Check for updates

Received: 23 April 2021 | Accepted: 12 August 2021

DOI: 10.1002/ajb2.1775

RESEARCH ARTICLE

American Journal of Botany

Plastid phylogenomics of the Gynoxoid group (Senecioneae, Asteraceae) highlights the importance of motif-based sequence alignment amid low genetic distances

Belen Escobari[1,2] | **Thomas Borsch**[1,3] | **Taylor S. Quedensley**[4] | **Michael Gruenstaeudl**[3]

[1]Botanischer Garten und Botanisches Museum Berlin, Freie Universität Berlin, Berlin 14195, Germany

[2]Herbario Nacional de Bolivia, Universidad Mayor de San Andres, Casilla, La Paz, 10077, Bolivia

[3]Institut für Biologie, Systematische Botanik und Pflanzengeographie, Freie Universität Berlin, Berlin 14195, Germany

[4]Department of Biology, Texas Christian University, Fort Worth, TX 76109, USA

Correspondence

Michael Gruenstaeudl, Institut für Biologie, Systematische Botanik und Pflanzengeographie, Freie Universität Berlin, Berlin 14195, Germany.
Email: m.gruenstaeudl@fu-berlin.de

Abstract

Premise: The genus *Gynoxys* and relatives form a species-rich lineage of Andean shrubs and trees with low genetic distances within the sunflower subtribe Tussilaginineae. Previous molecular phylogenetic investigations of the Tussilagininеае have included few, if any, representatives of this Gynoxoid group or reconstructed ambiguous patterns of relationships for it.

Methods: We sequenced complete plastid genomes of 21 species of the Gynoxoid group and related Tussilaginineae and conducted detailed comparisons of the phylogenetic relationships supported by the gene, intron, and intergenic spacer partitions of these genomes. We also evaluated the impact of manual, motif-based adjustments of automatic DNA sequence alignments on phylogenetic tree inference.

Results: Our results indicate that the inclusion of all plastid genome partitions is needed to infer well-supported phylogenetic trees of the Gynoxoid group. Whole plastome-based tree inference suggests that the genera *Gynoxys* and *Nordenstamia* are polyphyletic and form the core clade of the Gynoxoid group. This clade is sister to a clade of *Aequatorium* and *Paragynoxys* and also includes some but not all representatives of *Paracalia*.

Conclusions: The concatenation and combined analysis of all plastid genome partitions and the construction of manually-curated, motif-based DNA sequence alignments are found to be instrumental in the recovery of well-supported relationships of the Gynoxoid group. We demonstrate that the correct assessment of homology in genome-level plastid sequence data sets is crucial for subsequent phylogeny reconstruction and that the manual post-processing of multiple sequence alignments improves the reliability of such reconstructions amid low genetic distances between taxa.

KEYWORDS

Asteraceae, chloroplast genome, homoplasy, manual sequence alignment, noncoding DNA, phylogenetic inference, sequence partitioning, South America

The Andean region of northwestern South America is one of the most prominent regions of plant diversification in the neotropics (Luebert and Weigend, 2014) and a common location for plant radiations in sunflowers (Asteraceae; e.g., Vargas et al., 2017; Pouchon et al., 2018). The Gynoxoid group of the sunflower tribe Senecioneae is one such radiation (Vision and Dillon, 1996; Lundin, 2006). The group comprises 150–170 species of shrubs and trees that primarily inhabit high-elevation habitats in the northern and central Andes (Nordenstam et al., 2009). Under the current taxonomic circumscription, the species of the Gynoxoid group are separated into five closely related genera [i.e., *Aequatorium* B. Nord., *Gynoxys* Cass.,

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2021 The Authors. *American Journal of Botany* published by Wiley Periodicals LLC on behalf of Botanical Society of America.

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Nordenstamia Lundin, *Paracalia* Cuatrec., and *Paragynoxys* (Cuatrec.) Cuatrec.]. *Gynoxys* contains the majority of species in the group (Cuatrecasas, 1951; Nordenstam, 2007; Beck and Ibáñez, 2014), and novel species continue to be described (e.g., Beltran and Campos de la Cruz, 2009). Biogeographically, the group ranges from northern Venezuela to northern Argentina, and most of its species are characteristic elements of montane forests or solitary shrubs and trees in the paramo (Tinoco et al., 2013), while only a few exhibit a scandent growth form and inhabit lower-elevation montane forests (Beck and Ibáñez, 2014; Figure 1). Most of the species of the Gynoxoid group are restricted to relatively small distribution ranges and occur in habitats threatened by anthropogenic land use and climate change, thus making this a group of conservation concern (Beltran et al., 2006; Morillo and Briceno, 2000; Hind, 2007). Based on the most recent phylogenetic investigations of the Senecioneae, the Gynoxoid group is part of the subtribe Tussilagininae (Pelser et al., 2007, 2010), which represents one of four subtribes of the Senecioneae.

The phylogenetic relationships and species limits within the Gynoxoid group are poorly understood because few molecular phylogenetic investigations have included taxa of this group or focused on aspects other than their relationships. Kadereit and Jeffrey (1996) conducted a study on the phylogeny of the Senecioneae using chloroplast restriction site data and included one species of *Gynoxys* in their data set. Their results indicated that *Gynoxys* was most closely related to *Tussilago* L., *Roldana* La Llave, and *Brachyglottis* J.R. Forst. & G. Forst. Similarly, Pelser et al. (2007) aimed to infer relationships within the Senecioneae using the internal transcribed spacer (ITS) region and recovered a clade comprising the genera *Aequatorium*, *Gynoxys*, *Nordenstamia*, and *Paragynoxys*, but with mixed levels of branch support. Their results indicated that the current taxonomic circumscriptions within the Gynoxoid group were not fully substantiated by DNA sequence data, as *Nordenstamia* was found nested within a paraphyletic *Gynoxys*. Pelser et al. (2010) extended their previous taxon sampling of the Gynoxoid group by including one sample of *Paracalia* and recovered the group as monophyletic. Recently, Quedensley et al. (2018) included taxa of *Aequatorium* and *Gynoxys* in a study of North and Central American representatives of the Tussilagininеae and recovered a monophyletic Gynoxoid group with strong statistical support; however, no insight

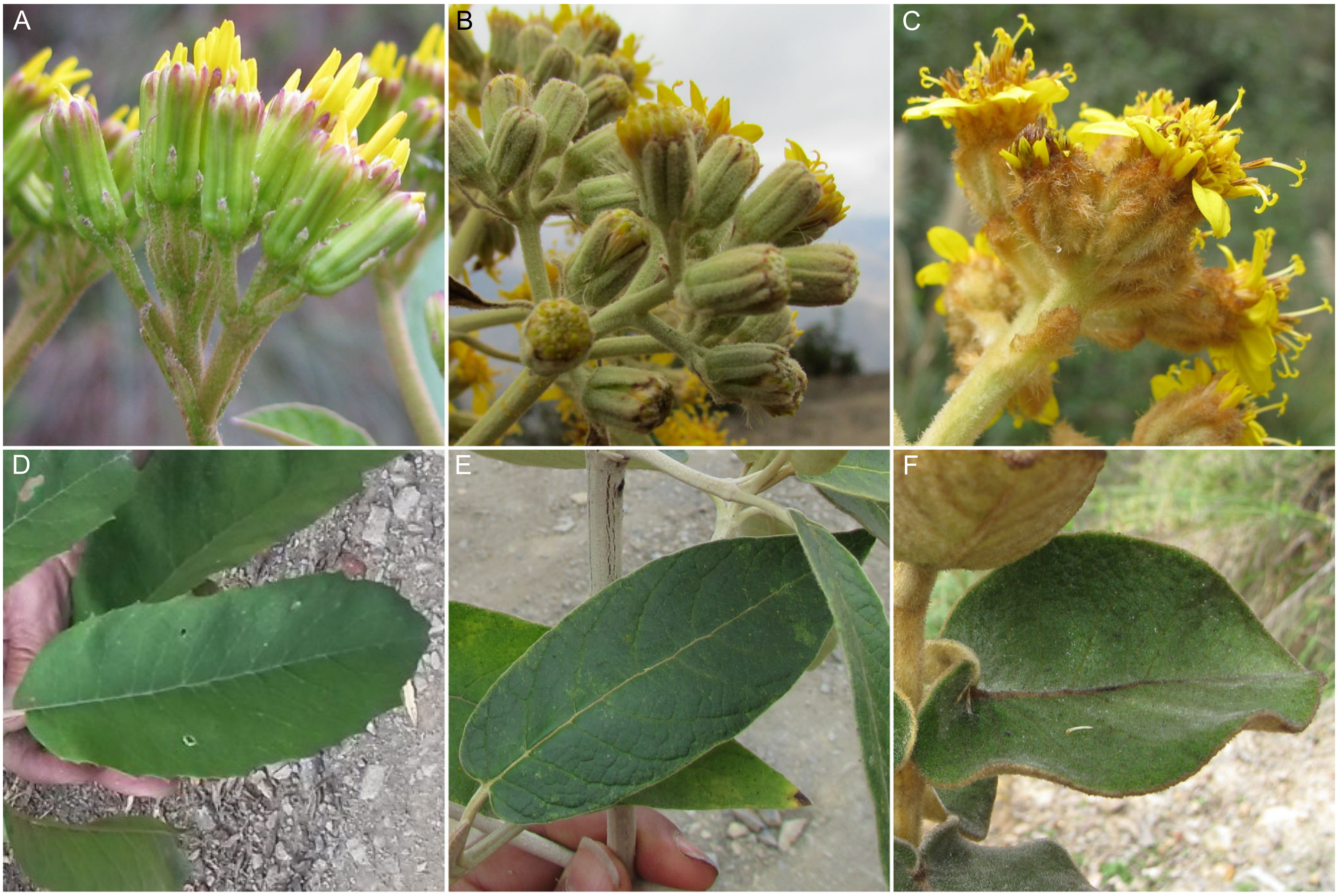

FIGURE 1 Morphological variability among three species of the Gynoxioid group. Displayed are the leaves and capitulescences of *Nordenstamia repanda* (A and D), which represents a characteristic element of upper cloud forests (bosque yungueño de ceja de monte); *Gynoxys asterotricha* (B and E), which represents a characteristic element of lower cloud forests; and *Gynoxys tomentosissima* (C and F), which represents a characteristic element of low montane forests

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

into the intergeneric relationships of the group was investigated. In summary, the intergeneric relationships of the Gynoxoid group have remained largely unresolved, and the delimitation of its genera is mostly unsubstantiated by phylogenetic methods.

Plastid phylogenomic studies have been shown to be efficient in resolving the phylogenetic relationships of species groups in the Asteraceae. Vargas et al. (2017) investigated the relationships of *Diplostephium* Kunth and related genera that exemplified low levels of molecular variability (Vargas and Madrinan, 2012). The authors sequenced complete plastid genomes of 14 different genera (91 samples) and inferred a highly resolved phylogeny. Similarly, Pouchon et al. (2018) sequenced plastid genomes in an analysis of the phylogenetic relationships of *Espeletia* Mutis ex Bonpl. and relatives, which previous studies could not resolve due to an insufficient number of informative DNA sequence characters available. Specifically, the authors sequenced complete plastid genomes for 41 species among eight genera and recovered well-supported clades. Zhang et al. (2019) sequenced plastid genomes to establish a robust phylogenetic framework for the species-rich genus *Saussurea* DC., for which previous work had generated conflicting infrageneric classifications. By analyzing complete plastid genomes of 136 species of *Saussurea*, the authors found that approximately 2000 parsimony informative sites were needed to produce a resolved and well-supported phylogeny. More recently, Knope et al. (2020) sequenced plastid genomes to reconstruct the phylogeny of Hawaiian endemics of the genus *Bidens* L. in light of a decade-long effort to clarify the evolutionary history of this rapidly radiating lineage and were able to generate a highly supported phylogeny. Evidently, the use of plastid genomes for phylogenetic inference can be instrumental in clarifying the relationships of plant groups that exhibit low molecular variability. This utility is not restricted to Asteraceae but has been demonstrated in numerous lineages of flowering plants (e.g., Givnish et al., 2018; Yao et al., 2019).

The process of sequencing and comparing plastid genomes for the reconstruction of phylogenetic relationships is often perceived as simple, but several studies have indicated that the evolution and structure of plastid genomes and, by extension, their application in phylogenetic inference, is complex. Based on observations that the majority of plastid genomes display strong structural conservation (Mower and Vickrey, 2018), are uniparentally (mostly maternally) inherited (Greiner et al., 2015), and do not experience biparental recombination (Marechal and Brisson, 2010), many investigations have operated under the assumption of a congruent phylogenetic signal across the entire plastid genome. However, several studies have cast doubt on the validity of this assumption and instead highlighted the presence of a discordant phylogenetic signal across different regions of the plastid genome. Investigations on the more variable regions of plastid genomes, for example, found mosaic-like patterns of molecular evolution (Borsch and Quandt, 2009), a hierarchical structure of phylogenetic signal (Müller et al., 2006; Barniske et al., 2012), and lineage-specific lengths and positions of these regions (Korotkova et al., 2014). Recent phylogenomic investigations corroborated these reports by identifying considerable phylogenetic incongruence across different regions of the plastid genome, which may result in inefficient or even incorrect phylogenetic inferences if the entire genome is analyzed under the same model parameters. Goncalves et al. (2019), for example, identified significant incongruence among the gene and species trees of different plastome regions in a phylogenomic study on rosids. The authors reported that the concatenation of all plastid coding regions produced highly supported phylogenies that were nonetheless incongruent to individual plastid gene trees. Similarly, Gruenstaeudl (2019) detected phylogenetic incongruence across different loci of the plastid genome in a phylogenomic investigation of water lilies and relatives. Walker et al. (2019) reported gene tree conflict among various plastid genes based on a broad sampling of angiosperm plastid genomes and noted numerous strongly supported but conflicting nodes between different gene trees. Similar observations were made in plastid phylogenomic analyses of Fabaceae (Zhang et al., 2020) and Bignoniaceae (Thode et al., 2020). Furthermore, Koehler et al. (2020) identified and ranked plastid genome regions by phylogenetic informativeness and found topological incongruence between the phylogenetic trees inferred from complete plastid genome sequences and those inferred from the five and 10 most variable plastid regions only. Based on these and similar studies, Goncalves et al. (2020) cautioned that "one or a few genes that have high phylogenetic signal may bias the inference" (Goncalves et al., 2020, p. 4) and, thus, recommended the continued exploration of more phylogenetic information among different regions of the plastid genome.

The positional homology among the nucleotides of a multiple sequence alignment (MSA) represents an essential aspect of phylogenetic tree inference and has not received sufficient attention in many phylogenomic investigations. Early plastid phylogenomic studies primarily employed the coding regions of the genomes for phylogenetic reconstruction (e.g., Leebens-Mack et al., 2005; Moore et al., 2010), which are largely conserved in their length and sequence and, thus, require relatively little adjustment upon standard MSA. Phylogenetic investigations that employ noncoding plastid DNA, by contrast, routinely inspect software-generated MSAs and adjust them according to the criterion of explicit sequence motifs to ensure correct positional homology (Kelchner, 2000; Loehne and Borsch, 2005; Morrison, 2006). In phylogenomic analyses, such adjustments are sometimes dismissed as impractical due to the large amounts of sequence data involved (Wu et al., 2012), but numerous investigations have demonstrated the impact of alignment errors on phylogenetic inference (reviewed by Wong et al., 2008). To reduce this impact while simultaneously avoiding time-expensive evaluations of the MSAs, many studies tend to automatically

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

exclude nucleotide positions in software-generated MSAs that are deemed unreliable (e.g., Bellot et al., 2020). In practice, such procedures may go as far as excluding alignment positions that exhibit a gap in only one of the aligned sequences, which substantially reduces the proportion of genome sequence employed for phylogenetic reconstruction (e.g., Gernandt et al., 2018). Along with the reduction of potential informativeness, such exclusions do not guarantee correct positional homology in the remaining alignment, as demonstrated for cases of small genomic inversions (Ochoterena, 2008). Given the larger share of noncoding compared to coding DNA in plastid genomes as well as the higher frequency of substitutions and microstructural mutations in noncoding DNA, manual adjustments of software-derived MSAs may even have a considerable impact on plastid phylogenomic reconstruction. In fact, in species groups with low genetic distances, a large proportion of potentially informative sequence characters will likely be encoded in the noncoding regions of the plastid genome.

The present investigation had two goals: (1) to infer the phylogenetic relationships within the Gynoxoid group of the Tussilagininеae using complete plastid genomes to account for the low genetic distances expected within this group, and (2) to assess the phylogenetic signal of different sections of the plastid genome, particularly with regard to motif-based adjustments of software-generated MSAs. To achieve these goals, we asked four questions: (1) Does phylogenetic analysis of complete plastid genomes yield resolved and well-supported phylogenetic trees for the Gynoxoid group? (2) Do different partitions of the plastid genome (i.e., coding sequences, intergenic spacers, and introns) support different phylogenetic hypotheses? (3) Does the manual adjustment of MSAs have a measurable impact on phylogenetic reconstruction? (4) Are the results of the plastome-based reconstructions congruent with the current generic classification of the Gynoxoid group? To address these questions, complete plastid genomes of 17 species of the Gynoxoid group and four species of closely related members of the Tussilagininеae were sequenced and annotated. These taxa form a taxon set that encompasses the typical genetic distances within the Gynoxoid group and to other members of the Tussilagininеae. We then used this data set to conduct phylogenetic tree inference based on different coding regions, introns, and intergenic spacers of the plastid genome before and after the manual adjustment of the MSAs and contrast the results.

MATERIALS AND METHODS

Taxon sampling and DNA extraction

A total of 21 samples of different species of the Tussilagininеae were collected for DNA extraction and subsequent plastid genome sequencing (Table 1). Of these, 17 samples represent genera of the Gynoxoid group (i.e., *Aequatorium*, *Gynoxys*, *Nordenstamia*, *Paracalia*, and *Paragynoxys*), while four represent other genera of the Tussilagininеae that form the sister clade to the Gynoxoid group (i.e., *Arnoglossum* Raf., *Roldana* La Llave, and *Telanthophora* H. Rob. & Brettell; see Quedensley et al. [2018] for details). Particular emphasis in our taxon sampling was placed on the inclusion of (1) more than one species per genus, where possible, to approximate the genetic variability within each genus, (2) multiple species of *Gynoxys* to represent its species diversity across the Andes and to accommodate previous results indicating that *Gynoxys* may be non-monophyletic, and (3) the type species of the genera *Gynoxys*, *Nordenstamia*, and *Paracalia* to enable comparisons of our reconstructions with the current taxonomic classification of the Gynoxoid group. We included the previously published plastid genome of *Ligularia fischeri* (Ledeb.) Turcz. (GenBank accession NC_039352; Chen et al., 2018) as an outgroup. The taxon names and generic concepts that we applied follow those of Nordenstam (2007). Herbarium vouchers of all newly sequenced samples of the Gynoxoid group were deposited in B, with duplicates in LPB, USM, HSP, or HUT.

Total genomic DNA was extracted from silica gel-dried leaf material using the NucleoSpin Plant II kit (Macherey-Nagel, Dueren, Germany) or from herbarium specimens using the CTAB DNA isolation method as modified by Borsch et al. (2003). A DNA sample for each specimen was deposited in the DNA bank of the Botanic Garden and Botanical Museum Berlin (BGBM; Table 1). Unless DNA isolates were fragmented due to age, samples were sheared to an average fragment size of 600 bp using a Covaris S220 sonicator (Covaris, Woburn, MA, USA). Upon shearing, all fragments between 400 and 900 bp were selected and maintained by applying the BluePippin protocol (Sage Science, Beverly, MA, USA). The final concentration of DNA samples was measured using Qubit 2.0 Fluorometer dsDNA BR Assay kits (Life Technologies-Thermo Fisher Scientific, Saint Aubin, France) and the final fragment size distribution using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Manufacturer protocols were followed for all steps of DNA extraction.

Genomic library preparation and DNA sequencing

Plastid genomes were sequenced via a genome skimming approach following the preparation of genomic libraries. For each DNA sample, a barcoded genomic library was constructed using the TruSeq DNA library preparation kit (Illumina, San Diego, CA, USA). Standard indexing adapters were ligated to the fragment ends to generate single-index libraries. Libraries were validated via qPCR on a Mastercycler ep realplex (Eppendorf AG, Hamburg, Germany) using the KAPA library quantification kit (KAPA Biosystems, Wilmington, MA, USA). Following qPCR, indexed DNA libraries were normalized and pooled in equal volumes. Pooled libraries were sequenced as paired-end

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Species name, collection location and number, herbarium voucher information, and GenBank accession number for each plastid genome sequenced in this investigation. n.a. = not applicable

Species	Taxonomic authority	Location	Collection no.	Herbarium voucher	DNA Bank code	GenBank
Aequatorium jamesonii	(S.F. Blake) C. Jeffrey	Ecuador: Napo, Parque Nacional Llanganates	Vargas H. et al. 2594	MO-2940002 (MO)	DB38638	MT528247
Arnoglossum atriplicifolium	(L.) H.Rob.	USA: Spring LakePark, Omaha, Nebraska	Quedensley T.S. s.n.	TSQ2011cp001 (TEX)	n.a.	MK170176
Gynoxys asterotricha	Sch. Bip.	Bolivia: Parque Nacional Madidi, Laji Sorapata	Escobari B. et al. 36	B-100720926 (LPB,B)	DB27411	MK044798
Gysnoxys baccharoides	(Kunth) Cass.	Ecuador: Azuay, Parque Nacional Cajas	Jorgensen P. et al. 1616	MO-1879904 (AAU,MO)	DB38633	MT528244
Gynoxys ignaciana	Cuatrec.	Ecuador: Loja, Guararas	Emperaire L. 1318	P-03833291 (P)	DB38767	MT528252
Gynoxys longifolia	Wedd.	Peru: Arequipa, Cailloma	Beck S. 26384	LPB0002593 (LPB,MO,US)	DB38565	MT528245
Gynoxys mandonii	Sch. Bip. ex Rusby	Bolivia: La Paz, Nor Yungas, Unduavi	Cayola L. 5570	MO-3151322 (LPB,MO)	DB27416	MK056106
Gynoxys megacephala	Rusby	Bolivia: Nor Yungas, Ecovia	Escobari B. et al. 46	LPB0002594 (LPB)	DB27426	MN328892
Gynoxys sp. SEN301		Peru: Amazonas, Leymebamba	Escobari B. et al. 593	B-101098545 (USM,HUT,B)	DB38744	MN328891
Gynoxys tomentosissima	Cuatrec.	Peru: Amazonas, Leymebamba	Escobari B. et al. 590	B-101098697 (USM,HUT,B)	DB38742	MN328890
Gynoxys violacea	Sch. Bip. ex Wedd.	Venezuela: Merida	Walter E. 166	B-101139853 (B)	DB38780	MT528243
Nordenstamia cajamarcensis	(H. Rob. & Cuatrec.) B. Nord.	Peru: Pasco, Oxapampa	Castillo G. 460	MO-2940516 (F,MO,USM)	DB38626	MT528248
Nordenstamia kingii	(H. Rob. & Cuatrec.) B.Nord.	Bolivia: Cochabamba, Monte Puncu	Solomon J. 18067	LPB0002595 (LPB,MO)	DB38764	MT528249
Nordenstamia repanda	(Wedd.) Lundin	Bolivia: Parque Nacional Madidi, Laji Sorapata	Cayola L. 5585	LPB0002597 (LPB,MO)	DB27409	MK086040
Paracalia jungioides	(Hook. & Arn.) Cuatrec.	Peru: Ancash, Huaraz	Diaz C. 1975	MO-3030471 (LPB,MO)	DB43895	MT528251
Paracalia pentamera	(Cuatrec.) Cuatrec.	Bolivia: Sud Yungas, Chulumani	Gallegos S. 3850	LPB0002596 (LPB)	DB40913	MT942595
Paragynoxys martingrantii	(Cuatrec.) Cuatrec.	Colombia: Cesar	Gentry A. 79156	MO-1962013 (MO)	DB43896	MT528250
Paragynoxys venezuelae	(V.M. Badillo) Cuatrec.	Venezuela: Merida, Paramo de Las Coloradas	Cuatrecasas J. et al. 28996	MA-01-00896838 (MA,S)	DB43902	MT528246
Roldana aschenborniana	(Schauer) H. Rob. & Brettell	Mexico: Oaxaca (cult. ex-situ in California)	Quedensley T.S. s.n.	TSQ2011cp002 (TEX)	n.a.	MK170177
Roldana barba-johannis	(DC.) H. Rob. & Brettell	Mexico: Oaxaca (cult. ex-situ in California)	Quedensley T.S. s.n.	TSQ2011cp003 (TEX)	n.a.	MK170178
Telanthophora grandifolia	(Less.) H. Rob. & Brettell	Mexico: Oaxaca (cult. ex-situ in California)	Quedensley T.S. s.n.	TSQ2011cp004 (TEX)	n.a.	MK170179

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

reads either on an Illumina MiSeq or an Illumina HiSeq X platform. Sequencing was performed in-house at the Berlin Center for Genomics in Biodiversity Research (Berlin, Germany) or the Genome Sequencing and Analysis Facility of the University of Texas at Austin, or outsourced to Macrogen (Seoul, Republic of Korea).

Genome assembly and annotation

After DNA sequencing, raw sequence reads were filtered for quality and successful pairing using scripts 1 and 2 of the pipeline of Gruenstaeudl et al. (2018), followed by a mapping of the quality-filtered reads to the plastid genome of *Jacobaea vulgaris* Gaertn. (accession NC_015543; Doorduin et al., 2011) to extract plastome reads using Bowtie2 v.2.3.4 (Langmead and Salzberg, 2012). Contigs were assembled de novo with either IOGA v.20160908 (Bakker et al., 2015) or NOVOPlasty v.2.7.2 (Dierckxsens et al., 2017) based on the subset of reads that mapped against the reference genome, using a range of different kmer values to optimize contig length (kmer = 33–97, in increments of 4). Unless already circular, final contigs were circularized manually with Geneious v.11.1.4 (Kearse et al., 2012), using the plastid genome of *Jacobaea vulgaris* as a reference for contig position and orientation. A circular, quadripartite structure of the plastid genome and equality of its inverted repeat (IR) regions were confirmed for each assembly through a blast search against itself using script 4 of the pipeline of Gruenstaeudl et al. (2018). Sequence ambiguities in the final assembly, if present, were resolved by mapping the quality-filtered reads against the circularized assembly using Bowtie2. Final assemblies were annotated via the annotation server DOGMA (Wyman et al., 2004), followed by a manual inspection and, where necessary, correction of the annotations in Geneious. Specifically, sequence annotations were corrected regarding the presence of start and stop codons, the absence of internal stop codons, and their length as a multiple of three for each coding region. This process resulted in a complete and fully annotated plastid genome for each taxon sampled for this study. Upon annotation, all new plastid genomes were deposited to GenBank; their accession numbers are listed in Table 1. Plastome maps were drawn with OGDRAW v.1.3.1 (Greiner et al., 2019).

Data partitioning, sequence alignment, and alignment adjustments

The coding and noncoding regions of the plastid genomes were extracted and aligned using a four-step procedure. First, one of the IRs was removed from each genome to avoid redundancy among the extracted loci. Second, all coding and noncoding regions (except tRNAs and rRNAs) were excised bioinformatically from each genome using script 9 of the pipeline of Gruenstaeudl et al. (2018) and then grouped by region name. Third, all sequences of the same region were aligned into preliminary MSAs using MAFFT v.7.394 (Katoh and Standley, 2013) under the default settings of the software. Specifically, MSAs of 81 coding regions, 20 introns, and 111 intergenic spacers, each consisting of the sequences of 22 taxa, were constructed. Fourth, these preliminary MSAs were evaluated by eye and, where necessary, adjusted manually to improve positional homology across nucleotides using PhyDE v.0.9971 (Müller et al., 2010). The adjustments followed the rules of Loehne and Borsch (2005) and included the masking of mutational hotspots for those sections of the MSAs where correct positional homology could not be established. This motif-based alignment approach was based on the assumption that insertions, deletions or inversions of genomic regions do not occur at random, but exhibit recurrent patterns (similar to those found in simple sequence repeats or hairpin-mediated inversions; e.g., Kelchner, 2002; Borsch and Quandt, 2009) and are often caused by structural and functional constraints (invoked, for example, during DNA replication and repair; Smith and Keeling, 2015). Such microstructural mutations may simultaneously encompass multiple nucleotides, contradicting earlier assumptions of a fifth character state per gap position (Barriel, 1994). When adjusting a MSA, microstructural mutations can be expressed by placing inserted elements in their own alignment columns, creating biologically meaningful gaps that can be utilized during indel coding (Simmons and Ochoterena, 2000). Occasionally, microstructural mutations are so frequent within a region that they create overlapping mutations for which positional homology can no longer be established by eye. Such cases are particularly common for mono- or dinucleotide microsatellites (e.g., poly-A repeats) where positional homology is often obscured by the short unit length. The evolution of plastid microsatellites has, thus, been reported as highly homoplastic (e.g., Tesfaye et al., 2007), involving insertions and deletions of one to several repeat units rather than following a stepwise model. Consequently, we excluded regions of uncertain homology from the process of indel coding and tree reconstruction in this investigation. Small sequence inversions, by comparison, were masked through a manual re-inversion of the sequence motifs, followed by a re-alignment to the other sequences and the recording of each inversion as a single-step event that was later added to the indel matrix as a binary character. If such inversions were left unchanged, the presence of incorrect nucleotide substitutions instead of the inversion itself would be implied, resulting in the loss of a relevant phylogenetic character (discussed by Loehne and Borsch, 2005). Examples of the masking of microsatellites and sequence inversions are illustrated in Figure 2; a summary of the positions and lengths of masked sequence regions within the MSAs is given in Appendix S1.

Following these alignment adjustments, all MSAs were assessed for length and sequence variability and excluded from the data set if any of the following criteria were met: lack of sequence variability, length less than 10 bp, more ambiguous than variable nucleotides if length less than 50 bp, and vicinity to trans-spliced genes. For example, the MSAs of the intergenic spacers *ndhH–ndhA*, *rpoB–rpoC1*, *ndhK–psbG*, and *psbF–psbE* were excluded because they

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

were only 1, 5, 7, and 9 bp long, respectively. Similarly, the spacer *psbT–psbN* was excluded due to the combination of a short alignment length and a higher number of ambiguous than variable nucleotides. All tRNA and rRNA genes were excluded due to minimal, if any, sequence variability. The intergenic spacers adjacent to *rps12* (i.e., *rpl20–rps12*, *rps12–clpP*, and *rps7–rps12*) were excluded because *rps12* comprises discontinuous group II introns that are often associated with complex secondary DNA structures and may bias the mutational dynamics of the intergenic spacers flanking the trans-spliced exons of this gene (Kelchner, 2002; Glanz and Kueck, 2009). All other MSAs were saved as NEXUS files for further processing. Insertions and deletions in each MSA were coded as binary characters using the simple indel coding (SIC) scheme of Simmons and Ochoterena (2000) as implemented in SeqState v.1.4.1 (Müller, 2005). The complete sets of MSAs representing 81 different coding regions, 20 different introns, and 103 different intergenic spacers were grouped by marker class (hereafter, plastid partitions; i.e., coding regions, introns, and intergenic spacers) and then concatenated with and without indel codes. Specifically, the MSAs were concatenated within each set as well as across the three sets, both with and without the presence of indel codes, generating a total of eight different matrices. All matrices were deposited in Zenodo (https://zenodo.org/record/4428211) and employed for phylogenetic tree reconstruction. For brevity in our analyses, coding regions are abbreviated with "CDS", introns with "INT", and intergenic spacers with "IGS". Similarly, the concatenation of all MSAs of the coding regions is abbreviated as "81 CDS concat", the concatenation of all MSAs of the introns as "20 INT concat", and the concatenation of all MSAs of the intergenic spacers as "103 IGS concat".

Alignment metrics and sequence variability

To assess the reliability of the MSAs and the impact of manual alignment adjustment on them, we calculated a total of eight different alignment metrics for each MSA before and after alignment adjustment. The inferred metrics were (1) alignment length, (2) GC content, (3) fraction of polymorphic sites, (4) fraction of parsimony-informative sites, (5–7) three homoplasy indices—(5) consistency index (CI; Kluge and Farris, 1969), (6) rescaled consistency index (RC), and (7) retention index (RI; both Farris, 1989) in their ensemble form, and (8) the largest uncorrected *p*-distance between all sequences as defined in equation 3.1 of Nei and Kumar (2000). Each metric was calculated in R (R Core Team, 2019) using the R packages ape v.5.2 (Paradis and Schliep, 2018) or phangorn v.2.4.0 (Schliep, 2011). For each MSA, the three homoplasy indices were calculated on the neighbor-joining tree of that MSA. The values of the homoplasy indices are negatively correlated with the level of homoplasy in the MSA, with high index values indicating low levels of homoplasy. Indels were taken into account during the calculation of alignment length and the fraction of polymorphic sites but were disregarded in the calculation of all other metrics in accordance with the original settings of the R functions. To quantify sequence variability across the plastid genomes under study, we calculated the nucleotide diversity index *π* (Nei and Li, 1979) with DnaSP v.6.12.03 (Rozas et al., 2017) for each MSA as well as their concatenation across all partitions using a sliding window algorithm with a step size of 200 bp and window size of 600 bp. To visualize sequence variability across the genomes, we generated variability plots using mVISTA (Frazer et al., 2004) following a global pairwise alignment of the sequences with LAGAN (Brudno et al., 2003). The IRa was excluded from each plastid genome before alignment and visualization with mVISTA; coding regions <10 nucleotides as well as the trans-spliced gene *rps12* were not annotated in the visualizations. The plastid genome of *Ligularia fischeri* was selected as a reference for the calculation of sequence similarity values in mVISTA.

Phylogenetic tree inference and hypothesis testing

Phylogenetic tree inference was conducted under the maximum likelihood (ML) and the Bayesian inference (BI) criterion on the concatenation of all plastid partitions before and after the manual adjustment of the MSAs. Tree inference under ML was also conducted on the independent concatenation of the coding regions, the introns, and the intergenic spacers, before and after the manual alignment adjustment. To infer the homoplasy indices, tree inference under ML was additionally conducted for each individual MSA before and after alignment adjustment. Tree inference under ML was performed using RAxML v.8.2.9 (Stamatakis, 2014), including the option for a thorough optimization of the best-scoring ML tree. Tree inference under BI was performed with MrBayes v.3.2.6 (Ronquist and Huelsenbeck, 2003) using four parallel Markov chain Monte Carlo (MCMC) runs for a total of 50 million generations. Branch support under ML was calculated through 1000 bootstrap (BS) replicates under the rapid BS algorithm (Stamatakis et al., 2008). Branch support under BI was calculated as posterior probability (PP) values. The nucleotide substitution model GTR + G + I was applied by default to model nucleotide substitution rates during tree inference under both optimality criteria. For indel characters, an F81-like binary substitution model with a gamma-shaped rate variation across sites was employed under both optimality criteria (Lewis, 2001). In analyses under BI, the sampling of independent generations and the convergence of the Markov chains were confirmed in Tracer v.1.7 (Rambaut et al., 2018); the initial 50% of all MCMC trees were discarded as burn-in, and post-burn-in trees were summarized as a 50% majority rule consensus tree. The significance of the topological differences between inferred trees was evaluated with the approximately unbiased (AU) test (Shimodaira, 2002). Specifically, we compared the

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

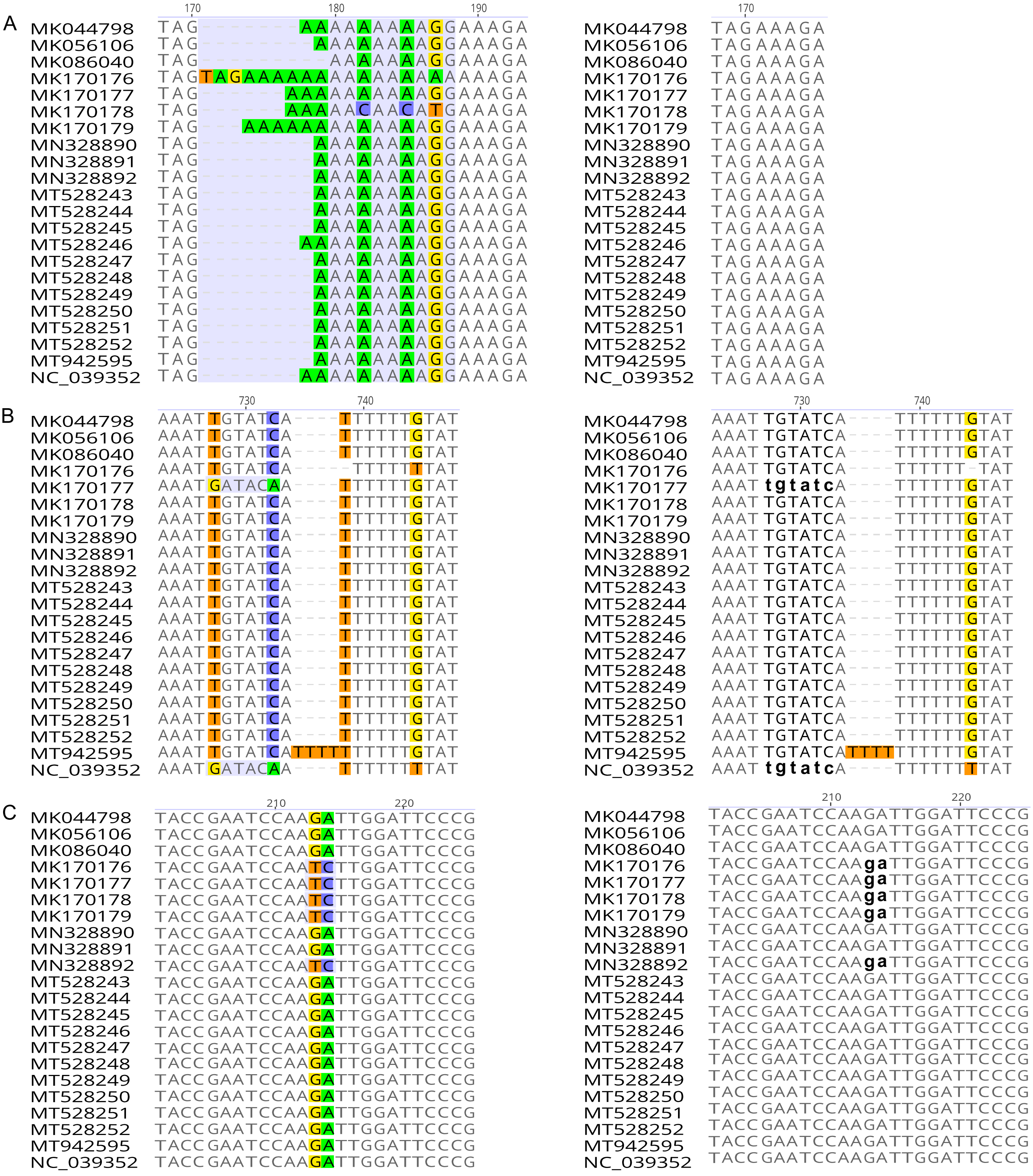

FIGURE 2 Illustration of the manual adjustment of sequence alignments as exemplified by three different intergenic spacers of the plastid genome. For each spacer, the MSA before adjustment is displayed on the left, the MSA after adjustment on the right. Polymorphic nucleotides are highlighted in color. Row (A) illustrates the adjustment of the MSA of spacer *atpI–atpH*, which contains a poly-A region with internal nucleotide polymorphism; this poly-A region (highlighted in blue) was removed due to uncertain positional homology during the adjustment. Row (B) illustrates the adjustment of the MSA of spacer *ndhC–trnV*, which contains shared sequence inversions of a length of six bp (highlighted in blue); these inversions were manually inverted (lower case nucleotides) during the adjustment to avoid incorrect positional homology. Row (C) illustrates the adjustment of the MSA of spacer *psbM–trnD*, which contains shared sequence inversions of a length of two bp (highlighted in blue) as part of a stem-loop structure; these inversions were manually inverted (lower case nucleotides) during the adjustment to avoid incorrect positional homology

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

likelihoods of competing tree topologies based on the concatenation of all plastid partitions using the software CONSEL v.0.20 (Shimodaira and Hasegawa, 2001) and a significance threshold of α = 0.05.

RESULTS

Genome structure and gene content

Genome structure and length, as well as the number of genes per genome, were found to be highly conserved across the plastid genomes of the Gynoxoid group. The genomes exhibit the standard circular and quadripartite structure, comprising one large (LSC) and one small single-copy (SSC) region, separated by two identical IRs, and display minor, if any, length variability. The variability in total sequence length between the largest and the smallest plastid genome in this group was less than 1 kb (except for *Gynoxys tomentosissima*; Appendix S2). All of the genomes of the Gynoxoid group consist of a total of 81 protein-coding regions (seven of which are duplicated in the IRs), 30 transfer RNA (tRNA) genes (seven duplicated in the IRs), and four ribosomal RNA (rRNA) genes (all duplicated in the IRs), resulting in a total of 133 functional coding regions per genome. Also, the same genes contain one or more introns across the genomes: *atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpoC1*, and *rps16* contain one intron each; *clpP* and *ycf3* contain two introns each. The plastid genome of *Gynoxys tomentosissima* is slightly different than the other genomes of the Gynoxoid group: with 155,060 bp, making it the largest sequenced in this study. Compared to the other plastid genomes of the Gynoxoid group, its sequence is approximately 4 kb longer (Appendices S2 and S3) due to the expansion of its IRs into the LSC, with the genes *rpl14*, *rpl16*, *rps3*, *rpl22*, and *rps19* additionally duplicated in the IRs. The GC content of all plastid genomes of the Gynoxoid group was between 37.2% and 37.9% and is, thus, within the typical bandwidth of plastid GC content (Smith, 2009).

Sequence variability across genomes

Sequence variability across the plastid genomes of the Gynoxoid group was low and located almost exclusively in the noncoding regions of the genomes, with coding regions exhibiting only occasional, if any, nucleotide polymorphism (Figure 3). Specifically, the intergenic spacers were the most variable among the three plastid partitions (average $n = 0.006$; Table 2A), followed by the introns ($n = 0.004$) and the coding regions ($n = 0.002$). Among the coding regions, only *ycf1* exhibited a modest number of differences across sequences (i.e., proportion of polymorphic sites of 0.08 and 0.07 before and after the alignment adjustment, respectively; Appendix S4). Among the intergenic spacers, by contrast, several MSAs exhibited a proportion of polymorphic sites above 0.10 both before and after the alignment adjustment (Appendix S5); the MSAs of the intergenic spacers *psbA–trnK-TTT*, *rpl16– rps3*, and *rps18–rpl20*, for example, contained the highest nucleotide diversity ($n = 0.20$). Some sequence variability was also observed in the MSAs of the introns (e.g., the actual noncoding domains of the intron of *trnK-TTT*), but most introns exhibited a modest number of differences across sequences (i.e., proportion of polymorphic sites below 0.10; Appendix S6). Interestingly, the visualization of sequence variability with mVISTA produced somewhat misleading results regarding their phylogenetic utility. The high level of sequence variability indicated for the intergenic spacer *trnT-GGT–psbD* (i.e., position 31,828–33,081 bp) was primarily the consequence of a DNA insertion shared by five sequences and only yields a single variable character for phylogenetic inference upon indel coding. In summary, the plastid genomes of the Gynoxoid group exhibited relatively low but nonetheless divergent levels of sequence variability across the three plastid partitions.

Effect of alignment adjustment on homoplasy indices

The evaluation and, when required, manual adjustment of the MSAs had a considerable effect on homoplasy in these alignments (Figure 4; Appendix S7). Specifically, the MSAs of the intergenic spacers *atpB–rbcL*, *ndhC–trnV-TAC*, *psaA–ycf3*, *trnC-GCA–petN*, and *trnL-TAG–rpl32* exhibited considerably reduced levels of homoplasy upon alignment adjustment, with some spacer regions having their homoplasy index values improve by more than 50%. Improvements were particularly noticeable for values of the RC and the RI among the intergenic spacers. The manual adjustment of the MSAs of the introns and the coding regions had less impact but nonetheless resulted in reduced levels of homoplasy for the alignment of three introns and one coding region. Overall, these changes in homoplasy levels due to alignment adjustment should be seen as conservative estimates of improvement, as the corrected positional homology may additionally benefit phylogenetic inference through, among other factors, a better fit of the employed nucleotide substitution models (e.g., Du et al., 2019).

Phylogenetic reconstructions

A comparison of the phylogenetic tree inferences based on each of the three plastid partitions to the inference conducted on their concatenation indicated that each partition contributed informative characters to the phylogenetic reconstruction, albeit in different proportions (Table 2). After alignment adjustment, the concatenation of all MSAs of the coding regions had a length of 68,076 bp, of which 1030 (1.51%) were polymorphic sites; the concatenation of all MSAs of the introns had a length of 14,184 bp, of which 318 (2.24%) were polymorphic sites; and the concatenation of all MSAs of the intergenic spacers had a length of 37,033 bp

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

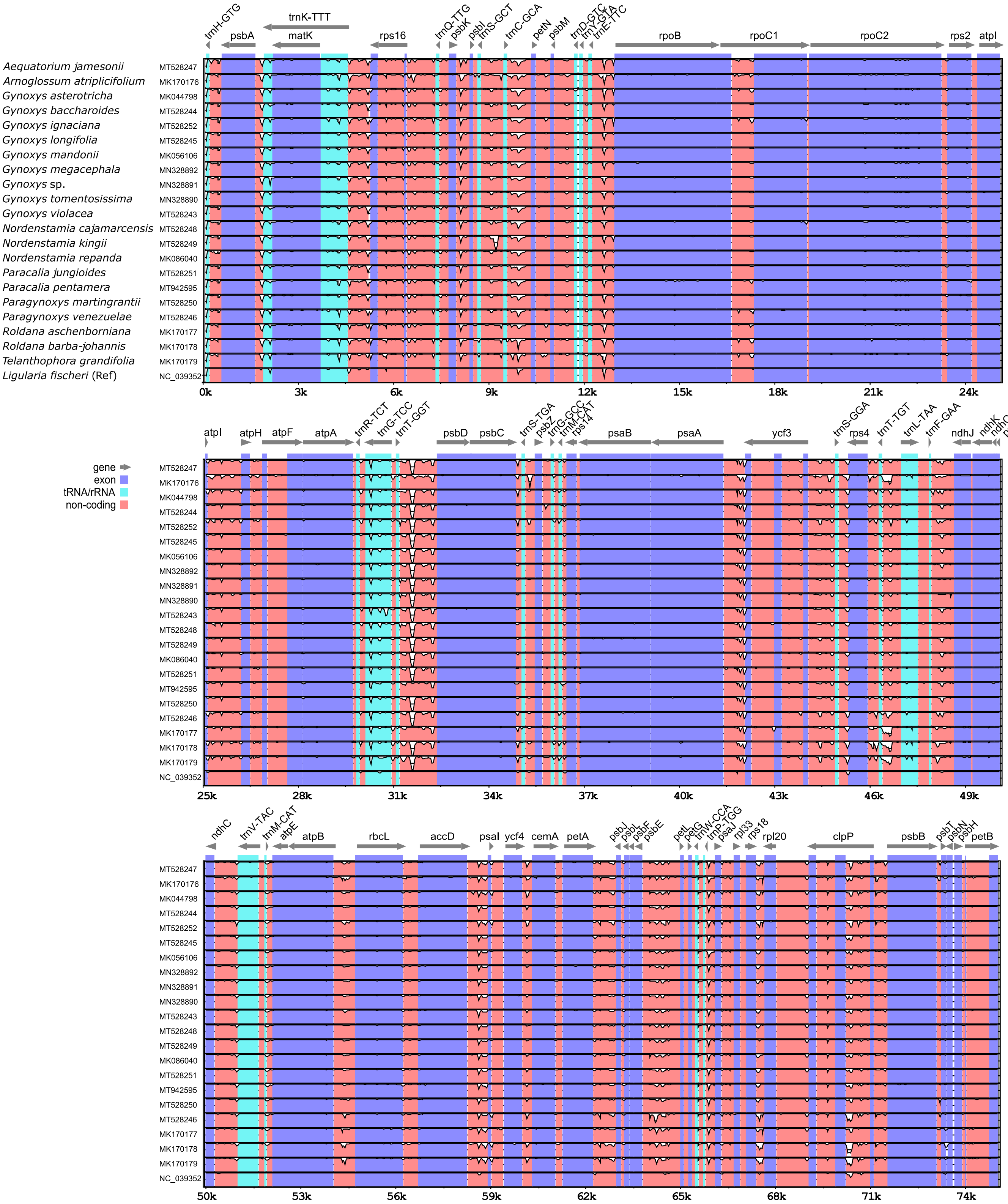

FIGURE 3 Visualization of sequence variability between the plastid genomes under study using mVISTA. The alignment was split into sequence batches of 25 kb length by mVISTA for easier visualization. Each lane represents a genome. In each lane, the proportion of missing similarity is indicated by white color, starting from the top of each lane. Coding regions are represented in blue, transfer and ribosomal RNAs in cyan, and non-coding regions in red. Gray arrows indicate the location and orientation of plastome genes

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

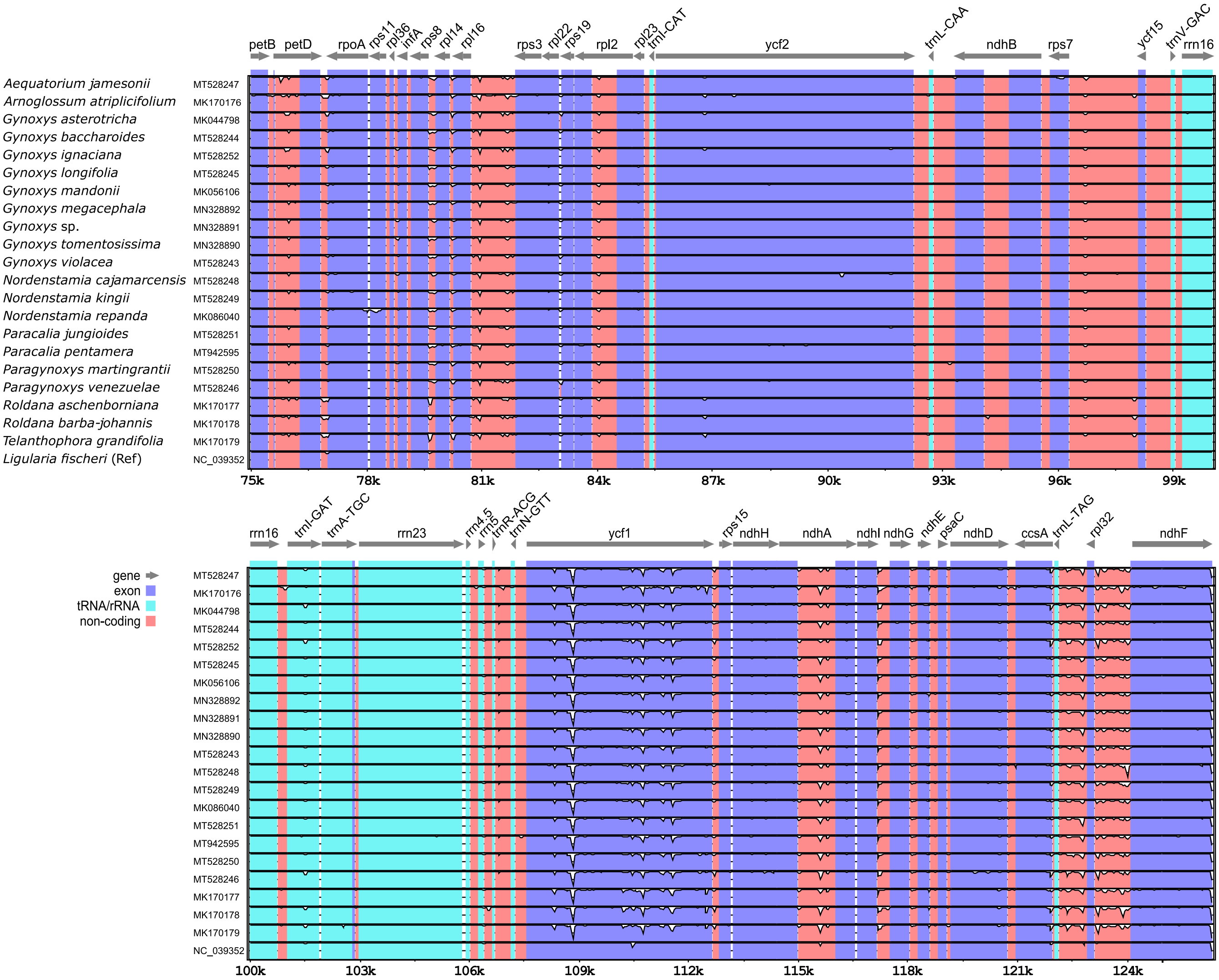

FIGURE 3 Continued

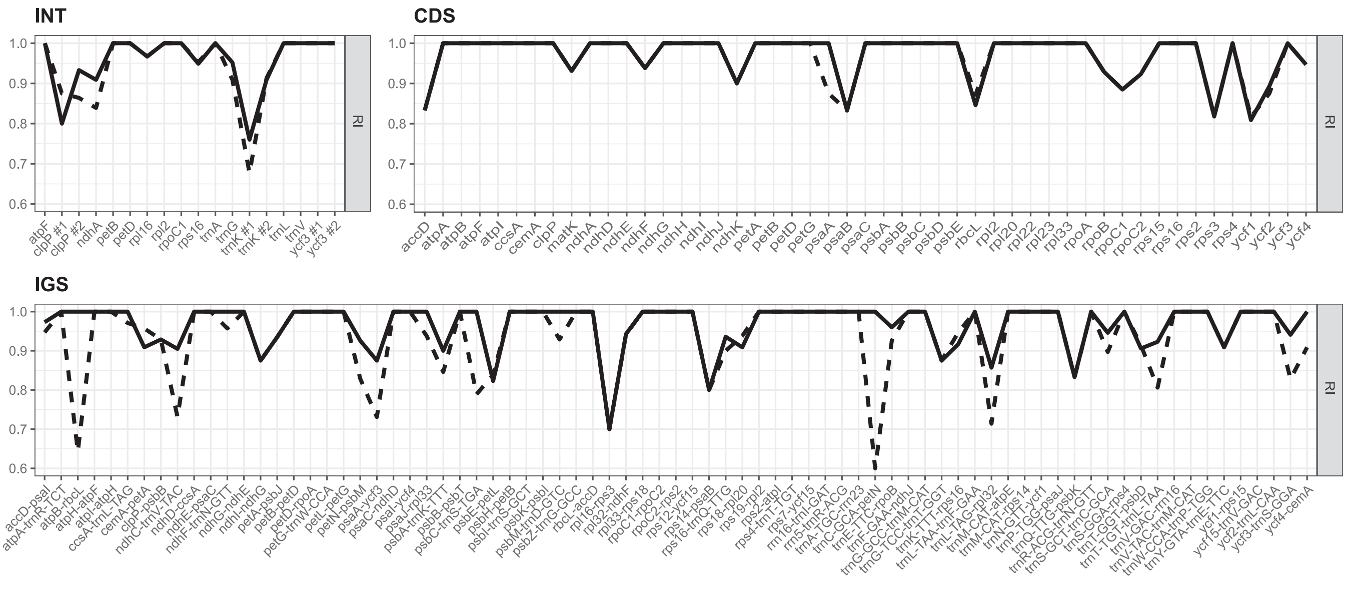

FIGURE 4 Comparison of the RI values across each MSA under study before (dashed lines) and after (solid lines) the alignment adjustment

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Alignment length, sequence variability, and maximum likelihood (ML) tree statistics of the different data sets under study. (A) Alignment length and sequence variability statistics (with percentages shown in parentheses); (B) statistics on ML tree inference before and after alignment adjustment. adj. = adjustment; align. = alignment; BS = bootstrap support; concat. = concatenated; nucl. = nucleotide; PI = parsimony informative

A. Partition	Align. adj.	Align. length (bp)	Gaps/ Missing (%)	Polymorphic sites (%)	PI sites (%)	Nucl. diversity
CDS	Before	68,117	824 (1.21)	1120 (1.64)	270 (0.40)	0.0023
	After	68,076	328 (0.48)	1030 (1.51)	248 (0.36)	0.0022
INT	Before	14,499	470 (3.24)	385 (2.66)	76 (0.52)	0.0037
	After	14,184	428 (3.02)	318 (2.24)	66 (0.47)	0.0032
IGS	Before	38,752	2623 (6.77)	1453 (3.75)	352 (0.91)	0.0056
	After	37,051	2458 (6.63)	1194 (3.22)	275 (0.74)	0.0048
Concat.	Before	121,368	3710 (3.11)	3958 (3.26)	698 (0.58)	0.0035
	After	119,302	3421 (2.82)	2542 (2.13)	589 (0.49)	0.0031

B. Partition	Align. adj	Indel coding	Align. length (bp)	Best ML tree likel. (-lnL)	Best ML tree length (bp)	Avg. BS per node
CDS	Before	no	68,117	103,459.8	0.0193	68
	After	no	68,076	102,518.0	0.0177	66
	Before	yes	68,150	103,736.0	0.0199	68
	After	yes	68,108	102,783.0	0.0182	65
INT	Before	no	14,499	23,560.9	0.0370	45
	After	no	14,184	22,187.8	0.0273	44
	Before	yes	14,658	25,318.6	2.1675	42
	After	yes	14,259	22,760.0	0.0344	43
IGS	Before	no	38,752	66,780.7	0.0580	65
	After	no	37,042	60,729.1	0.0430	63
	Before	yes	39,318	71,559.8	0.0831	62
	After	yes	37,385	63,146.6	0.0556	68
Concat.	Before	no	121,368	194,716.3	0.0337	67
	After	no	119,302	186,078.8	0.0266	73
	Before	yes	122,126	201,615.7	0.0432	68
	After	yes	119,752	189,627.2	0.0315	73

after alignment adjustment, of which 1194 (3.22%) were polymorphic sites. Consequently, the sequence matrix representing the concatenation of the three partitions had a total length of 119,311 bp after alignment adjustment, of which 113,059 (94.8%) were monomorphic sites, 2542 (2.1%) were polymorphic sites, and 3710 (3.1%) gaps or sites of missing data. Among the polymorphic sites, a total of 589 (23.2%) were parsimony informative, of which 248 (42.1%) originated in the coding regions, 275 (46.7%) in the intergenic spacers, and 66 (11.2%) in the introns.

Considerable differences in the results of our phylogenetic reconstructions were identified based on the different plastid partitions as well as the adjustment of alignments. First, the phylogenetic reconstructions before and after the adjustment of alignments generally resulted in trees with different topologies and branch support. For example, the best ML tree inferred under the concatenation of all three plastid partitions was significantly different before and after alignment adjustment (Figure 5A; Table 3). Similarly, the best ML trees based on the concatenation of all coding regions produced different topologies before and after alignment adjustment, while BS support was >50% for all but one node (Appendix S8). In the tree inferred before alignment adjustment, *Gynoxys longifolia* was sister to *Nordenstamia kingii* (BS 59%), and a clade comprising *Paragynoxys* and *Aequatorium* was sister to a clade of *G. baccharoides* and *G. violacea* (BS 52%). In the tree inferred after alignment adjustment, fewer nodes with BS support >50% were retrieved, and none of the previously stated relationships were supported. The best ML trees of the concatenation of all introns also exhibited different topologies before and after alignment adjustment, with BS support primarily >50% (Appendix S9).

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

By contrast, the best ML trees of the concatenation of all intergenic spacers inferred before and after alignment adjustment were highly similar in topology and generally recovered BS support >50% (Appendix S10). Second, the phylogenetic reconstructions based on different plastid partitions resulted in trees with different topologies and branch support. For example, the best ML tree inferred under the concatenation of the coding regions was significantly different from the best ML tree inferred under the concatenation of all three plastid partitions (Figure 5B; Table 3). Similarly, the tree topologies inferred under the concatenation of all introns were notably different from the topologies inferred under the concatenation of all coding regions and intergenic spacers, and also exhibited lower branch support (Appendices S8–S10). The average BS support per node in the best ML tree inferred under the concatenation of all introns was considerably lower than the average BS support per node in the trees inferred on the concatenation of all coding regions and all intergenic spacers. Only the sister relationship between the Gynoxoid group and the North and Central American members of the Tussilagininae was consistently retrieved across the plastid partitions. In summary, the reconstruction of the phylogenetic relationships of the Gynoxoid group was strongly

TABLE 3 Statistical comparison of competing phylogenetic trees as displayed in Figure 5 using the AU test on the concatenation of all three plastid partitions upon the coding of indels and after alignment adjustment. Significant *P* values are indicated with an asterisk

Constraint	*P*
Figure 5, left tree (positive control)	0.545
Figure 5A, right tree	0.036*
Appendix S11A, right tree	0.037*
Appendix S11A, right tree	0.545
Appendix S11B, right tree	0.461

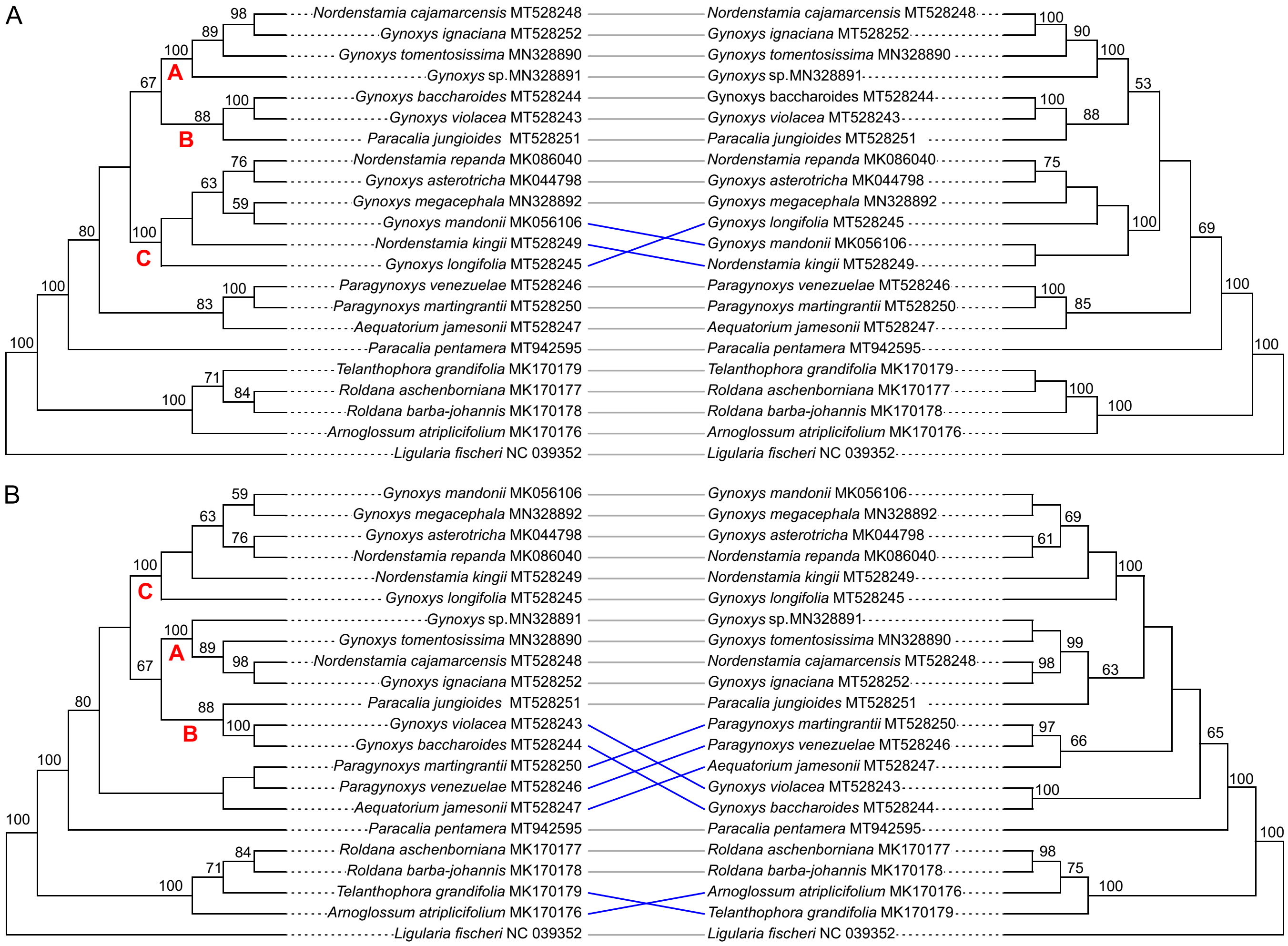
A
98
89
100
A
67
88
B
76
63
59
80
100
C
100
83
100
100
71
84
100
Nordenstamia cajamarcensis MT528248
Gynoxys ignaciana MT528252
Gynoxys tomentosissima MN328890
Gynoxys sp.MN328891
Gynoxys baccharoides MT528244
Gynoxys violacea MT528243
Paracalia jungioides MT528251
Nordenstamia repanda MK086040
Gynoxys asterotricha MK044798
Gynoxys megacephala MN328892
Gynoxys mandonii MK056106
Nordenstamia kingii MT528249
Gynoxys longifolia MT528245
Paragynoxys venezuelae MT528246
Paragynoxys martingrantii MT528250
Aequatorium jamesonii MT528247
Paracalia pentamera MT942595
Telanthophora grandifolia MK170179
Roldana aschenborniana MK170177
Roldana barba-johannis MK170178
Arnoglossum atriplicifolium MK170176
Ligularia fischeri NC 039352
Nordenstamia cajamarcensis MT528248
Gynoxys ignaciana MT528252
Gynoxys tomentosissima MN328890
Gynoxys sp.MN328891
Gynoxys baccharoides MT528244
Gynoxys violacea MT528243
Paracalia jungioides MT528251
Nordenstamia repanda MK086040
Gynoxys asterotricha MK044798
Gynoxys megacephala MN328892
Gynoxys longifolia MT528245
Gynoxys mandonii MK056106
Nordenstamia kingii MT528249
Paragynoxys venezuelae MT528246
Paragynoxys martingrantii MT528250
Aequatorium jamesonii MT528247
Paracalia pentamera MT942595
Telanthophora grandifolia MK170179
Roldana aschenborniana MK170177
Roldana barba-johannis MK170178
Arnoglossum atriplicifolium MK170176
Ligularia fischeri NC 039352
100
90
100
53
100
88
75
100
69
100
85
100
100
100
100
B
59
63
76
100
C
100
A
89
98
67
88
B
100
80
100
100
84
71
100
Gynoxys mandonii MK056106
Gynoxys megacephala MN328892
Gynoxys asterotricha MK044798
Nordenstamia repanda MK086040
Nordenstamia kingii MT528249
Gynoxys longifolia MT528245
Gynoxys sp.MN328891
Gynoxys tomentosissima MN328890
Nordenstamia cajamarcensis MT528248
Gynoxys ignaciana MT528252
Paracalia jungioides MT528251
Gynoxys violacea MT528243
Gynoxys baccharoides MT528244
Paragynoxys martingrantii MT528250
Paragynoxys venezuelae MT528246
Aequatorium jamesonii MT528247
Paracalia pentamera MT942595
Roldana aschