion

4.1 Pseudogenization of ndh genes

Independent pseudogenization or complete loss of *ndh* genes has been shown in many plant lineages, including holoparasitic, hemiparastitic, and carnivorous plant lineages (Barrett et al., 2014; Lin et al., 2017; Cao et al., 2019; Gruzdev et al., 2019; Nevill et al., 2019). Functional *ndh* genes are rarely found in non-photosynthetic parasitic plants, and the loss of *ndh* genes is strongly correlated with the transition to heterotrophy in parasitic plant lineages (Wicke et al., 2016). Since plastid encoded *ndh* genes are thought to optimize photosynthetic chemistry in fluctuating or stressful environments [reviewed in (Sabater, 2021)], the loss of *ndh* genes in parasitic lineages that are no longer fully dependent on photosynthesis as a source of carbon is unsurprising. In carnivorous plants, however, evidence for significant heterotrophic uptake of carbon is limited (Rischer et al., 2002), and a transition to full heterotrophy seems unlikely, so this line of reasoning does not explain the independent pseudogenization of functional *ndh* genes across carnivorous plant lineages. It is possible that the acquisition of organic nitrogen has an interaction with photosynthetic chemistry that relaxes the need for *ndh*. As Nevill et al. (2019) noted, organic nitrogen acquisition

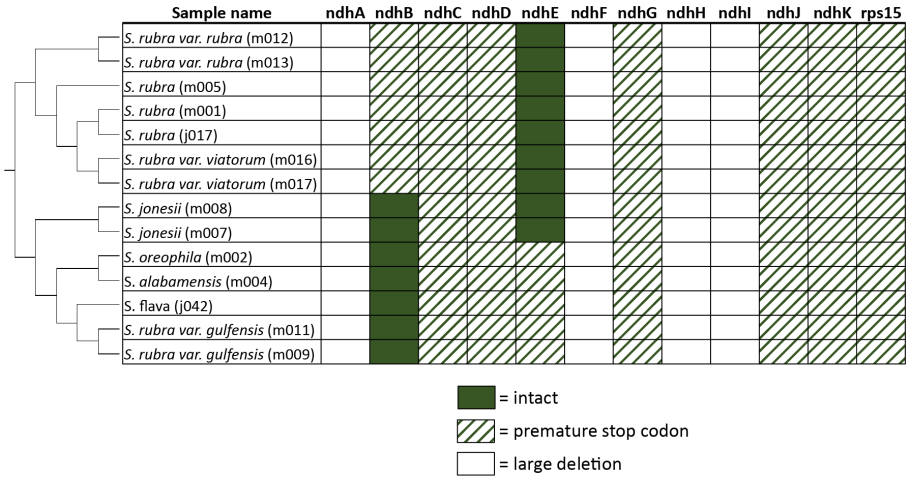


FIGURE 1
Status of *ndh* genes in all complete plastome assemblies. Filled cells represent intact genes, dashed cells represent premature stop codons, and blank cells represent genes with large deletions. Plastome tree trimmed from these samples is shown on the left.

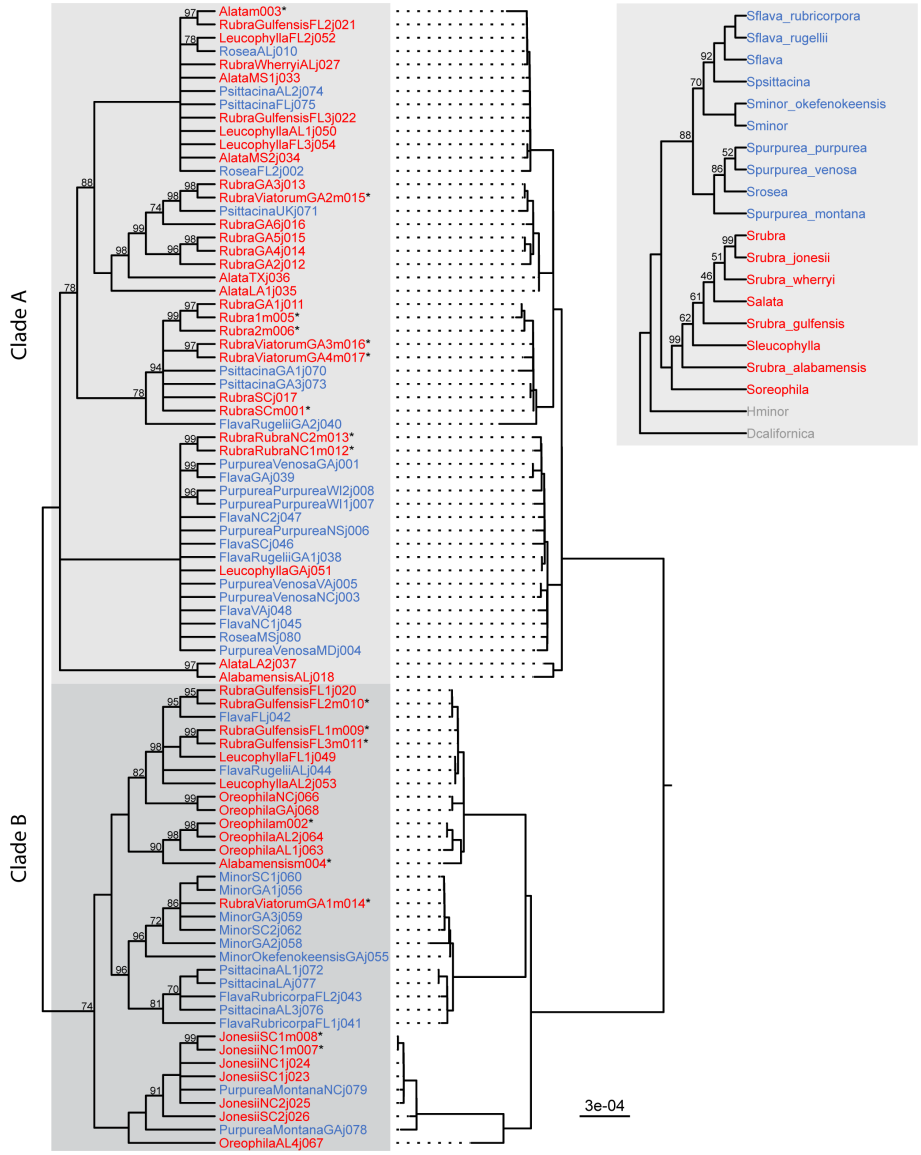


FIGURE 2
Maximum likelihood plastome cladogram (left) and phylogram (center) and species tree (inset, top right) from Stephens et al. (2015). Nodes on the cladogram with bootstrap values less than 70 are collapsed. Uncollapsed cladogram nodes with bootstrap values less than 100 are labelled. Tip names are either red or blue based on which of the two major clades the species belongs to in the species tree. Asterisks next to tip labels indicate samples that were newly sequenced for this study.

bypasses the need to assimilate nitrate using photosynthetically-derived reductant. Alternatively, the pseudogenization of *ndh* genes in parasitic plants and carnivorous plants could be due to unrelated mechanisms. The pseudogenization of almost all of the *ndh* genes across the genus *Sarracenia* shown here provides further evidence that carnivorous plants do not require these genes. Sequencing of full plastomes from other carnivorous species would reveal if the pseudogenization of *ndh* occurs early in carnivorous plant evolution.

4.2 Cytonuclear discordance

The plastome phylogeny in this study shows a similarly extreme level of discordance with the species tree as that of the Stephens et al. (2015) plastome phylogeny. That study ascribed the discordance to a combination of chloroplast capture and a lack of informative polymorphisms in the chloroplast sequence. A third source of discordance, ILS, is considered here. A lack of informative polymorphisms is not an issue here, as almost all the plastome coding sequences are used and the resulting phylogeny has high bootstrap values across the spine, suggesting that there is sufficient evidence that major clades within the tree are correct.

To distinguish between the two remaining sources of discordance, plastome phylogenies under ILS were simulated. The simulated phylogenies showed much lower levels of discordance with the species tree than the empirically estimated plastome. To simulate the plastome phylogenies, the branch lengths of the guide tree were multiplied by four due to the assumption that the

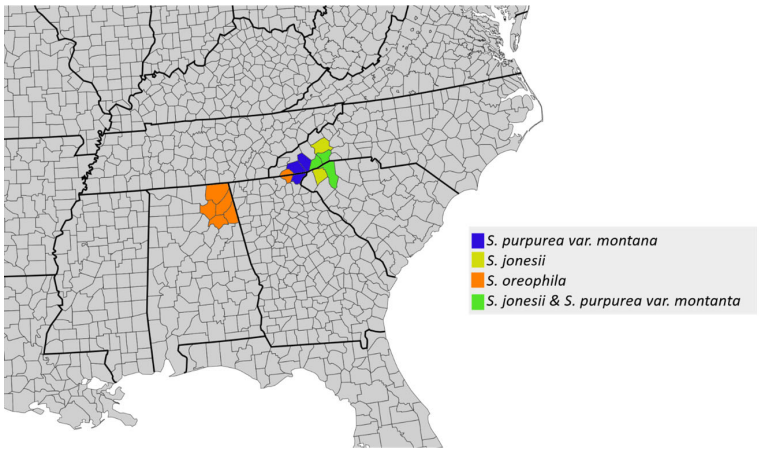


FIGURE 3
A county level distribution map for *S. purpurea var montana*, *S. jonesii*, and *S. oreophila*.

chloroplast is inherited matrilineally in *Sarracenia* like most seed plants (Mogensen, 1996). Since this assumption hasn't been empirically proven and biparental inheritance of the chloroplast is possible, simulations with branch lengths multiplied by two were performed and show similar results (Supplemental Data).

There is ample signal of introgression in the plastome, but Stephens et al. (2015) reported no evidence of gene flow in the nuclear data. A search through the nuclear gene trees revealed that none of the trees had a similar topology to the plastome tree, but some trees did exhibit a high degree of discordance with the species tree, possibly due to occasional nuclear gene introgression. Cytonuclear discordance is commonly observed and is attributed to introgression in plant and animal systems (Rieseberg and Soltis, 1991; Berthier et al., 2006; Gernandt et al., 2018), including several instances where there is limited signal for introgression in nuclear data (Winkler et al., 2013; Good et al., 2015; Folk et al., 2016; Rose et al., 2020). However, the mechanism for organellar introgressions without accompanying nuclear loci is poorly understood (Rieseberg and Soltis, 1991; Folk et al., 2018). *Sarracenia* is a genus where hybridization is common and thus some level of nuclear introgression might be expected. The extreme level of chloroplast capture and lack of signal for nuclear gene flow in *Sarracenia* illustrates the comparative ease of introgression of organelles over nuclear loci.

4.2.1 Geographic patterns of plastome introgression

Although the lack of monophyletic species in the plastome tree makes it difficult to interpret specific instances of plastome introgression, a handful of such instances can be elucidated using geographic context. For example, all accessions of *S. purpurea* var. *montana* and *S. jonesii*, two taxa restricted to a small region in the southern Appalachians, form a well-supported clade within clade B.

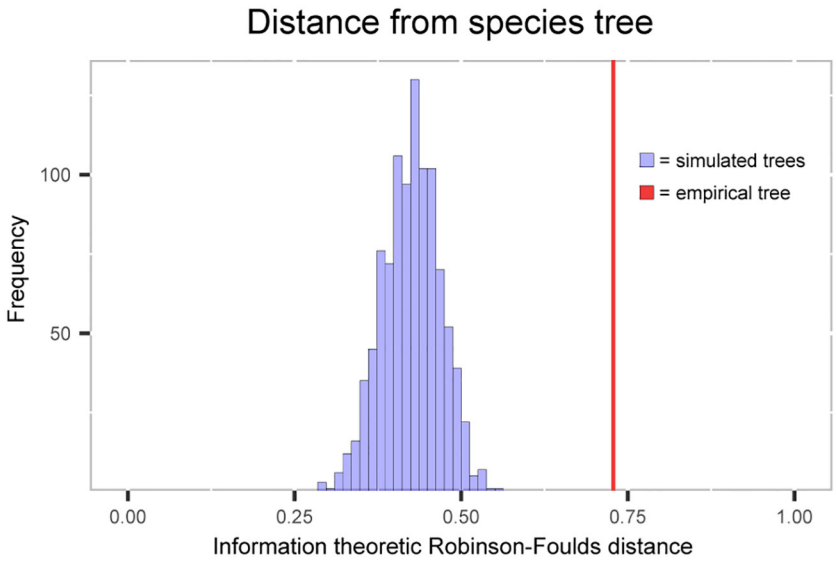


FIGURE 4
Histogram of information-based generalized Robinson-Foulds distance between the simulated plastome trees and the species tree. Red line shows the distance between empirically estimated plastome tree and the species tree.

Given that all other *S. purpurea* accessions are placed in clade A, it is likely that a plastome derived from *S. jonesii* was introgressed into *S. purpurea* var. *montana*. Similarly, an accession of *S. rubra* that was sampled from the Georgia fall line near *S. minor* populations is placed within the *S. minor* clade. Again, we hypothesize this to be an instance of *S. minor* plastome being introgressed into *S. rubra*. More generally, the weak species clustering in the plastome tree implies a long history of interspecific exchange of cytoplasmic genomes in *Sarracenia*.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/bioproject/; PRJNA884359.

Author contributions

EB, MM, JL-M contributed to study design. EB performed all data analysis and wrote initial manuscript draft. MM collected samples and generated data. MM, JL-M contributed to refinement of manuscript and approved final submission. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by an National Science Foundation grant (DEB 2110875) to JL-M.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1237749/full#supplementary-material

SUPPLEMENTARY FIGURE 1
png - Histograms of information-based generalized Robinson-Foulds distance between the simulated plastome trees and the species tree. Panel titles indicate how the branches of the guide tree were scaled when plastome trees were simulated. Red line shows the distance between empirically estimated plastome tree and the species tree.

alignment.fasta
Concatenated gene alignments used to estimate plastome tree.

Plastome.newick
Plastome tree in newick format.

Simtrees.1.newick
Plastome trees simulated with unscaled guide tree.

Simtrees.2.newick
Plastome trees simulated with guide tree branch lengths scaled by two.

Simtrees.4.newick
Plastome trees simulated with guide tree branch lengths scaled by four.

Fragmented_assemblies.zip
Zipped folder containing all fragmented assemblies in fasta format.

References

Albert, V. A., Williams, S. E., and Chase, M. W. (1992). Carnivorous plants: phylogeny and structural evolution. *Science* 257 (5076), 1491–1495. doi: 10.1126/science.1523408

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19 (5), 455–477. doi: 10.1089/cmb.2012.0021

Barrett, C. F., Freudenstein, J. V., Li, J., Mayfield-Jones, D. R., Perez, L., Pires, J. C., et al. (2014). Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Mol. Biol. Evol.* 31 (12), 3095–3112. doi: 10.1093/molbev/msu252

Bell, C. R. (1952). Natural hybrids in the genus sarracenia: I. *Hist. Distribut. Taxonomy. J. Elisha Mitchell Sci. Soc.* 68 (1), 55–80. Available at: https://www.jstor.org/stable/24334305.

Berthier, P., Excoffier, L., and Ruedi, M. (2006). Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species Myotis myotis and Myotis blythii. *Proc. R. Soc. B: Biol. Sci.* 273 (1605), 3101–3123. doi: 10.1098/rspb.2006.3680

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Cao, M., Li, Z., Dai, X., Wu, X., Li, Y., and Wu, S. (2019). The complete plastid genome of carnivorous pitcher plant Cephalotus follicularis. *Mitochondrial DNA Part B.* 4 (1), 2025–2027. doi: 10.1080/23802359.2019.1617054

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17 (4), 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Degnan, J. H., and Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 24 (6), 332–340. doi: 10.1016/j.tree.2009.01.009

Ellison, A. M., Butler, E. D., Hicks, E. J., Naczi, R. F. C., Calie, P. J., Bell, C. D., et al. (2012). Phylogeny and biogeography of the carnivorous plant family Sarraceniaceae. *PloS One* 7 (6), e39291–e3929e. doi: 10.1371/journal.pone.0039291

Folk, R. A., Mandel, J. R., and Freudenstein, J. V. (2016). Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Systemat. Biol.* 66 (3), 320–337. doi: 10.1093/sysbio/syw083

Folk, R. A., Soltis, P. S., Soltis, D. E., and Guralnick, R. (2018). New prospects in the detection and comparative analysis of hybridization in the tree of life. *Am. J. Botany.* 105 (3), 364–375. doi: 10.1002/ajb2.1018

Furches, M. S., Small, R. L., and Furches, A. (2013). Hybridization leads to interspecific gene flow in Sarracenia (Sarraceniaceae). *Am. J. Bot.* 100 (10), 2085–2091. doi: 10.3732/ajb.1300038

Gernandt, D. S., Aguirre Dugua, X., Vázquez-Lobo, A., Willyard, A., Moreno Letelier, A., Pérez de la Rosa, J. A., et al. (2018). Multi-locus phylogenetics, lineage sorting, and reticulation in Pinus subsection Australes. *Am. J. Bot.* 105 (4), 711–725. doi: 10.1002/ajb2.1052

Glenn, T. C., Nilsen, R. A., Kieran, T. J., Sanders, J. G., Bayona-Vásquez, N. J., Finger, J. W., et al. (2019). Adapterama I: universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). *PeerJ* 7, e7755. doi: 10.7717/peerj.7755

Good, J. M., Vanderpool, D., Keeble, S., and Bi, K. (2015). Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. *Evolution* 69 (8), 1961–1972. doi: 10.1111/evo.12712

Grant, P. R., and Grant, B. R. (2019). Hybridization increases population variation during adaptive radiation. *Proc. Natl. Acad. Sci.* 116 (46), 23216–23224. doi: 10.1073/pnas.1913534116

Gruzdev, E. V., Kadnikov, V. V., Beletsky, A. V., Kochieva, E. Z., Mardanov, A. V., Skryabin, K. G., et al. (2019). Plastid Genomes of Carnivorous Plants Drosera rotundifolia and Nepenthes × ventrata Reveal Evolutionary Patterns Resembling Those Observed in Parasitic Plants. *Int. J. Mol. Sci.* 20 (17), 4107. doi: 10.3390/ijms20174107

Jin, J.-J., Yu, W.-B., Yang, J.-B., Song, Y., dePamphilis, C. W., Yi, T.-S., et al. (2020). GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biol.* 21 (1), 241. doi: 10.1186/s13059-020-02154-5

Johnson, M. G., Pokorny, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., et al. (2019). A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Syst. Biol.* 68 (4), 594–606. doi: 10.1093/sysbio/syy086

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30 (4), 772–780. doi: 10.1093/molbev/mst010

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5 (2), R12. doi: 10.1186/gb-2004-5-2-r12

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34 (18), 3094–3100. doi: 10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Lin, C.-S., Chen, J. J. W., Chiu, C.-C., Hsiao, H. C. W., Yang, C.-J., Jin, X.-H., et al. (2017). Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. *Plant J.* 90 (5), 994–1006. doi: 10.1111/tpj.13525

McPherson, S., and Schnell, D. E. (2011). *Sarraceniaceae of North America: redfern natural history productions*.

Meier, J. I., Stelkens, R. B., Joyce, D. A., Mwaiko, S., Phiri, N., Schliewen, U. K., et al. (2019). The coincidence of ecological opportunity with hybridization explains rapid adaptive radiation in Lake Mweru cichlid fishes. *Nat. Commun.* 10 (1), 5391. doi: 10.1038/s41467-019-13278-z

Minh, B. Q., Nguyen, M. A., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30 (5), 1188–1195. doi: 10.1093/molbev/mst024

Mogensen, H. L. (1996). The hows and whys of cytoplasmic inheritance in seed plants. *Am. J. Botany*. 83 (3), 383–404. doi: 10.1002/j.1537-2197.1996.tb12718.x

Nevill, P. G., Howell, K. A., Cross, A. T., Williams, A. V., Zhong, X., Tonti-Filippini, J., et al. (2019). Plastome-wide rearrangements and gene losses in carnivorous droseraceae. *Genome Biol. Evol.* 11 (2), 472–485. doi: 10.1093/gbe/evz005

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32 (1), 268–274. doi: 10.1093/molbev/msu300

Pease, J. B., Haak, D. C., Hahn, M. W., and Moyle, L. C. (2016). Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PloS Biol.* 14 (2), e1002379. doi: 10.1371/journal.pbio.1002379

Qu, X. J., Moore, M. J., Li, D. Z., and Yi, T. S. (2019). PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* 15, 50. doi: 10.1186/s13007-019-0435-7

Rentsch, J. D., and Holland, R. C. (2020). Population Genetic Structure and Natural Establishment of Hybrids Between Sarracenia flava and Sarracenia minor in Francis Marion National Forest. *Castanea* 85 (1), 108–21, 14. doi: 10.2179/0008-7475.85.1.108

Rieseberg, L. H., and Soltis, D. (1991). Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends Plants* 5, 65–84. Available at: https://www.researchgate.net/publication/262005952_Phylogenetic_consequences_of_cytoplasmic_gene_flow_in_plants.

Rischer, H., Hamm, A., and Bringmann, G. (2002). Nepenthes insignis uses a C2-portion of the carbon skeleton of l-alanine acquired via its carnivorous organs, to build up the allelochemical plumbagin. *Phytochemistry* 59 (6), 603–609. doi: 10.1016/S0031-9422(02)00003-1

Rose, J. P., Toledo, C. A. P., Lemmon, E. M., Lemmon, A. R., and Sytsma, K. J. (2020). Out of sight, out of mind: widespread nuclear and plastid-nuclear discordance in the flowering plant genus polemonium (Polemoniaceae) suggests widespread historical gene flow despite limited nuclear signal. *Systemat. Biol.* 70 (1), 162–180. doi: 10.1093/sysbio/syaa049

Sabater, B. (2021). On the edge of dispensability, the chloroplast ndh genes. *Int. J. Mol. Sci.* 22 (22), 12505. doi: 10.3390/ijms222212505

Smith, M. R. (2020). Information theoretic generalized Robinson–Foulds metrics for comparing phylogenetic trees. *Bioinformatics* 36 (20), 5007–5013. doi: 10.1093/bioinformatics/btaa614

Stephens, J. D., Rogers, W. L., Heyduk, K., Cruse-Sanders, J. M., Determann, R. O., Glenn, T. C., et al. (2015). Resolving phylogenetic relationships of the recently radiated carnivorous plant genus Sarracenia using target enrichment. *Mol. Phylogenet. Evol.* 85, 76–87. doi: 10.1016/j.ympev.2015.01.015

Sukumaran, J., and Holder, M. T. (2010). DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26 (12), 1569–1571. doi: 10.1093/bioinformatics/btq228

Wicke, S., Müller, K. F., dePamphilis, C. W., Quandt, D., Bellot, S., and Schneeweiss, G. M. (2016). Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proc. Natl. Acad. Sci. U. S. A.* 113 (32), 9045–9050. doi: 10.1073/pnas.1607576113

Winkler, M., Tribsch, A., Schneeweiss, G. M., Brodbeck, S., Gugerli, F., Holderegger, R., et al. (2013). Strong nuclear differentiation contrasts with widespread sharing of plastid DNA haplotypes across taxa in European purple saxifrages (Saxifraga section Porphyrion subsection Oppositifoliae). *Botan. J. Linn. Soc.* 173 (4), 622–636. doi: 10.1111/boj.12104

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 October 2024
DOI 10.3389/fpls.2024.1445713

OPEN ACCESS

EDITED BY
Mohamed Mannaa,
Pusan National University, Republic of Korea

REVIEWED BY
Henda Mahmoudi,
International Center for Biosaline Agriculture (ICBA), United Arab Emirates
Magdalena Wróbel-Kwiatkowska,
Wroclaw University of Environmental and Life Sciences, Poland

*CORRESPONDENCE
Magdy Alabady
malabady@uga.edu

†Deceased

RECEIVED 14 June 2024
ACCEPTED 30 September 2024
PUBLISHED 15 October 2024

CITATION
Cai J, Mohsin I, Rogers W, Zhang M, Jiang L, Malmberg R and Alabady M (2024) The microbiome and metatranscriptome of a panel from the *Sarracenia* mapping population reveal complex assembly and function involving host influence. *Front. Plant Sci.* 15:1445713. doi: 10.3389/fpls.2024.1445713

The microbiome and metatranscriptome of a panel from the *Sarracenia* mapping population reveal complex assembly and function involving host influence

Jiazhang Cai[1], Iqra Mohsin[2], Willie Rogers[3], Mengrui Zhang[4], Lin Jiang[5], Russell Malmberg[3†] and Magdy Alabady[3*]

[1]Department of Statistics, University of Georgia, Athens, GA, United States, [2]Department of Biology, University of Georgia, Athens, GA, United States, [3]Department of Plant Biology, University of Georgia, Athens, GA, United States, [4]Quantitative Sciences Unit, Department of Medicine, Stanford University, Stanford, CA, United States, [5]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, United States

Sarracenia provide an optimal system for deciphering the host-microbiome interactions at various levels. We analyzed the pitcher microbiomes and metatranscriptomes of the parental species, and F1 and F2 generations from the mapping population (*Sarracenia purpurea* X *Sarracenia psittacina*) utilizing high-throughput sequencing methods. This study aimed to examine the host influences on the microbiome structure and function and to identify the key microbiome traits. Our quality datasets included 8,892,553 full-length bacterial 16s rRNA gene sequences and 65,578 assembled metatranscripts with microbial protein annotations. The correlation network of the bacterial microbiome revealed the presence of 3-7 distinct community clusters, with 8 hub and 19 connector genera. The entire microbiome consisted of viruses, bacterial, archaea, and fungi. The richness and diversity of the microbiome varied among the parental species and offspring genotypes despite being under the same greenhouse environmental conditions. We have discovered certain microbial taxa that are genotype-enriched, including the community hub and connector genera. Nevertheless, there were no significant differences observed in the functional enrichment analysis of the metatranscriptomes across the different genotypes, suggesting a functional convergence of the microbiome. We found that the pitcher microcosm harbors both rhizosphere and phyllosphere microbiomes within its boundaries, resulting in a structurally diverse and functionally complex microbiome community. A total of 50,424 microbial metatranscripts were linked to plant growth-promoting microbial proteins. We show that this complex pitcher microbiome possesses various functions that contribute to plant growth promotion, such as biofertilization, bioremediation, phytohormone signaling, stress regulation, and immune response stimulation. Additionally, the pitcher microbiome exhibits traits related to microbe-microbe interactions, such as colonization of plant systems, biofilm formation, and

microbial competitive exclusion. In summary, the demonstrated taxonomical divergence and functionally convergence of the pitcher microbiome are impacted by the host genetics, making it an excellent system for discovering novel beneficial microbiome traits.

KEYWORDS

carnivorous plants, pitcher plant, sarracenia, microbiome, metatranscriptome, plant-microbiome interaction, structural and functional core microbiome

Introduction

Plants live in association with a wide range of microorganisms, including bacteria, fungi, protists, nematodes, and viruses, all referred to as the plant microbiome. In natural habitats, through this association, microorganisms play a significant role in enhancing plant growth, productivity, and health (Trivedi et al., 2020). Experimental evidence has demonstrated that plants have the ability to modify the abundance and diversity of microbiomes in their environments (Mahnert et al., 2015). Nevertheless, the precise mechanisms through which this occurs remain unidentified. Plants may have acquired genetic traits that impact their interaction with their microbiomes. The identification of these traits and their mechanisms are essential for the success of the ongoing efforts to harness and manipulate microbiomes in order to optimize their advantages for the host plants (Ke et al., 2021; Afridi et al., 2022; Morales Moreira et al., 2023).

Carnivory evolved independently on 13 occasions in flowering plants, which includes four instances in monocots (three in Poales and one in Alismatales) and nine instances in eudicots (three in Caryophyllales, three in Lamiales, two in Ericales, and one in Oxalidales). There is a total of 810 species belonging to 21 plant genera, which are distributed across 13 families and six orders (reviewed in (Fu et al., 2023)). As a result of convergent evolution, some carnivorous species have developed modified leaves that allow them to capture and kill insect prey. These species also host a complex microbiome that aids in the digestion of the captured prey and support plant growth through an array of direct and indirect traits. These traits might be considered as extensions of the host plant traits, which are equally significant to the inherent plant traits (the concept is reviewed in (Whitham et al., 2006, 2008)). The "pitfall" carnivorous plants evolved to trap insects in three separate lineages (Albert et al., 1992). Using modified leaves called 'pitchers', they digest insect prey with the aid of a complex microbiome that develops inside their pitcher's fluid. In the genus *Sarracenia*, species differ in their pitcher morphology, which in turn affects the insect species they capture for nutrients (Ellison et al., 2003a, 2004), and also affects the initial assembly of the pitcher microbiomes. The factors involved in shaping the microbiome community likely involve host plant genetics, local environmental conditions, as well as prey's microbiomes. The interactions between the *Sarracenia*, its microbiome, and captured prey likely played a role in driving the process of plant speciation. In return, the host species genetics may have evolved traits to impact the composition and function of their microbiome communities. The host impact can be measured by examining the taxonomic divergence and functional convergence of the pitcher microbiome, which are two prevalent phenomena in microbial systems (Louca et al., 2016, 2018).

The bacteria, archaea and eukaryotes that reside in the *Sarracenia* pitchers have been described by organismal methods (Ellison et al., 2003b; Baiser et al., 2013) and by environmental sequencing (Adlassnig et al., 2011; Boyer and Carter, 2011; Koopman and Carstens, 2011). Multiple studies have explored the impact of the *Sarracenia* host plant on the microbiome assembly by comparing the microbial communities in natural and artificial pitchers (Koopman et al., 2010; Bittleston et al., 2018; Ellison et al., 2021; Grothjan and Young, 2022), as well as among various *Sarracenia* species (Heil et al., 2022). These studies concluded that *Sarracenia* may play an active role in shaping the structure of their microbiome. However, is this impact is regulated in part by the genetic makeup of the host (i.e., deterministic impact)? To address this question, we are using a *Sarracenia* mapping population in a controlled environment to characterize the microbiome diversity, structure and function among genotypes of same genetic background.

Malmberg et al. developed a *Sarracenia* genetic linkage map of 437 SNP and SSR markers with a total length of 2017 cM using an F2 generation of 280 plants from a genetic cross between *Sarracenia purpurea* (*Spu*) and *Sarracenia psittacina (Spa)* (Malmberg et al., 2018). They mapped 64 pitcher QTLs traits that were involved in prey attraction, capture and kill, and digestive processes/mechanisms. Despite the fact that the morphological characteristics of these traits are indicative of their adaptation to prey capture, their impact on the structure and function of the microbial community remains uncertain. We hypothesize that *Sarracenia* evolved genetic features for interacting with their vital microbiomes in a manner similar to the convergent evolution observed in their leaf structure.

Pitcher plants are ideal for studying host-microbiome and microbiome-microbiome interactions at the experimental and molecular levels for several reasons. Firstly, each fluid-containing pitcher is a natural microcosm with well-defined boundaries, making it experimentally amenable. This differs from other symbiotic microbial systems that are less amenable to

experimentation, such as gut microbiomes, or those that do not have well-defined boundaries such as plant rhizosphere and phyllosphere microbiomes. Secondly, different species within the *Sarracenia* genus have varying leaf traits associated with differing microbial communities, that attract and digest different species of insects (Grothjan and Young, 2019). These genetically determined characteristics may have allowed pitcher plants to impose control over the microbiome assembly in their pitchers. Thirdly, the various species of *Sarracenia* can hybridize readily, offering a unique opportunity to incorporate a genetic approach to examine plant genome influences on microbiome assembly (Furches et al., 2013). Although pitcher plants have long attracted biologists, we know little about the assembly mechanisms involved in the microbiome assembly and structure in their pitchers (Grothjan and Young, 2019). It has been suggested that factors other than prey capture and colonization by eukaryotic species may affect the recruitment of bacteria to their pitchers (Grothjan and Young, 2019).

The objectives of this study are to evaluate the variations in the microbiome structure and function among different genotypes of the *Sarracenia* mapping populations, namely *Spu*, *Sps*, F1 and F2, and to characterize the composition and functions of the *Sarracenia* microbiome. For a panel of these genotypes, we utilized the single molecule (Oxford Nanopore Technology, ONT) and short reads (Illumina) technologies to sequence the full-length 16S rRNA gene (microbiome) and total mRNA (metatranscriptome), respectively, from their pitcher fluids (Figure 1A). We employed a variety of specialized databases, as well as bioinformatical tools to analyze the data (Figures 1B–D).

Materials and Methods

Genotypes

The samples of parental species are true biological replicates because they are clones of the same individuals, but the samples from the F1 and F2 generations are separate individuals. To simplify the study, we treated the F1 individuals as biological replicates of each other and referred to them as "F1". Similarly, the F2 individuals were treated as biological replicates and referred to them as "F2". Comprehensive information regarding this mapping population is published in our earlier paper (Malmberg et al., 2018). All plants were grown under the same conditions in the greenhouse. Briefly: plants were grown in 4-6-inch pots on flats without holes and filled with 1/2 to 3/4 inch of water to simulate a tiny bog. The plants are top watered daily in the summer and less often in the autumn, winter, and early spring. The bottom inch of soil was immersed in standing water in the flats, keeping it saturated while the top 3 inches were moist but never saturated.

Collecting *Sarracenia* pitcher fluid

We collected the fluid from the pitchers of *Spu*, *Sps*, F1 and F2 plants (Figure 1A). All pitchers were at least 8 weeks old to ensure stable microbiome communities. The pitcher fluid was collected using sterile glass pipettes. For *Spu*, F1 and F2 plants, the glass pipettes were easily inserted into the pitchers as the flap is open. The *Sps* plants have a narrow body and a closed flap, which required a small incision in the flap to insert the glass pipette into the pitcher (Figure 1A). Due to the morphology of *Sps*, some pitchers were dry or contained a small amount of fluid. In these cases, sterilized water was added to elute the microbiome that is attached to pitcher's internal surface. The fluid samples were collected into 50 ml conical sterilized tubes and stored in a -20 °C freezer.

Sample preparation

We thawed the pitcher fluid samples on ice for 30-40 minutes, followed by filtration using the Steriflip Sterile Centrifuge Tube Top Filter Unit (Cat# SCNY00020) to remove any plant material and dead insect matter. A schematic diagram of the sample preparation is shown in Figure 1B. The filtered fluid was a cloudy and homogenous mixture, indicating the presence of microbial growth. We centrifuged the samples at 3000 rpm at 4°C for 20 minutes to precipitate the microbiota, discarded the supernatant, suspended the pellet in 450 µl of PBS buffer, and mixed by pipetting. We added 750 µl of DNA/RNA Shield (Cat# R2002) to the suspension and mixed by pipetting and inverting. The material was transferred into a bead-bashing lysis tube and battered at high frequency for 2 minutes. Next, the samples were then centrifuged for 30 seconds at 16,000 rpm, divided into 400 µl aliquots in 2 ml tubes with 2 volumes of the DNA lysis buffer added to each tube. Finally, we used the ZymoBIOMICS DNA/RNA Miniprep Kit (Cat# R2002) to isolate both DNA and RNA.

Microbiome DNA isolation

We transferred the entire sample into a SpinAway Yellow Filter unit, followed by centrifugation at 16,000 rpm for 30 seconds. The flow-through, which contained the RNA, was saved in a 15 ml tube. An equal volume of 100% ethanol was added, and the tube was placed on ice. After processing the entire sample, a new collecting tube was used and 400 µl of DNA/RNA Prep Buffer was added to the column, followed by centrifugation at 16,000 rpm for 30 seconds, adding 700 µl of DNA/RNA wash buffer to the column, and centrifugation at 16,000 rpm for 30 seconds. Another 400 µl of the wash buffer was added to the column followed by centrifugation at 16,000 rpm for 2 minutes. We then carefully transferred the yellow filter into a nuclease-free 1.5 ml tube and added 50 µl of DNase/RNase-free water directly in the center of the yellow column matrix and incubated it at room temperature for 5 minutes. While waiting, we assembled a Zymo-Spin III-HRC Filter unit and added 600 µl of ZymoBIOMICS HRC Prep Solution into it, followed by centrifugation at 8000 rpm for 3 minutes. The elution column was then centrifuged at 16,000 rpm for 30 seconds, and the eluted DNA was transferred into the prepared HRC filter unit and centrifuged for 16,000 rpm for 3 minutes to elute the final DNA into1.5 ml tube.

Microbiome RNA extraction

We transferred the RNA sample into a SpinAway Green Filter unit, followed by centrifugation at 16,000 rpm for 30 seconds, replacing the collection tube, adding 400 µl of DNA/RNA buffer to the column, mixing by pipetting, and centrifugation at 16,000 rpm for 30 seconds. We added 700 µl of DNA/RNA wash buffer to the column and centrifuged at 16,000 rpm for 30 seconds, followed by adding another 400 µl of the wash buffer to the column, centrifugation at 16,000 rpm for 2 minutes. We then carefully transferred the green filter into a nuclease-free 1.5 ml tube, added 30 µl of DNase/RNase-free water directly in the center of the yellow column matrix, and incubated it at room temperature for 5 minutes. While waiting, we assembled the Zymo-Spin III-HRC unit, added 600 µl of HRC prep solution, and centrifuged at 8000 rpm for 3 minutes. The green column matrix was centrifuged at 16,000 rpm for 30 seconds to elute the RNA into the nuclease-free tube. The eluted RNA was transferred into the prepared HRC filter in another 1.5 ml nuclease-free tube and centrifuged at 16,000 rpm for 3 minutes to elute the final RNA in 1.5 ml tube. The RNA tubes were labeled and stored at -80°C.

DNA and RNA quality assessment

We used the Fragment Analyzer Long Fragment (LF) assay to examine the DNA quality and the Bioanalyzer Pico RNA Assay to

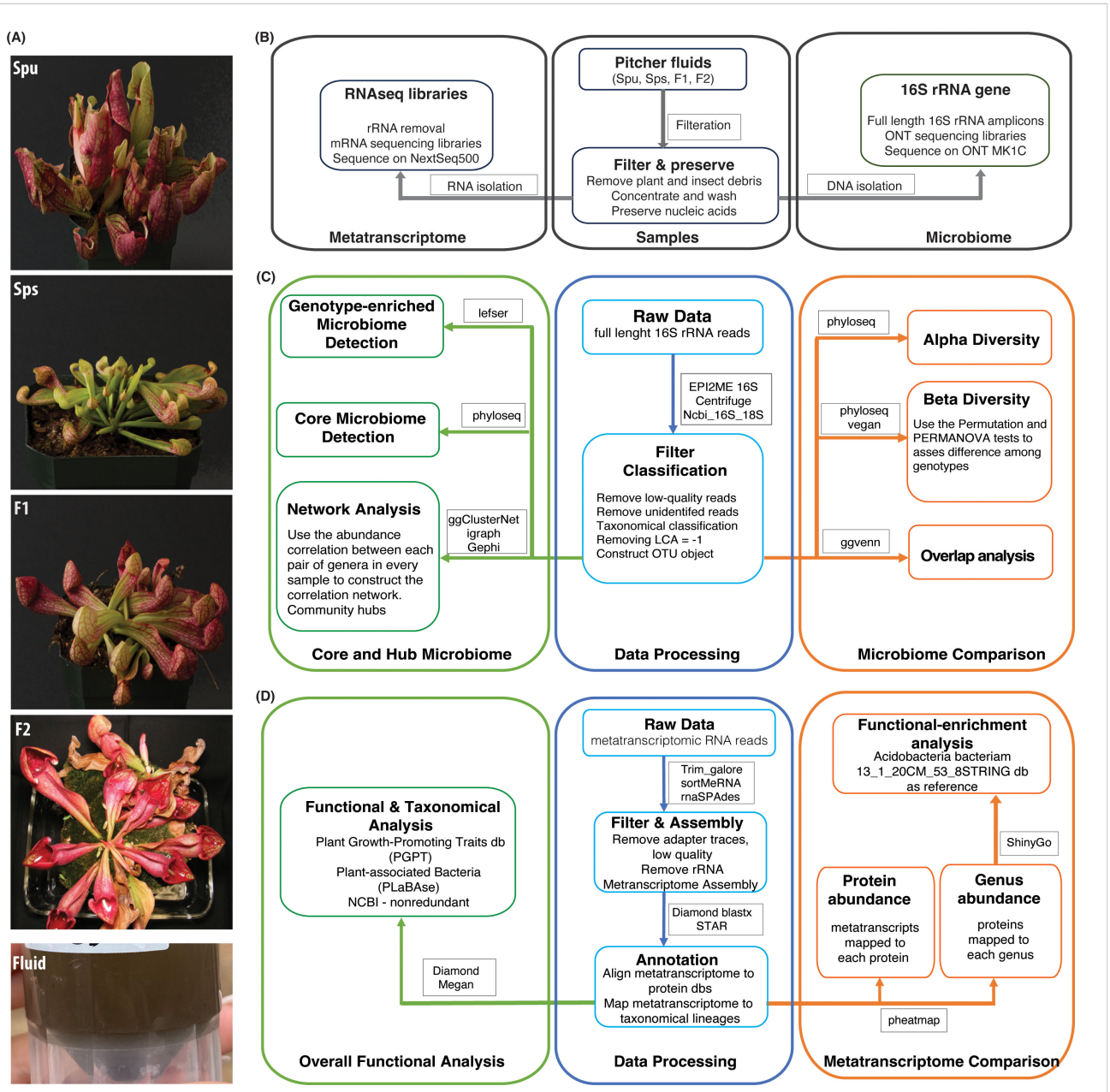


FIGURE 1
Summary of the experiment and analysis pipelines. **(A)** Photos of the parents (Spu and Sps) and selected F1 and F2 individuals of the Spu X Sps mapping population, and a sample of the collected pitcher fluid. **(B)** The experimental process of preparing the pitcher fluid samples and sequencing. **(C)** Modules of the 16S rRNA microbiome analysis pipeline. **(D)** Modules of the metatranscriptome RNA analysis pipeline.

assess the RNA integrity. The concentrations were assessed using the Qubit 4 fluorometer high-sensitivity assays.

Full-length 16S rRNA gene libraries and sequencing

we used the Oxford Nanopore Technology (ONT) amplicon sequencing method to build the 16s rRNA gene libraries and sequence them on the MinIon Mk IC platform. First, we built the libraries using the 16S barcoding kit (SQK-16S024) following the manufacturer's protocol. Briefly, the ONT primers (F: TTTCTGTTGGTGCTGATATTGCAGRGTTYGATYMTG GCTCAG, and R: ACTTGCCTGTCGCTCTATCTTCRGYTACCT TGTTACGACTT) were used to amplify the full-length rRNA gene from the isolated microbiome DNA samples. The amplicons (~ 1500 bp) were barcoded by PCR using the ONT barcoding primers (SQK-16S024), followed by multiplexing, and ONT adapter ligation. The Final library pool was sequenced on the Mk IC platform using FLO-MIN106 flow cells. The final pool of 16S full-length ONT libraries included 13 samples (4 *Spu*, 3 *Sps*, 3 F1, and 3 F2). One of the three F2 samples failed in the sequencing, leaving only two F2 samples.

Metatranscriptomic library preparation and sequencing

We treated the metatranscriptome total RNA with the FastSelect 5S/16S/23S reagent (Qiagen, cat# 335921) to deplete the rRNA following the manufacturer protocol. We started with 12.5 µl total RNA, fragmented for 2 min at 89 °C, and eluted the fragment RNA in 18 µl RNase-free water. The rRNA depleted was used immediately to prepare the metatranscriptome libraries using the SEQuoia complete RNAseq kit (Bio-Rad cat# 17005726) following the manufacturer's procedures without any modifications. The final sequencing libraries were assessed on the Fragment analyzer, pooled equimolarly, and sequenced on the NextSeq 500 sequencer using a mid-output kit. The sequencing run produced over 171 million reads.

Microbiome data analysis

Taxonomical classification and abundance

We carried out the primary analysis using the EPI2ME 16S workflow (ONT 16S workflow 2019) to generate the taxonomic classification and abundance of the 16S genes for all samples. The 16S workflow uses the Centrifuge algorithm (Kim et al., 2016) to map the full-length 16s reads to the NCBI 16S-18S database and classify them according to the taxonomy of their best matches. The obtained taxonomical matrix for all samples was then used for more in-depth analysis using R as described in the results section. The overall analysis pipeline (Figure 1C) is divided into three modules: data preprocessing, microbiome comparison, and core and hub microbiome analyses. The R packages and analysis tools utilized in this study pipeline are listed in Figure 1C and will be elaborated on in the results section.

Statistical analysis of alpha and beta diversity

We utilized the Alpha diversity test to evaluate microbial diversity within each group, that is, among replicates of the same genotype (4 Spu, 3 Sps, 3 F1, and 2 F2). The alpha diversity was assessed using three methods: Chao1 (Chao, 1984), Pielou's Evenness (Volvenko, 2014), and Shannon (Shannon, 2001). The pairwise Wilcoxon test (Wilcoxon, 1992) was then used to test the significance of the variance of the alpha diversity among genotypes. The difference in the microbiome composition among genotypes was calculated using beta diversity analysis, followed by visualization using *PCoA* (Principal Coordinate Analysis) (Gower, 1966). The sample distances were calculated using Bray-Curtis distances (Bray and Curtis, 1957; Kers and Saccenti, 2022).

Ubiquitous, genotype-enriched, and core microbiomes

We used an abundance-based approach to identify ubiquitous microbiomes, which is the microbiome existing in all samples above the background noise. This was followed by an overlap analysis using the ggvenn R package (Yan, 2021) to identify the genotype-enriched microbiomes. As for the core microbiome, we used the microbiome R package (Lahti and Shetty, 2018) to identify the core pitcher microbiome after optimizing the prevalence and limit of detection factors. We then employed the Linear Discriminant Analysis Effect Size (LEfSe) (Segata et al., 2011) method to identify the most distinctive microbial genera showing significant differential abundance between each pair of genotypes. The LefSe method is powerful because it combines statistical significance with the evaluation of biological consistency (effect size) using several tools including the Kruskal-Wallis test (Kruskal and Wallis, 1952) and Linear Discriminant Analysis (LDA) (Fisher, 1936).

Network and hub microbiome

We implemented network analysis to gain insights into the structure of the pitcher microbiome community. We deployed ggClusterNet (Wen et al., 2022) to infer the correlation network followed by visualization using Gephi (Bastian et al., 2009). Subsequently, we computed three scores for each node: hub score, within-module connectivity Z-score, and participation coefficient scores. The scores reflect the node's connectivity to all other nodes in the network, to nodes within the same community, and the evenness of edges distribution within each community, respectively. The network structure, and hence all three node scores, will change when the number of nodes, the threshold for correlations, and the *p-values* are altered. Therefore, we examined the changes in the network structure across various correlation criteria, *p*-values, and node counts (Supplementary Figure S1). As a result, we set the correlation criterion to 0.7 and the p-value cutoff to 0.05.

Metatranscriptome data analysis

Metatranscriptome assembly and annotation

The metatranscriptome analysis pipeline is shown in Figure 1D. First, we trimmed the sequencing reads to remove adapter traces as well as short and low-quality reads using Trim_galore. Then

we used the rRNA database v4 and sortMeRNA (Kopylova et al., 2012) to remove rRNA reads from each sample. The remaining metatranscriptomic mRNA reads from all samples were combined and assembled using rnaSPAdes (Bushmanova et al., 2019). The assembled metatranscriptome was deduplicated to remove identical metatranscripts followed by annotation against a customized uniprot_sprot protein database, which included bacterial, fungal, viral and archaea sprot sequences (uniprot_sprot_microbial) using diamond blastx (*–top 1 –evalue 0.05*). Also, we mapped the metatranscriptome to the NCBI nonredundant and GTDB protein databases using diamond blastx to obtain accurate taxonomical lineages. Subsequently, we mapped the cleaned sequence reads of each sample to the annotated metatranscriptome using STAR (Dobin et al., 2013) and compiled the count matrix for all samples with taxonomical information. Lastly, we converted the count matrix to CPM matrix.

Abundance of genera and metaproteins

In order to determine the abundance of taxa, the count matrix was transformed from reads per protein to genus abundance. The genus abundance was determined by counting all metatranscripts mapped to proteins from the same genus (according to the uniprot_sprot annotation) for the *Spu*, *Sps*, F1, and F2 samples. Next, we calculated the *z*-core and *p*-value for each genus count in each sample and filtered out all genera with *p*-values > 0.05. The TMM method was used to normalize the abundances of the filtered genera. Similarly, in order to identify prevalent proteins, the metatranscripts count matrix was transformed into a metaprotein abundance matrix. The metaprotein abundance was determined by counting the metatranscripts mapped to the same protein, based on the uniprot_sprot annotation, from multiple genera. Next, we calculated the *z*-core and *p*-value for each metaprotein count in each sample and filtered out all metaproteins with *p*-values > 0.01, followed by a TMM normalization.

Functional core microbiome

We used two criteria to identify the functional core microbiome: prevalence > 0.4 and limit of detection > 50 metatranscripts. Unlike the structural core microbiome, which was identified based on the 16S rRNA gene data, the functional core microbiome relies on the prevalence and abundance of protein-coding metatranscripts, hence their functions.

Genotype functional-enrichment analysis

We conducted function category enrichment analysis on the proteins that were enriched in each genotype (*p*-value < 0.05) using ShinyGO 0.80 (Ge et al., 2020) and *Acidobacteria Bacteriam* 13_1_20CM_53_8STRINGdb as a reference. The pathway databases of the enrichment analysis included GO biological processes, GO cellular components, GO molecular functions, annotated UniProt keywords, and local network clusters (STRING). We then manually filtered the list of enriched functions to remove redundancy, which existed as a result of using multiple databases.

Overall microbiome functional analysis

We used the intensively curated Plant Growth-Promoting Traits (PGPT) (Patz et al., 2024) from the Plant-associated Bacteria (PLaBAse) project (Patz et al., 2021) to gain insights into the major functions of the pitcher microbiome community. The metatranscriptome was mapped to the mgPGPT protein database using the Diamond blastx (mini. E-value ≤ 0.05) and used the Megan software (Gautam et al., 2023) to further analyze and visualize the results.

Results

Microbiome sequencing data

The ONT sequencing run produced 9,772,540 full-length 16S rRNA gene sequences of which 9,742,538 reads were taxonomically classified (~ 99%) with an average classification accuracy of 93%. The read length average and mode are 1,244bp, and 1,780bp, respectively. The quality score average and mode are 13.91, and 14.05 respectively. Reads without valid barcodes were removed leaving a total of 8,999,370 sequences. The EPI2ME 16S workflow assigned a lowest common ancestor (LCA) score (Munch et al., 2008) to each full-length 16S read (Supplementary Table S1). The LCA score is used instead of the classical "num_genus_taxid" and is used to store records of the classification status. It has three values: 1 for a species-level classification without accurate genus classification; 0 for both species and genus classification; -1 for unsuccessful species classification. We removed 106,817 reads that had a -1 LCA score. The remaining 8,892,553 reads (minimum classification accuracy is 77%) were used in the subsequent analyses.

Microbiome diversity and intergenotype variance

The use of full-length 16S rRNA genes in this analysis offers a heightened degree of precise classification analysis, hence eliminating the necessity for data rarefaction. The non-rarefied count matrix was further normalized as Counts Per Million (CPM) and was utilized to create the phyloseq (McMurdie and Holmes, 2013) object for further examination in the subsequent analysis.

Alpha diversity

The alpha diversity is measured using Chao1 (Chao, 1984), Pielou's Evenness (Volvenko, 2014), and Shannon (2001) methods (Figure 2A). Each method measures particular aspects of microbial diversity. Chao1 assesses the species richness, Pielou's Evenness focuses on the evenness of species distribution, and Shannon measures both richness and evenness. All three measures reveal noticeable variations in microbial richness and evenness between the genotypes (intergenotype) and among the replicates of the same group (intragenotype). When both richness and evenness are assessed simultaneously using Shannon's approach, the median alpha diversities of all samples were more closely aligned. The

observed variations may be attributed to the limited number of replicates, variations in sequence depth among samples, genetic differences among the replicates of the F1 and F2 samples, as well as the differences in microbial diversity. We then used the pairwise Wilcoxon method (Wilcoxon, 1992) to test if the alpha diversity among the parents, F1 and F2 is significantly different. All of the *p*-values were close to 1 (Supplementary Table S1), indicating that none of the pairs' diversity showed a significant difference. This suggests that the low replication and sequencing depth variation did not result in any statistically significant difference in the diversity and species evenness observed across the samples. In the subsequent analysis, we assume there is no significant difference in the alpha diversity among the parents, F1, and F2 generations.

Beta diversity

In Figure 2B, we illustrate the beta diversity among genotypes using the *PCoA* (Principle Coordinate Analysis) (Gower, 1966), where the colored ellipse depicts the dispersion of all samples from each genotype, encompassing 95% of the diversity, assuming a normal distribution. Since there are only two replicates in the F2, it is infeasible to construct an ellipse for it, and hence it uses a straight line to indicate the dispersion. The results show that the microbiome compositions of the two parent species, *Spu* and *Sps*, are different from the descendant generations, F1 and F2. To further explore the inter-genotype microbiome differences, we implemented *PERMANOVA* (Anderson, 2001) to test the difference in the distribution between genotypes. The *p*-value for the test is 0.011, which indicates significant differences in microbial community structure between genotypes. We also implemented the PERMDISP test (Anderson, 2006) to assess inter-genotypes microbiome dispersion. The *p*-value of the test is 0.825, indicating the microbiome dispersion is not significantly different between genotypes. Combining the results from the two tests, we can conclude that there is a significant difference in the microbial community structure between the parental species and their F1, and F2 genotypes.

The genotype-enriched and core microbiomes

We identified the genotype-enriched and core microbiomes in the pitcher fluids of the parents, F1, and F2 groups. Specifically, a genus is deemed genotype-enriched only if it is represented in every replicate of that genotype by at least one full-length 16S rRNA read. Accordingly, out of 1,960 identified bacterial genera, the pitcher microbiomes of *Spu*, *Sps*, F1, and F2 contained 180, 177, 213, and 155 bacterial genera, respectively (Table 1; Supplementary Table S1). There is a total of 342 non-redundant bacterial genera constituting

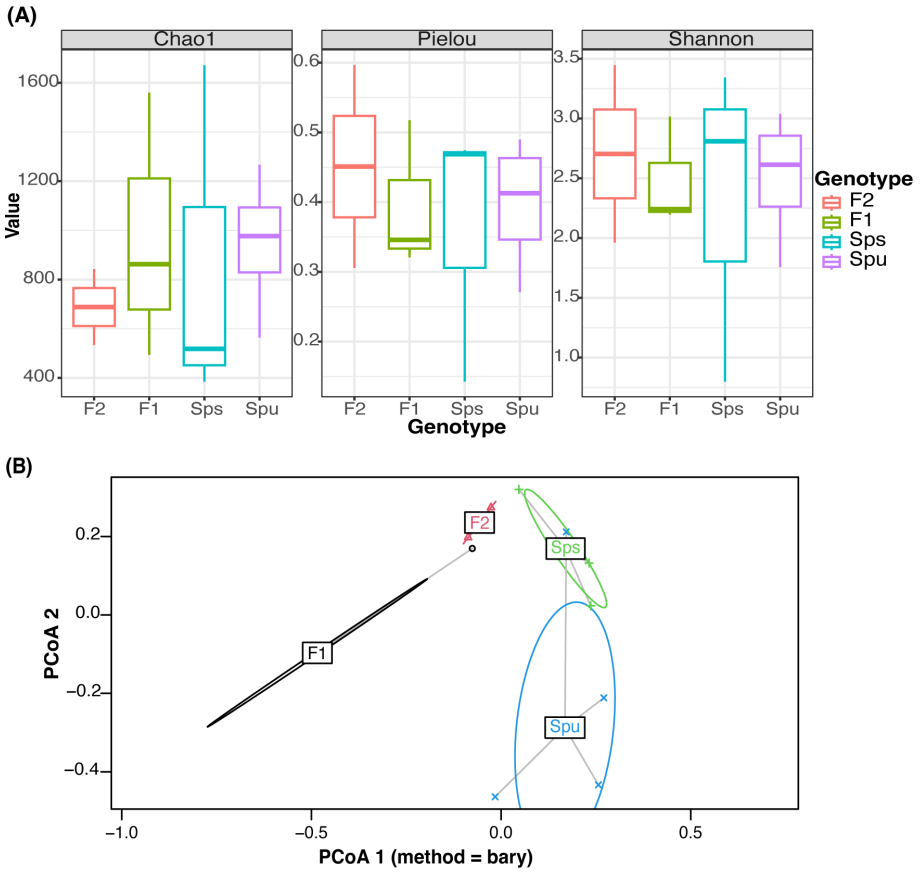


FIGURE 2
Microbiome diversity assessment. **(A)** Comparison of alpha diversity of different genotypes using Chao, Pielou, and Shannon tests to assess richness, evenness, and both, respectively. **(B)** Visualization of the relationship between every sample using Principal Coordinate Analysis. The color represents the genotype of each sample. The ellipses mark the variation of each genotype.

both the ubiquitous (62), and genotype-specific (145) bacterial taxa, as well as taxa existed in two to three genotypes (135). We utilized the ggvenn R package (Yan, 2021) to identify the microbiome overlap among genotypes (Figures 3A, C). Specifically, 62 (18.1%) bacterial genera exist in all samples of *Spu*, *Sps*, F1, and F2, among which the top 5 abundant genera are *Aeromonas*, *Rhodopseudomonas*, *Achromobacter*, *Paraburkholderia*, and *Azospirillum*.

For the core microbiome, we considered a genus to be a member of the core microbiome if its least prevalence under different detection thresholds is greater than 40%, as described in the microbiome analysis R package (Lahti and Shetty, 2018). Unlike the ubiquitous microbiome, the core microbiome requires a specific abundance (the detection threshold) in a specific proportion of samples (prevalence) instead of in every sample. The heatmap in Figure 3B illustrates the prevalence of the core microbiome under different detection thresholds (100 to 1,000). With a prevalence equal to 0.4 and a detection limit equal to 500, a total of 58 bacterial genera constituted the core microbiome of *Spu*, *Sps*, F1, and F2 (Supplementary Table S1). The top 5 genera with the highest CPM normalized count in the core microbiome are *Azospirillum*, *Achromobacter*, *Rhodopseudomonas*, *Mucilaginibacter*, and *Aeromonas*. There are 42 genera that are shared by the ubiquitous and the core microbiomes, representing 67.7% and 72.4% of each, respectively (Supplementary Table S1).

TABLE 1 summary of the results.

Bacterial 16s rRNA gene reads classification and analysis	
Full-length classified 16S rRNA gene	8,892,553 reads
Classified bacteria	1960 genera
Ubiquitous and genotype-enriched (non-redundant):	432 genera
Spu	180 genera
Sps	177 genera
F1	213 genera
F2	155 genera
Structural core microbiome	58 genera
Community clusters	3-7 clusters
Community hubs	4 genera
Community Connectors	23 genera
Microbial Metatranscriptome classification and functions	
Metatranscripts with microbial protein annotation	65,578 metatranscripts
Microbial genera based on metatranscriptome:	2157
Viral	121 genera
Bacterial	1067 genera
Archaeal	11 genera
Fungal	270 genera
Functional core microbiome	90 genera
Bacterial	68 genera
Fungal	22 genera
Genotype-enriched microbes (p-value < 0.05)	25 genera
Non-redundant microbial protein matches	9,693 proteins
Genotype-enriched microbial proteins (p-value < 0.01)	212 proteins
Metatranscripts mapped to PGPT[1] proteins:	50424 metatranscripts
Direct effect traits	13,519 metatranscripts
Indirect effect traits	29,029 metatranscripts

[1]Plant-grwoth promoting traits (PGPT).

In the genotype-enriched microbiome, 35 genera (10.2%) are unique to the *Spu* genotype, 23 genera (6.7%) to the *Sps* genotype, 68 genera (19.9%) to the F1 genotype, and 19 genera (5.6%) to the F2 genotype (Figure 3A). The results show that the microbiome of the F1 genotype differs significantly from that of the other genotypes. A similar observation can be obtained from the comparisons presented in Figure 3C. The microbiomes of the two parents (*Sps* and *Spu*) exhibit a greater similarity to each other than to either the F1 or F2 microbiomes (Figure 3A). The proportion of shared genera between *Sps* and *Spu* microbiomes is 37.8% when compared to F1 and 44.5% when compared to F2 microbiomes, which is much higher than the overlap between F1 and F2 microbiomes (Figure 3C). The overlaps between F1 and F2 microbiomes are 35.4 relative to the *Spu* and 36.8 relative to the *Sps* microbiomes. The results confirm that the *Sps* microbiome is similar to that of *Spu*, while the F1 microbiome is more closely related to the F2 microbiome (Figure 2B).

We employed LEfSe analysis (Segata et al., 2011) to examine the microbiomes of *Spu*, *Sps*, F1, and F2 to identify the most distinctive microbial genera showing significant differences in abundance across all genotype. Figure 3D displays the LEfSe analysis results for the comparisons *Sps* vs. F1, *Spu* vs. F1, and *Spu* vs. *Sps*. The F2 microbiome was excluded from the analysis due to having only two replicates, as previously explained. At a log10 LDA score of 2, there are more distinct genera in the microbiomes of F1 compared to *Spu* and *Sps*, than between *Spu* and *Sps* (Figure 3C). These results are consistent with the PCoA analysis (Figure 2B) and microbiome overlapping analysis (Figures 3A, C).

The microbiome communities networks and hubs

This analysis included 582 genera (out of 1,960) with an accumulative normalized abundance of more than 100 across all samples. In the correlational network (Figure 4A), nodes with correlation ≥ 0.7 and p-value ≤ 0.05 are connected with edges. The greedy algorithm (Clauset et al., 2004) assigned the network nodes to 5 microbiome communities (clusters), which are indicated by the node and edge colors (Figure 4A-left). The edge colors represent the correlation values of connected nodes, where blue and red edges signify positive and negative correlations, respectively. Two communities contain less than 10 genera (Cluster 1, 2), and the majority of the genera (97%) are assigned to the other three communities (Cluster 3, 4, 5). Cluster 1 (dark green) is located in the negatively correlated part of the network and serves as the hub

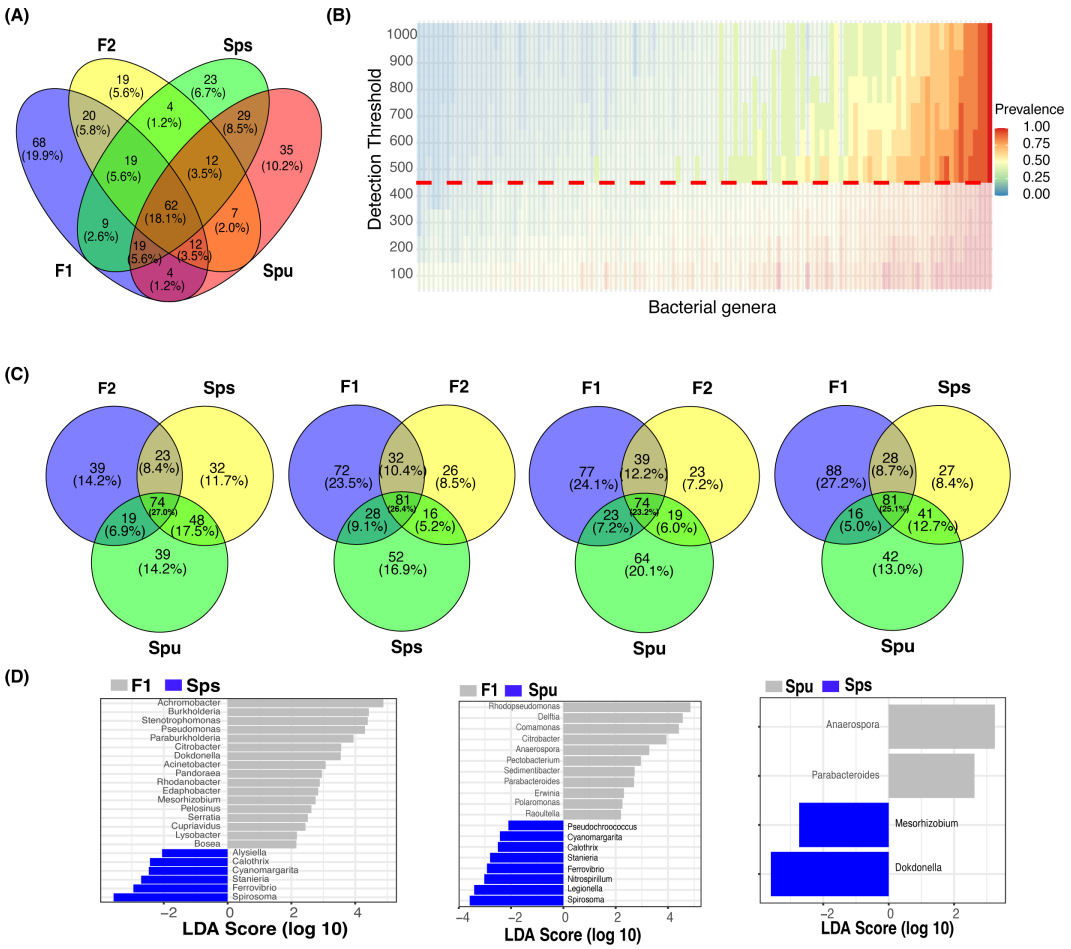


FIGURE 3
The core and genotype enriched microbiome analysis. **(A)** A Venn diagram showing the number of overlapped genera among Spu, Sps, F1, and F2. **(B)** Heatmap of the prevalence of different microbiomes under different detection thresholds. The detection value and prevalence cut-off limits are represented by the horizontal red line and transparent area, respectively. **(C)** Four Venn diagrams comparing the overlapping genera among every three genotypes to highlight the genotype specific numbers. **(D)** Genotype-enriched genes detected by Linear discriminant analysis Effect Size.

node for the other three major microbiome communities. Cluster 2 (orange) has very few connections to the rest of the network. Cluster 3 (pink), Cluster 4 (light green), and Cluster 5 (blue) construct the main part of the microbiome network. The top 5 genera with the highest normalized count in these five communities are: *Sphingobacterium*, *Pedobacter*, *Tissierella*, *Epilithonimonas*, and *Diaphorobacter* for Cluster 1; *Gracilibacter*, *Desulfonatronobacter*, *Desulfococcus*, *Maritalea*, and *Desulfoconvexum* for Cluster 2; *Azospirillum*, *Mucilaginibacter*, *Legionella*, *Terriglobus*, and *Methylacidimicrobium* for Cluster 3; *Achromobacter*, *Rhodopseudomonas*, *Aeromonas*, *Comamonas*, and *Burkholderia* for Cluster 4; *Sphingomonas*, *Luteibacter*, *Granulicella*, *Reyranella*, and *Terrimonas* for Cluster 5.

We also computed the within-model connectivity *z*-score and participation coefficient (Guimera and Nunes Amaral, 2005) for each node in the whole network, which indicates the connectivity within and between communities, respectively (Figure 4A-right). Nodes are classified into four categories based on their scores: peripheral, connections, module hubs, and network hubs (Guimera and Nunes Amaral, 2005). A network hub node has a stronger connection to the network than the ordinary node. Connector and module hub nodes have more connectedness across and within communities than the average of community nodes. The remaining genera are classified as peripheral. As shown in Figure 4A-right, the whole network has 2 module hubs and 10 connectors. The module hubs are *Schlesneria* and *Methylophaga*, whereas the connectors are *Dendronalium*, *Calothris*, *Lacrimispora*, *Sphingobium*, *Mesorhizbobium*, *Parabacteroides*, *Legionella*, *Leifsonia*, *Lactococcus*, and *Desulfococcus*.

To gain a closer insight into the microbiome communities, we zoomed in on the top 100 genera ranked by their hubscore, within-model connectivity *z*-score, normalized abundance, and participation coefficient (Figures 4B–E). In the co-occurrence subnetworks (Figures 4B–E, left), node colors denote the greedy algorithm-detected clusters and node sizes represent the accumulated abundance in all 12 samples. The connectivity scores of the four subnetworks are illustrated in Figures 4B–E (right panel), respectively. The inferred networks of the top 100 genera ranked by hubscore and within-model *z*-score have similar topology including 3 highly connected community clusters each,

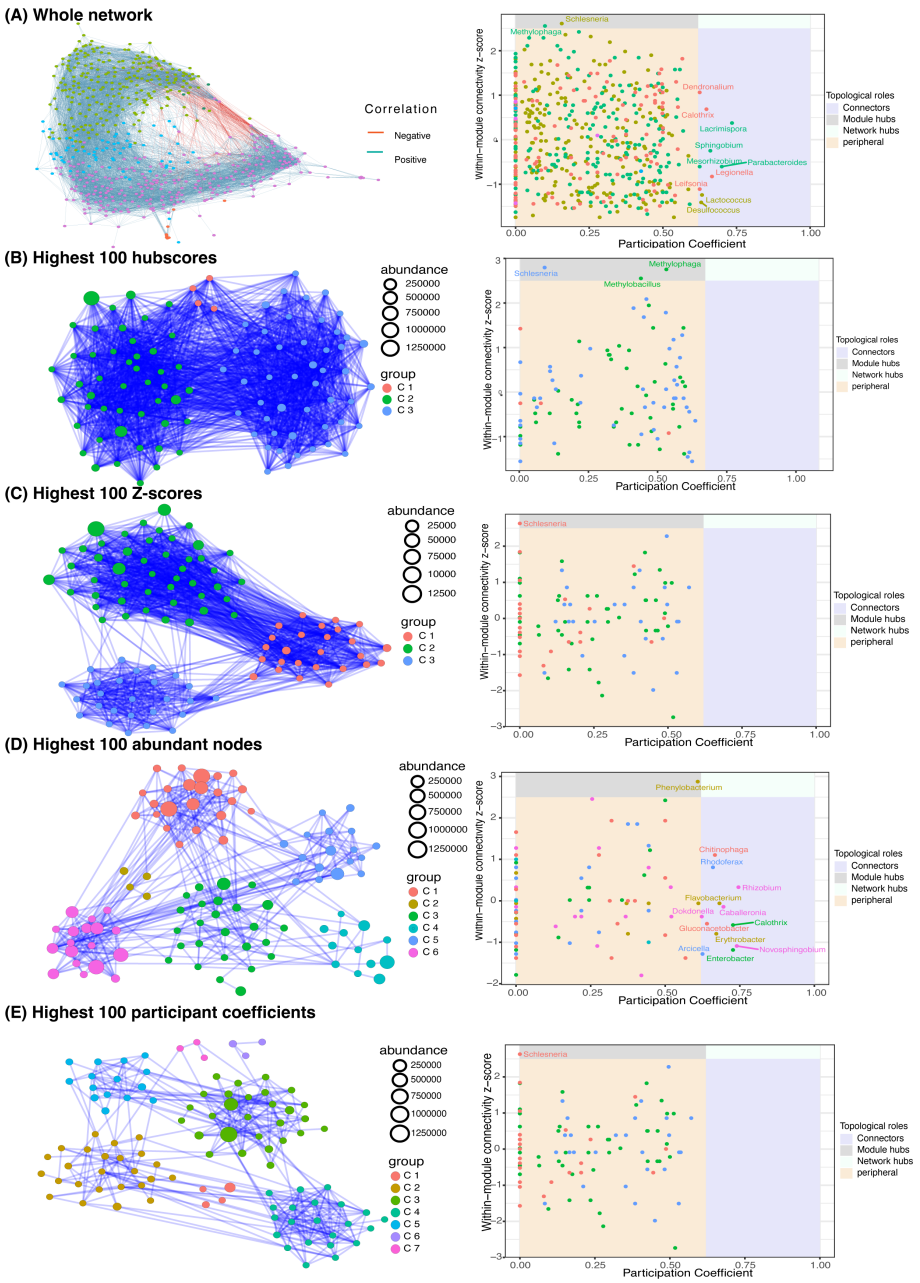


FIGURE 4
Network analysis. **(A)** the whole network of 582 filtered genera (left) and connectivity scores (Right). (**B-E** left) the subnetworks of top 100 genera ranked by their hubscore **(B)**, z-score **(C)**, normalized abundance **(D)**, and participant coefficient **(E)**. The connectivity scores of the four subnetworks are in their corresponding right panels.

two module hubs (*Schlesneria* and *Methylophaga*), and no connectors or network hubs (Figures 5B, C). The high connectivity in these two subnetworks can be explained by the fact that both the hubscore and within-model connectivity *z*-score prioritize the connections of each node, resulting in a highly interconnected network with fewer distinct clustering patterns. Similarly, the inferred subnetworks of the top 100 genera ranked by normalized abundance and participation coefficient (Figure 4E, E) have similar structures with 6 and 7 clusters, and *Phenylobacterium* and *Schlesneria* as module hubs, respectively. The abundance-based subnetwork has the following 11 connector nodes: *Phenylobacterium*, *Schlesneria*, *Chitinophaga*, *Rhodoferax*, *Rhizobium*, *Flavobacterium*, *Dokdonella*, *Caballeronia*, *Gluconacetobacter*, *Calothrix*, *Erythobacter*, *Novosphingobium*, *Arcicella*, and *Entrobacter* (Figure 4E-right). The abundance and participation coefficient subnetworks seem to capture more of the network topological information and identify more community connectors and hubs.

The overlap among the four subnetworks is presented in Figure 5. A total of 67 genera (50.4%), including a module hub (*Schlesneria*), appeared in both hubscore and *z*-score subnetworks (Figure 5A). On the other hand, only 22 genera (12.4%) are common to both the

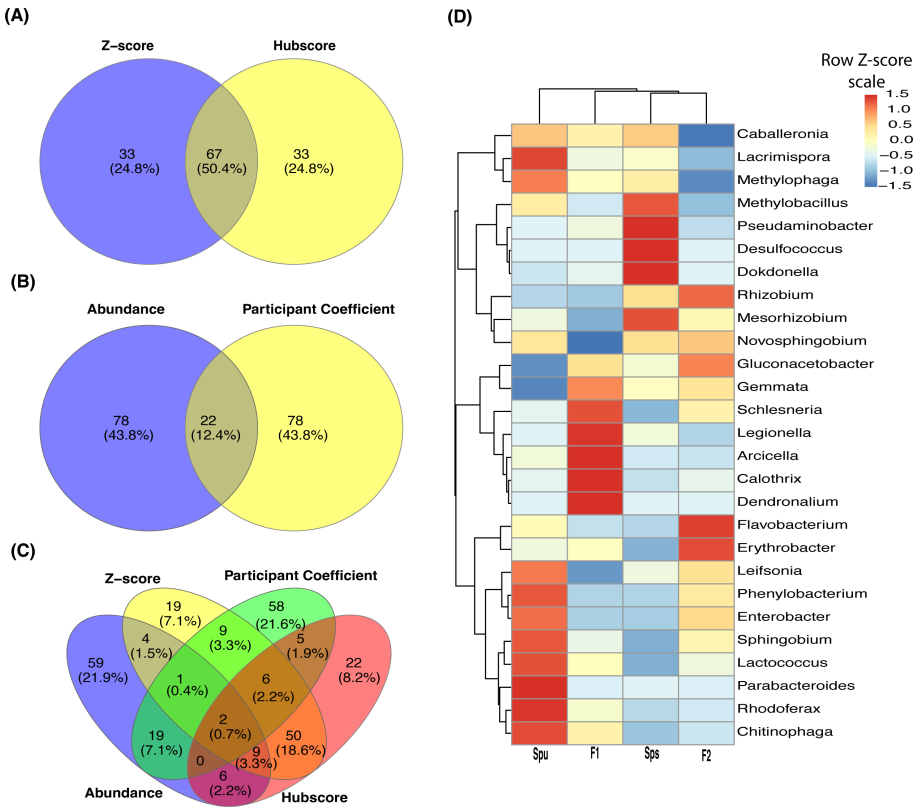


FIGURE 5
Genera overlapping among subnetworks and abundance of hub genera. **(A-C)** Venn diagrams showing the number of overlapped nodes (genera) among the four subnetworks. Overlapping between the hubscore and z-score top most 100 genera **(A)**, normalized abundance and participant coefficients **(B)**, and all four subnetworks **(C)**. **(D)** A heatmap showing the normalized abundance of all hub and connector genera in the Spu, Sps, F1, and F2 samples.

abundance and participation coefficient subnetworks (Figure 5B). Interestingly, no module hubs or connectors were shared between these two subnetworks. Only two genera overlapped among the four subnetworks (Figure 5C). The participation coefficient quantifies the overall connectivity of each node, making it a valuable tool for capturing the structure of the network of the top 100 genera. While there is minimal overlap between the genera in the abundance and participation coefficient subnetworks (Figure 5B), both networks have a similar topology, suggesting that these two parameters represent comparable community clusters. The heatmap in Figure 5D illustrates the TMM normalized abundance of the discovered hubs and connector genera in the parents, F1, and F2 data. It reveals three clusters and six subclusters. Also, it reveals the presence of a cluster of extremely prevalent hub and connector nodes in each of the genotypes, indicating that the genotype of the host may impact the formation of the hubs and connectors within the microbiome community.

Characterization of the pitcher metatranscriptome

The assembled pitcher metatranscriptome included a total of 132,710 metatranscripts, with an N50, average and maximum length of 211 bp, 219 bp, and 8048 bp, respectively. About 65,578 (49.4%), 62,382 (47%), and 63,105 (47.6%) metatranscripts have blastx protein hits (e-value ≤ 0.05) in the uniprot_sprot_microbial, NCBI nonredundant, and GTDB protein databases, respectively. Figure 6A shows a random example of the blastx alignment of metatranscripts to reference protein (min. coverage ≥ 70%). The total number of non-redundant proteins is 9693. These best protein hits come from 2157 microbial species representing 121 viral, 1067 bacterial, 11 archaeal and 270 fungal lineages (Table 1; Supplementary Table S2). At the microbial level, the highest represented viral clade is Riboviria (71 metatranscripts), the highest bacterial taxon is proteobacteria (17,517 metatranscripts), and an unclassified Archaea (3975 metatranscripts). About 9021 metatranscripts mapped to proteins from eukaryotic organisms, the highest among them were *Cryptophyceae* (432), *Discoba* (302), *Opisthokonta* (503), and *Sar* (4765) (Figure 6B). The metatranscriptome data revealed a functional core microbiome consisting of 90 microbial species, including 68 bacterial taxa and 22 fungal taxa (Supplementary Table S2). Interestingly, sixteen bacterial taxa are part of both the structural and functional microbiomes (Supplementary Table S2). The taxonomy tree generated from the metatranscripts (Figure 6B) offers a visual representation of the general classification of the microbial taxa in the pitcher microbiome. It is worth noting that there can be heterogeneity in the microbiome structure from one individual to another. However, as a whole, the pitcher microbiome appears to have a very intricate structure with multiple levels of interactions.

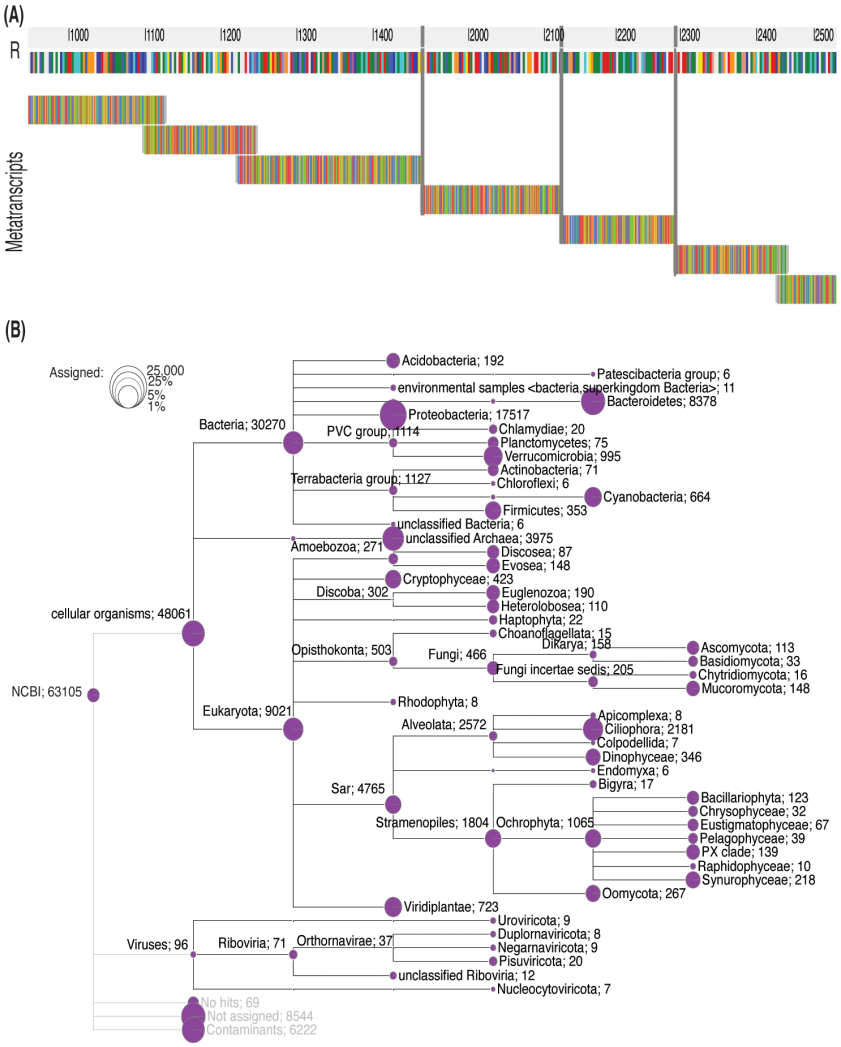


FIGURE 6
Taxonomy tree of the identified taxa based on the pitcher metatranscriptome alignment to proteins. **(A)** Metatranscripts of variable lengths aligned to the protein MAG TPA: carbamoyl-phosphate synthase large subunit (*Sphingobacterium* sp.). The vertical gray lines represent gaps. **(B)** The taxonomy tree of the identified taxa. The circles represent the log scale of number of metatranscripts assigned to the tree nodes and leaves. The numbers listed after each taxon name is the number of metatranscripts mapped to the proteins of this taxon.

Genotype-based differences in the pitcher metatranscriptomes

We examined the differences in the pitcher metatranscriptome among the genotypes (*Spu*, *Sps*, F1, and F2) using the genera and metaproteins enrichment analysis. For each genotype, we selected the genera with p-value < 0.05 and metaproteins with p-values < 0.01. There were 25 genera and 212 metaproteins meeting these criteria. The normalized abundance of these genera and proteins is illustrated in the heatmaps in Figure 7. It is evident from the heatmap clustering that there are differences in the genera abundance (Figure 7A) and metaprotein accumulation (Figure 7B) among the genotypes. Note that in Figure 7B, the heatmap includes only 25 proteins, but the complete list of proteins is provided in Supplementary Table S2. Despite the intergenotypic difference in the genera diversity and metaproteins, the enriched functional categories in all genotypes seem to be very similar, with small variations in the fold changes and number of genes in each category (Figure 8). For each of genotype, proteins with p-values < 0.01 (see materials and methods) were used in the functional category enrichment analysis. These sets included 199, 237, 236, and 201 proteins in *Spu*, *Sps*, F1, and F2 respectively. The analysis revealed a significant (FDR < 0.01) enrichment (more than two folds) of multiple functional categories in all genotypes. Nevertheless, there were no discernible differences in the enriched categories across the different genotypes (Figure 8), suggesting a microbial function convergence in the pitcher microcosm despite variations in structure and diversity.

Phyllosphere and rhizosphere microbiomes

We compared the identified pitcher microbial taxa with previously reported phyllosphere and rhizosphere microbiomes (reviewed in (Saeed et al., 2021; Bashir et al., 2022)). As presented

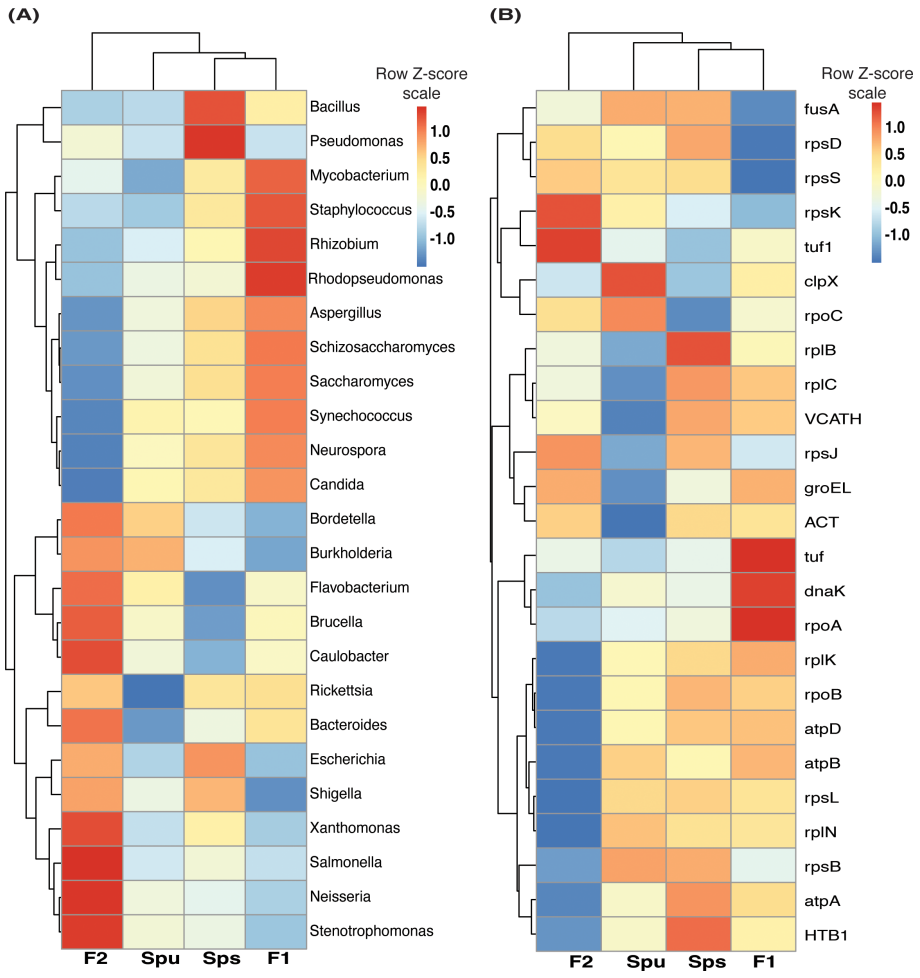


FIGURE 7
The pitcher metatranscriptome: the abundant genera and proteins showing species-related patterns. **(A)** heatmap showing the TMM normalized abundance of the genera in the samples of *Spu*, *Sps*, F1, and F2 (p-values < 0.05). **(B)** similar to "A" but showing the abundance of the 25 protein with lowest *P*-values.

in Table 2, groups of 30, 23, and 20 microbial taxa that were reported to be performing different functions in the phyllosphere, rhizosphere, and both, respectively, were also present in the pitcher microbiomes in this study. Furthermore, many of these microbes were community connector hubs, indicating that they play a major role in microbiome assembly and functions (Table 2). The richness of the pitcher microbiome with both phyllosphere and rhizosphere microbes indicates its intricate nature and suggests that its functions extend beyond the traditional role of supporting plant nutrition.

The pitcher metatranscriptome functions

Our analysis of the pitcher metatranscriptome using the PGPT database revealed major function categories (Table 1; Figure 9). About 50,424 metatranscripts were mapped to PGPT proteins that are associated with different plant growth-promoting functional categories. Globally, 13,519 and 29,028 metatranscripts are mapped to functional groups with direct and indirect effects on plant growth traits, respectively (Figure 9). Among the microbiome traits that have direct effects on plant growth are biofertilization, bioremediation, and phytohormonal plant signals. The microbiome traits that have indirect effect on plant growth include stress control, immune response stimulation, colonizing plant systems and microbial competitive exclusion. As shown in Figure 9, the size of the green circle represents the log scale of the number of metatranscripts assigned to the corresponding function. Altogether, there are 44 major microbiome traits that have direct effect on the host-microbiome and microbiome-microbiome interactions. Traits that are represented by over a thousand metatranscripts include nitrogen acquisition (1003), phosphate solubilization (1751), heavy metals detoxifications (2413), plant vitamin production (1069), neutralizing biotic stress (1578), neutralizing abiotic stress (6230), universal stress response (1062), chemotaxis (1033), plant-derived substrate usage (7067), colonization surface attachment (1485), cell envelop remodeling (1618), biofilm formation (2313), bacterial fitness (2526), and bacterial secretion (1188). These findings reveal the complex functions of the pitcher microbiome at several levels, providing a chance to investigate a wide range of microbiome traits and their host interactions.

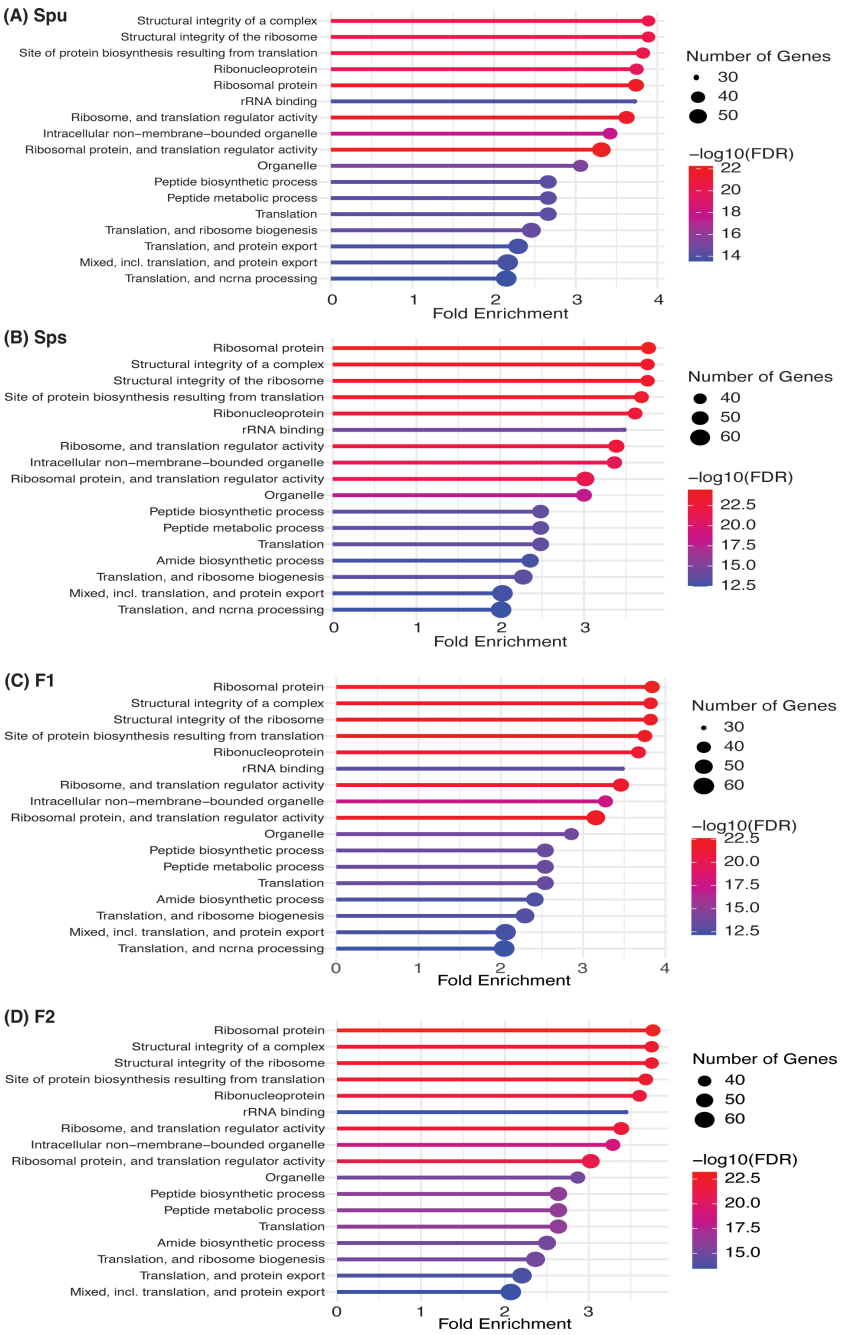


FIGURE 8
Function enrichment of the genotype-enriched metaproteins against the Go functional categories, uniprot keywords, STRING functional lists of *Acidobacteria*. The number of genes, FDR and fold enrichment of the significantly enriched functional categories in *Spu*, *Sps*, F1, and F2 metatranscriptomes are shown in **(A-D)**, respectively.

Discussion

The *Sarracenia* genus relies on its microbiome to facilitate the digestion of trapped prey and to provide important nutrients such as nitrogen (N) for protein synthesis, phosphorus (P) for nucleic acid synthesis, and magnesium (Mg) and iron (Fe) for chlorophyll synthesis. Previous research suggested that factors other than prey capture and colonization by eukaryotic species may affect the recruitment of bacteria to the pitchers (Grothjan and Young, 2019). Also, It has been demonstrated that plants have the ability to modify the composition of the microbiomes in their surroundings (Mahnert et al., 2015) and host genetic variations have an impact on the structure and functions of the microbiome in humans (Turnbaugh et al., 2009; Blekhman et al., 2015). Extensive research has been conducted on the convergent evolution of carnivorous plant leaves, particularly those belonging to the

TABLE 2 The phyllosphere (PS) and Rhizosphere (RS) bacterial (B) and fungal (F) taxa in the pitcher microbiome and metatranscriptome.

Literature[1]			*This study*		
Microbes	Plant space	Type	Microbiome	Metatranscriptome	Core/Hub
Aureobasidium	PS	F		D	
Azorhizobium	PS	B	D	D	
Beijerinckia	PS	B	D	D	
Chaetomium	PS	F		D	
Cladosporium	PS	F		D	
Clonostachys	PS	F		D	
Colletotrichum	PS	F		D	
Cryptococcus	PS	F		D	
Curtobacterium	PS	B	D		
Enterobacteria	PS	B		D	
Erwinia	PS	B	D	D	
Exiguobacterium	PS	B	D	D	
Fusarium	PS	F		D	C
Gluconacetobacter	PS	B	D	D	Co
Halomonas	PS	B	D	D	
Hymenobacter	PS	B	D	D	
Hyphomicrobium	PS	B	D	D	C
Massilia	PS	B	D		
Methylobacterium	PS	B	D	D	C
Microbacterium	PS	B	D	D	
Mycobacterium	PS	B	D	D	C
Nocardia	PS	B	D	D	
Pseudochrobactrum	PS	B	D		
Pseudoxanthomonas	PS	B	D	D	C
Rathayibacter	PS	B	D		
Rhodopseudomonas	PS	B	D	D	C
Saccharothrix	PS	B	D		
Sphingomonas	PS	B	D	D	C
Stenotrophomonas	PS	B	D	D	C
Xanthomonas	PS	B	D	D	C
Acidovorax	RS	B	D	D	C
Agromyces	RS	B	D		
Azoarcus	RS	B	D	D	
Azospirillum	RS	B	D	D	C
Bradyrhizobium	RS	B	D	D	C
Brevibacillus	RS	B	D	D	
Brevibacterium	RS	B	D	D	
Brevundimonas	RS	B	D	D	C

(Continued)

TABLE 2 Continued

Literature[1]			*This study*		
Microbes	Plant space	Type	Microbiome	Metatranscriptome	Core/Hub
Corynebacterium	RS	B	D	D	
Cronobacter	RS	B	D	D	
Debaryomyces	RS	F		D	C
Dietzia	RS	B	D		
Escherichia	RS	B	D	D	C
Halobacillus	RS	B	D	D	
Kluyvera	RS	B	D		
Leptolyngbya	RS	B		D	
Micrococcus	RS	B	D	D	
Neurospora	RS	F		D	C
Nocardioides	RS	B	D	D	
Paenibacillus	RS	B	D	D	
Rahnella	RS	B	D	D	
Ralstonia	RS	B	D	D	C
Rhizopus	RS	F		D	C
Achromobacter	PS, RS	B	D	D	C
Acinetobacter	PS, RS	B	D	D	C
Alcaligenes	PS, RS	B	D	D	
Arthrobacter	PS, RS	B	D	D	
Bacillus	PS, RS	B	D	D	C
Burkholderia	PS, RS	B	D	D	C
Candida	PS, RS	B		D	C
Citrobacter	PS, RS	B	D	D	
Enterobacter	PS, RS	B	D	D	Co
Flavobacterium	PS, RS	B	D	D	Co
Klebsiella	PS, RS	B	D	D	C
Pantoea	PS, RS	B	D	D	
Penicillium	PS, RS	F		D	C
Pseudomonas	PS, RS	B	D	D	C
Psychrobacter	PS, RS	B	D	D	
Rhizobium	PS, RS	B	D	D	Co
Rhodococcus	PS, RS	B	D	D	
Saccharomyces	PS, RS	F		D	C
Staphylococcus	PS, RS	B	D	D	C
Streptomyces	PS, RS	B	D	D	C
Variovorax	PS, RS	B	D	D	C

[1] (Saeed et al., 2021; Bashir et al., 2022).

The listed microbial taxa have been found in the phyllosphere and/or rhizosphere of many plant species. Here we report the coexistence of both PS and RS taxa in the Sarracenia pitcher fluids. D, detected; Co, community connector, and C, core genera.

Sarracenia genus (Thorogood et al., 2018; Wheeler and Carstens, 2018; Chomicki et al., 2024). The evolution of the composite trait of the insect-trapping pitchers in carnivorous plants is believed to have occurred through separate modifications of shape, coloration, and biosynthetic pathways (reviewed in (Chomicki et al., 2024)). We are exploring the concept of considering the pitcher microbiome properties as an extension of plant traits that have evolved alongside carnivory as a part of the insect-trapping composite trait. Our hypothesis is that carnivorous plants have gained genetic traits during the convergent evolution of their pitcher trait, enabling them to interact with their vital microbiomes. Our results show that pitcher plants impact the structure and function of their microbiomes, which in turn have a broader effect on their growth, extending beyond mere nutritional support.

The host genotype significantly influence the microbiome assembly

In their natural habitats, the microbiome diversity diverged among the *Sarracenia* host species (Heil et al., 2022). In this study, we detected statistically significant differences in the structure of the pitcher microbiome communities among the parental species and their F1 and F2 genotypes under the same greenhouse conditions. The finding is supported by the *PERMANOVA* analysis, which indicated that the variations in distribution among genotypes are statistically significant (p-value = 0.011). The *PERMDISP* test indicated that there is no significant difference in the dispersion (p-value = 0.825) between genotypes, ruling out the possibility that difference in sequencing depth between samples could have produced the observed distribution differences. Both Alpha-diversity and PCoA (Figure 2) analyses unveiled differences in the microbiome diversity among individuals of the same group and between the genotypes. Although the majority of the microbial taxa existed in the microbiomes of all species and genotypes, there were genotype-enriched microbial taxa (Figures 3A–D). The microbiome of the F1 generation differed significantly from all other samples, which could be attributed to its dominant genetics that are expressed in more distinctive traits. The variation in the microbiota between the parental species is smaller than the variation observed between either of them and their F1 and F2 genotypes. This observation is reinforced by the number of distinctive genera, as identified by the LEfSe analysis, between the parental species and their F1 and F2 generation (Figure 3C). The F2 generation exhibits increased genotype complexity and genetic variation, resulting in a wide range of traits, including those that influence the assembly and structure of their microbiome. The abundance patterns of the overall microbiome community hub and connector genera varied significantly among the different genotypes (Figures 4, 5). This suggests that the genotypes may influence the development of their respective local community hubs and connectors. By utilizing the *Sarracenia* mapping population, our results suggest that host genetic factors are impacting the observed significant variance in the microbiome structure.

The core microbiomes and plant growth-promoting traits

The richness and diversity of the *Sarracenia* core microbiome suggests that it possesses a wide range of complex traits and levels of interactions. Over 1900 bacterial genera were identified using 16S the rRNA gene, including 62 ubiquitous genera, 145 genotype-enriched genera, and 58 genera constituting the structural core microbiome. More than 2100 microbial genera were found using the metatranscriptomic data, including 68 bacterial and 22 fungal genera constituting the functional core microbiome. There are 16 genera existed in both structural and functional core microbiomes, including *Acidovorax*, *Bradyrhiobium*, *Bukholderia*, *Caulobacter*, *Chromobacterium*, *Clostridium*, *Cupriavidus*, *Magnetospirilllum*, *Mesorhizobium*, *Novosphingobium*, *Paraburkholderia*, *Pseudomonas*, *Rhizobium*, *Rhodopesudomonas*, *Rickettsia*, and *Stenotrophomonas*. The functional core microbiome relies on the prevalence and abundance of protein-coding metatranscripts, while the identification of the structural microbiome relies on the prevalence and abundance of 16S rRNA gene reads. Many of the core genera especially the prevalent ones are among the group of plant-associated microbes known as plant growth-promoting Rhizobacteria (PGPR) (Beneduzi et al., 2012; Agarwal et al., 2020). As shown in Table 2, we identified the fungal and bacterial taxa in the pitcher microbiome, which have previously been discovered in rhizospheres and/or phyllosphere of other plant species (reviewed in (Saeed et al., 2021; Bashir et al., 2022)). While PGPR has traditionally been found in the rhizosphere, it is not unexpected to discover them in the pitcher microbiome (phyllosphere) due to the fact that this plant species depends on pitcher leaves for acquiring essential nutrients. This finding prompts inquiries into the functional and structural interactions between the rhizosphere and phyllosphere microbiomes while coexisting in the confined microenvironment of the pitchers. The functional traits of these PGPR genera include the ability to enhance plant growth through biofertilization, act as biocontrol agents to combat pests, and serve as biological fungicides to protect their specific host plants.

Microbiome community clusters and hubs

The network analysis (Figures 4, 5) unveiled many clusters of microbial communities (Supplementary Table S1), resulting in the identification of community hubs and connector genera. The number of community clusters varied between 3 and 7, depending on the network factors used to identify the clusters. For example, the analysis revealed the presence of five distinct community clusters in the entire network. Additionally, the networks of the top 100 abundant and top 100 participation coefficient nodes (genera) exhibited the identification of six and seven community clusters, respectively. But regardless of the method, there are always distinctive community clusters in the inferred microbiome network. The key characteristic of these

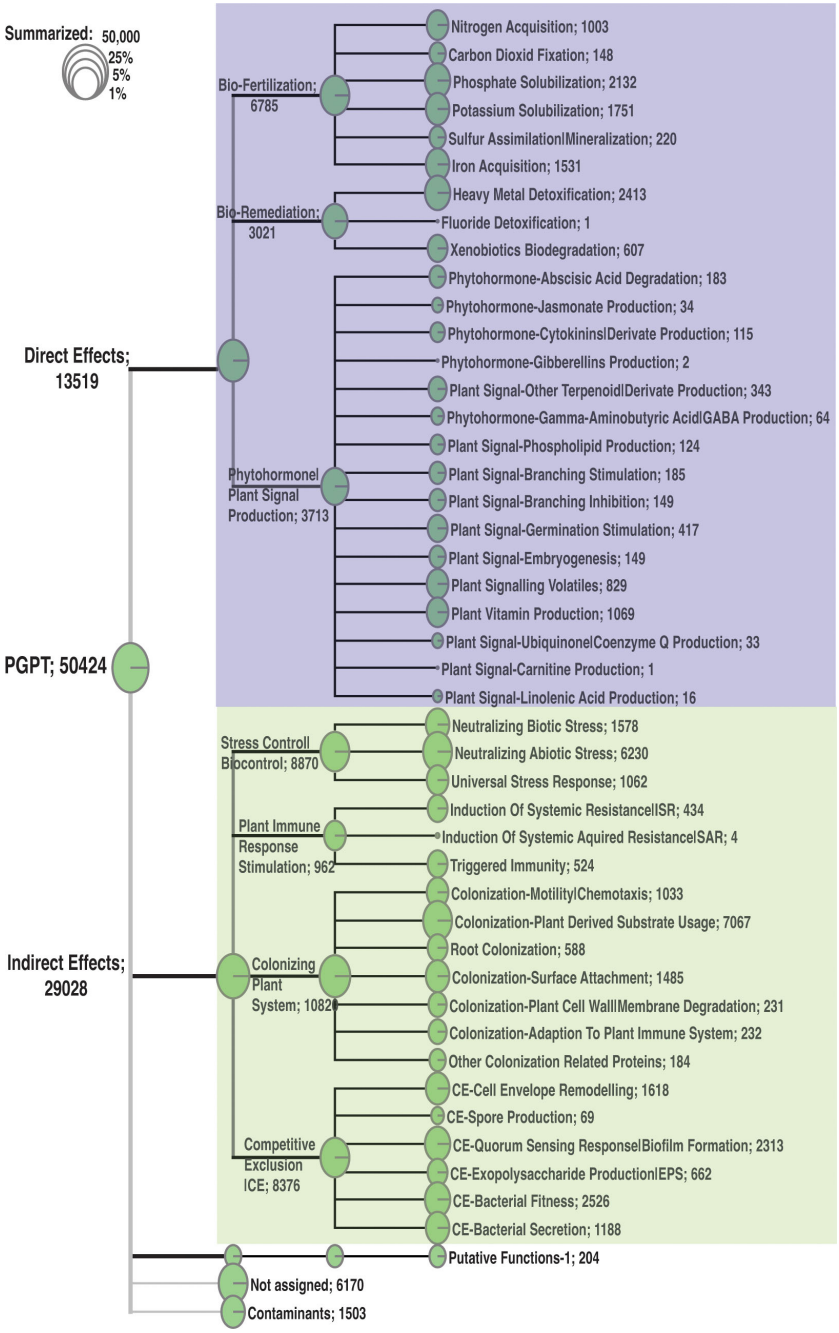


FIGURE 9
Classification of the major function categories of the pitcher microbiome based on the number of metatranscripts mapped to PGPT ontology. The circles represent the log scale of number of metatranscripts assigned to each function. The number following each function is the number of metatranscripts mapped to this function.

communities is the discovery of a total of 27 genera playing the hub and connector roles in these communities. It is crucial to note that the hub and connector genera, based on hubscore, within-model connectivity z-score, and participation coefficient, have a low abundance. This indicates that the genera that perform important roles in the microbiome community are not always the most prevalent genera. Interestingly, the analysis of the normalized abundance of the 27 hub and connector genera in the parents, F1, and F2 genotypes uncovers distinct clusters of genera that vary significantly in their abundance among the genotypes (Figure 5D). This suggests that the host genetic makeup influences the selection of the microbiome hubs and connectors. For instance, the genus *Schlesneria* appeared as a hub genus in the whole network, as well as in the subnetworks, despite not being among the highly abundant genera compared to the other hub and connectors genera (Figure 5D; Supplementary Table S1). As shown in Table 2, the

majority of the hub and connector genera are commonly reported as plant-associated genera living in the plant's phyllosphere and rhizosphere (Mahnert et al., 2018). Overall, the inferred microbiome networks offered insights into the assembly and organization of the pitcher microbiome. This involved the identification of many subcommunities, each with their hub genera, as well as connector genera that facilitate communications between these subcommunities. Interestingly, the community networks also show that the hub and connector genera exhibit varying levels of abundance across different genotypes (Figure 5D), suggesting that the host genotypes may have an impact on their development.

The complex function of pitcher microbiome

Although our findings demonstrate a considerable influence of the host on the diversity and organization of the microbiome, we did not discover any notable genotype-related variations in the enriched functional categories in the metatranscriptome (Figure 8). This analysis, however, lacks the level of detail necessary to discover any fine functional differences. Nevertheless, it suggests that even though there may be fine difference in the microbiome functions among genotypes, the overall microbiome traits and functions remain consistent. This could be explained by functional redundancy among the taxa of the pitcher microbiome and a microbiome functional convergence to achieve similar functional traits (Louca et al., 2016, 2018). A similar conclusion was reached by analyzing the microbial communities and functions in *Spu* populations in natural habitats (Grothjan and Young, 2019). The functional analysis, however, revealed complex and multifaceted functional traits of the pitcher microbiome (Figure 9). At the host-microbiome interaction level, approximately 13.5% of the metatranscripts that have been classified functionally are related to genes involved in biofertilization activities. These activities encompass nitrogen acquisition, phosphate solubilization, and iron acquisition. The presence of a significant portion of the pitcher microbiome that contributes to plant biofertilization is expected, given that acquiring nutrients from leaves is a key characteristic associated with the carnivorous behavior. Nevertheless, the discovery of the pitcher microbiome's bioremediation (6%) and phytohormonal plant signal generation (7%) activities are novel additions to our understanding of the *Sarracenia*-microbiome interaction. The microbiome has a significant role in the detoxification of heavy metals (2413 metatranscripts) and the degradation of xenobiotics (607 metatranscripts) in the pitcher fluid, which serves to safeguard the host plant. These functions directly impact the fitness and defense of the pitcher plant and can be considered extended plant traits. Approximately 58% of the pitcher metatranscriptome is assigned to a set of functions categorized as traits that have an indirect impact on the host plan (Figure 9). The most prevalent functions, as indicated by the abundance of metatranscripts, are those associated with stress regulation, promotion of plant immune response, and colonization of plant systems. On the microbiome-microbiome interaction level, the analysis revealed that major portions of the metatranscriptome are involved in microbial competitive exclusion activities (8376 metatranscripts) and usage of plant-derived substrates (7067 metatranscripts).

These results offer a comprehensive understanding of the functions of the microbiome in connection to the host plant, microbial community, and environmental conditions inside the pitcher microcosm. Furthermore, they highlight some strategies utilized by the host plant to regulate the assembly of the microbiome, such as plant immunity and the release of extracellular substrates. The pitcher contains a varied microbial population with complicated multiscale functions, which can be used for metagenomic mining to discover new beneficial microbial taxa.

Conclusions and prospects

The variations in the assembly and structure of the microbial communities within the pitchers of the *Sarracenia* mapping population can be attributed, at least in part, to the genetic characteristics of the host. The pitcher microbiome performs multiple functions that have both direct and indirect effects on the host. Nevertheless, the precise genetic and biochemical pathways via which the host plant interacts with the microbiome are still unidentified. Our goal is to link the extended traits of the *Sarracenia* plant (namely, microbiome traits and functions) to the genetic elements of *Sarracenia* using cross-species QTL linkage mapping. This will enable us to gain insights into the mechanisms underpinning the interactions between the host and its microbiome, which is an essential requirement for the pursuits of microbiome engineering.

Data availability statement

The raw sequencing data are deposited in NCBI RSA under bioporjct PRJNA1163405. The RSA Ids of the datasets are 43849136 through 43849147.

Author contributions

JC: Data curation, Formal analysis, Software, Visualization, Writing – original draft, Writing – review & editing. IM: Investigation, Methodology, Writing – review & editing. WR: Investigation, Methodology, Writing – review & editing, Resources. MZ: Writing – review & editing, Formal analysis. LJ: Writing – review & editing, Resources. RM: Resources, Writing – review & editing. MA: Conceptualization, Funding acquisition, Resources, Writing –review & editing, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1445713/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Examining the changes in the network structure under various correlation, p-value, and node counts criteria.

SUPPLEMENTARY TABLE S1
microbiome data.

SUPPLEMENTARY TABLE S2
metatranscriptome data.

References

Adlassnig, W., Peroutka, M., and Lendi, T. (2011). Traps of carnivorous pitcher plants as a habitat: composition of the fluid, biodiversity and mutualistic activities. *Ann. Bot.* 107, 181–194. doi: 10.1093/aob/mcq238

Afridi, M. S., Ali, S., Salam, A., César Terra, W., Hafeez, A., Sumaira,, et al. (2022). Plant microbiome engineering: hopes or hypes. *Biol. (Basel)* 11, 1782. doi: 10.3390/biology11121782

Agarwal, P., Giri, B. S., and Rani, R. (2020). Unravelling the role of rhizospheric plant-microbe synergy in phytoremediation: A genomic perspective. *Curr. Genomics* 21, 334–342. doi: 10.2174/1389202921999200623133240

Albert, V. A., Williams, S. E., and Chase, M. W. (1992). Carnivorous plants: phylogeny and structural evolution. *Science* 257, 1491–1495. doi: 10.1126/science.1523408

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26, 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x

Anderson, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics.* 62. doi: 10.1111/j.1541-0420.2005.00440.x

Baiser, B., Buckley, H. L., Gotelli, N. J., and Ellison, A. M. (2013). Predicting food-web structure with metacommunity models. *Oikos* 122, 492–506. doi: 10.1111/j.1600-0706.2012.00005.x

Bashir, I., War, A. F., Rafiq, I., Reshi, Z. A., Rashid, I., and Shouche, Y. S. (2022). Phyllosphere microbiome: Diversity and functions. *Microbiol. Res.* 254, 126888. doi: 10.1016/j.micres.2021.126888

Bastian, M., Heymann, S., and Jacomy, M. (2009). "Gephi: An Open Source Software for Exploring and Manipulating Networks." in *Proceedings of the International AAAI Conference on Web and Social Media*. 3, 361–362. doi: 10.1609/icwsm.v3i1.13937

Beneduzi, A., Ambrosini, A., and Passaglia, L. M. P. (2012). Plant growth-promoting rhizobacteria (PGPR): Their potential as antagonists and biocontrol agents. *Genet. Mol. Biol.* 35, 1044–1051. doi: 10.1590/S1415-47572012000600020

Bittleston, L. S., Wolock, C. J., Yahya, B. E., Chan, X. Y., Chan, K. G., Pierce, N. E., et al. (2018). Convergence between the microcosms of Southeast Asian and North American pitcher plants. *eLife* 7, e36741. doi: 10.7554/eLife.36741.023

Blekhman, R., Goodrich, J. K., Huang, K., Sun, Q., Bukowski, R., Bell, J. T., et al. (2015). Host genetic variation impacts microbiome composition across human body sites. *Genome Biol.* 16, 191. doi: 10.1186/s13059-015-0759-1

Boyer, T., and Carter, R. (2011). Community analysis of green pitcher plant (Sarracenia oreophila) bogs in Alabama. *Castanea* 76, 364–376. doi: 10.2179/10-048.1

Bray, J. R., and Curtis, J. T. (1957). An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* 27, 325–349. doi: 10.2307/1942268

Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A. D. (2019). rnaSPAdes: a *de novo* transcriptome assembler and its application to RNA-Seq data. *GigaScience* 8, giz100. doi: 10.1093/gigascience/giz100

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian J. Stat* 11, 265–270. Available online at: https://www.jstor.org/stable/4615964.

Chomicki, G., Burin, G., Busta, L., Gozdzik, J., Jetter, R., Mortimer, B., et al. (2024). Convergence in carnivorous pitcher plants reveals a mechanism for composite trait evolution. *Science* 383, 108–113. doi: 10.1126/science.ade0529

Clauset, A., Newman, M. E. J., and Moore, C. (2004). Finding community structure in very large networks. *Phys. Rev. E* 70, 066111. doi: 10.1103/PhysRevE.70.066111

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Ellison, A. M., Buckley, H. L., Miller, T. E., and Gotelli, N. J. (2004). Morphological variation in *Sarracenia purpurea* (Sarraceniaceae): geographic, environmental, and taxonomic correlates. *Am. J. Bot.* 91, 1930–1935. doi: 10.3732/ajb.91.11.1930

Ellison, A. M., Gotelli, N. J., Błędzki, L. A., and Butler, J. L. (2021). Regulation by the Pitcher Plant Sarracenia purpurea of the Structure of its Inquiline Food Web. *amid* 186, 1–15. doi: 10.1674/0003-0031-186.1.1

Ellison, A. M., Gotelli, N. J., Brewer, J. S., Cochran-Stafira, D. L., Kneitel, J. M., Miller, T. E., et al. (2003a).The evolutionary ecology of carnivorous plants. In: *Advances in Ecological Research* (Academic Press). Available online at: https://www.sciencedirect.com/science/article/pii/S0065250403330090 (Accessed March 16, 2024).

Ellison, A. M., Gotelli, N. J., Brewer, J. S., Cochran-Stafira, D. L., Kneitel, J. M., Miller, T. E., et al. (2003b). The evolutionary ecology of carnivorous plants. *Adv. Ecol. Res.* 33, 1–74. doi: 10.1016/S0065-2504(03)33009-0

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugenics* 7, 179–188. doi: 10.1111/j.1469-1809.1936.tb02137.x

Fu, C.-N., Wicke, S., Zhu, A.-D., Li, D.-Z., and Gao, L.-M. (2023). Distinctive plastome evolution in carnivorous angiosperms. *BMC Plant Biol.* 23, 660. doi: 10.1186/s12870-023-04682-1

Furches, M. S., Small, R. L., and Furches, A. (2013). Hybridization leads to interspecific gene flow in Sarracenia (Sarraceniaceae). *Am. J. Bot.* 100, 2085–2091. doi: 10.3732/ajb.1300038

Gautam, A., Zeng, W., and Huson, D. H. (2023).DIAMOND + MEGAN Microbiome Analysis. In: *Metagenomic Data Analysis* (New York, NY: Springer US) (Accessed May 6, 2024).

Ge, S. X., Jung, D., and Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36, 2628–2629. doi: 10.1093/bioinformatics/btz931

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53, 325–338. doi: 10.1093/biomet/53.3-4.325

Grothjan, J. J., and Young, E. B. (2019). Diverse microbial communities hosted by the model carnivorous pitcher plant Sarracenia purpurea: analysis of both bacterial and eukaryotic composition across distinct host plant populations. *PeerJ* 7, e6392. doi: 10.7717/peerj.6392

Grothjan, J. J., and Young, E. B. (2022). Bacterial recruitment to carnivorous pitcher plant communities: identifying sources influencing plant microbiome composition and function. *Front. Microbiol.* 13, 791079. doi: 10.3389/fmicb.2022.791079

Guimera, R., and Nunes Amaral, L. A. (2005). Functional cartography of complex metabolic networks. *nature* 433, 895–900. doi: 10.1038/nature03288

Heil, J. A., Wolock, C. J., Pierce, N. E., Pringle, A., and Bittleston, L. S. (2022). Sarracenia pitcher plant-associated microbial communities differ primarily by host species across a longitudinal gradient. *Environ. Microbiol.* 24, 3500–3516. doi: 10.1111/1462-2920.15993

Ke, J., Wang, B., and Yoshikuni, Y. (2021). Microbiome engineering: synthetic biology of plant-associated microbiomes in sustainable agriculture. *Trends Biotechnol.* 39, 244–261. doi: 10.1016/j.tibtech.2020.07.008

Kers, J. G., and Saccenti, E. (2022). The power of microbiome studies: some considerations on which alpha and beta metrics to use and how to report results. *Front. Microbiol.* 12, 796025. doi: 10.3389/fmicb.2021.796025

Kim, D., Song, L., Breitwieser, F. P., and Salzberg, S. L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res.* 26, 1721–1729. doi: 10.1101/gr.210641.116

Koopman, M. M., and Carstens, B. C. (2011). The microbial phyllogeography of the carnivorous plant sarracenia alata. *Microb. Ecol.* 61, 750–758. doi: 10.1007/s00248-011-9832-9

Koopman, M. M., Fuselier, D. M., Hird, S., and Carstens, B. C. (2010). The carnivorous pale pitcher plant harbors diverse, distinct, and time-dependent bacterial communities. *Appl. Environ. Microbiol.* 76, 1851–1860. doi: 10.1128/AEM.02440-09

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28, 3211–3217. doi: 10.1093/bioinformatics/bts611

Kruskal, W. H., and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* 47, 583–621. doi: 10.1080/01621459.1952.10483441

Lahti, L., and Shetty, S. (2018). *Introduction to the microbiome R package*. Available online at: https://microbiomegithubio/tutorialshttps://s3.jcloud.sjtu.edu.cn/899a892efef34b1b944a19981040f55b-oss01/bioconductor/3.10/bioc/vignettes/microbiome/inst/doc/vignette.html (Accessed December 18, 2023).

Louca, S., Jacques, S. M. S., Pires, A. P. F., Leal, J. S., Srivastava, D. S., Parfrey, L. W., et al. (2016). High taxonomic variability despite stable functional structure across microbial communities. *Nat. Ecol. Evol.* 1, 1–12. doi: 10.1038/s41559-016-0015

Louca, S., Polz, M. F., Mazel, F., Albright, M. B. N., Huber, J. A., O'Connor, M. I., et al. (2018). Function and functional redundancy in microbial systems. *Nat. Ecol. Evol.* 2, 936–943. doi: 10.1038/s41559-018-0519-1

Mahnert, A., Haratani, M., Schmuck, M., and Berg, G. (2018). Enriching beneficial microbial diversity of indoor plants and their surrounding built environment with biostimulants. *Front. Microbiol.* 9, 2985. doi: 10.3389/fmicb.2018.02985

Mahnert, A., Moissl-Eichinger, C., and Berg, G. (2015). Microbiome interplay: plants alter microbial abundance and diversity within the built environment. *Front. Microbiol.* 6. Available at: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2015.00887/full.

Malmberg, R. L., Rogers, W. L., and Alabady, M. S. (2018). A carnivorous plant genetic map: pitcher/insect-capture QTL on a genetic linkage map of Sarracenia. *Life Sci. Alliance* 1. doi: 10.26508/lsa.201800146

McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS One* 8, e61217. doi: 10.1371/journal.pone.0061217

Morales Moreira, Z. P., Chen, M. Y., Yanez Ortuno, D. L., and Haney, C. H. (2023). Engineering plant microbiomes by integrating eco-evolutionary principles into current strategies. *Curr. Opin. Plant Biol.* 71, 102316. doi: 10.1016/j.pbi.2022.102316

Munch, K., Boomsma, W., Willerslev, E., and Nielsen, R. (2008). Fast phylogenetic DNA barcoding. *Phil Trans. R Soc. B* 363, 3997–4002. doi: 10.1098/rstb.2008.0169

ONT 16S workflow 16S Workflow (2019). *EPI2ME WIMP workflow: quantitative, real-time species identification from metagenomic samples* (Oxford Nanopore Technologies). Available online at: https://nanoporetech.com/resource-centre/epi2me-wimp-workflow-quantitative-real-time-species-identification-metagenomic (Accessed February 23, 2024).

Patz, S., Gautam, A., Becker, M., Ruppel, S., Rodríguez-Palenzuela, P., and Huson, D. (2021). PLaBAse: A comprehensive web resource for analyzing the plant growth-promoting potential of plant-associated bacteria. 2021.12.13.472471. doi: 10.1101/2021.12.13.472471

Patz, S., Rauh, M., Gautam, A., and Huson, D. H. (2024). mgPGPT: Metagenomic analysis of plant growth-promoting traits. *bioRxiv 2021.12.13.472471*. doi: 10.1101/2024.02.17.580828

Saeed, Q., Xiukang, W., Haider, F. U., Kučerik, J., Mumtaz, M. Z., Holatko, J., et al. (2021). Rhizosphere bacteria in plant growth promotion, biocontrol, and bioremediation of contaminated sites: A comprehensive review of effects and mechanisms. *Int. J. Mol. Sci.* 22, 10529. doi: 10.3390/ijms221910529

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12, R60. doi: 10.1186/gb-2011-12-6-r60

Shannon, C. E. (2001). A mathematical theory of communication. *SIGMOBILE Mob Comput. Commun. Rev.* 5, 3–55. doi: 10.1145/584091.584093

Thorogood, C. J., Bauer, U., and Hiscock, S. J. (2018). Convergent and divergent evolution in carnivorous pitcher plant traps. *New Phytol.* 217, 1035–1041. doi: 10.1111/nph.14879

Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T., and Singh, B. K. (2020). Plant-microbiome interactions: from community assembly to plant health. *Nat. Rev. Microbiol.* 18, 607–621. doi: 10.1038/s41579-020-0412-1

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., et al. (2009). A core gut microbiome in obese and lean twins. *Nature* 457, 480–484. doi: 10.1038/nature07540

Volvenko, I. V. (2014). General patterns of spatial distribution of the integral characteristics of benthic macrofauna of the northwestern pacific and biological structure of ocean. *Open J. Ecol.* 4, 196–213. doi: 10.4236/oje.2014.44020

Wen, T., Xie, P., Yang, S., Niu, G., Liu, X., Ding, Z., et al. (2022). ggClusterNet: An R package for microbiome network analysis and modularity-based multiple network layouts. *iMeta* 1, e32. doi: 10.1002/imt2.32

Wheeler, G. L., and Carstens, B. C. (2018). Evaluating the adaptive evolutionary convergence of carnivorous plant taxa through functional genomics. *PeerJ* 6, e4322. doi: 10.7717/peerj.4322

Whitham, T. G., Bailey, J. K., Schweitzer, J. A., Shuster, S. M., Bangert, R. K., LeRoy, C. J., et al. (2006). A framework for community and ecosystem genetics: from genes to ecosystems. *Nat. Rev. Genet.* 7, 510. doi: 10.1038/nrg1877

Whitham, T. G., DiFazio, S. P., Schweitzer, J. A., Shuster, S. M., Allan, G. J., Bailey, J. K., et al. (2008). Extending genomics to natural communities and ecosystems. *Science* 320, 492–495. doi: 10.1126/science.1153918

Wilcoxon, F. (1992).Individual Comparisons by Ranking Methods. In: *Breakthroughs in Statistics* (New York, NY: Springer New York). Available online at: http://link.springer.com/10.1007/978-1-4612-4380-9_16 (Accessed December 18, 2023).

Yan, L. (2021). *ggvenn: Draw Venn Diagram by 'ggplot2.' R Package Version 19.*

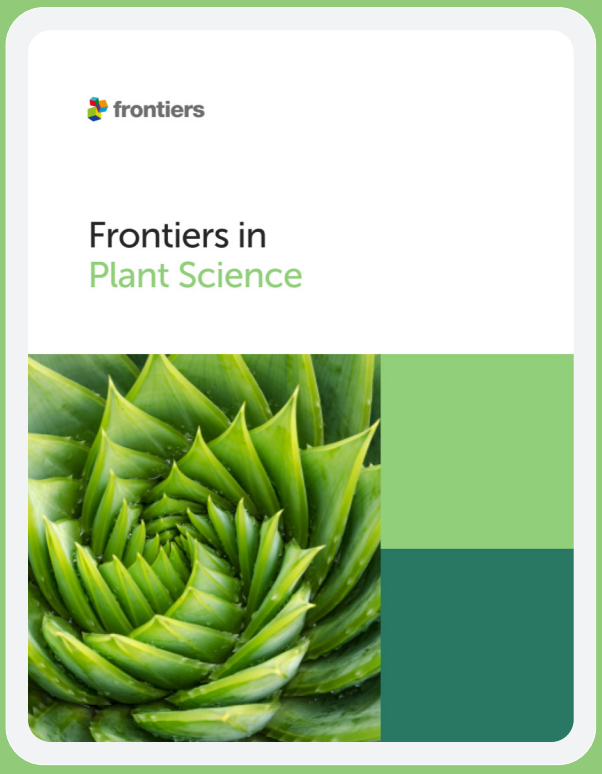
Frontiers in Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact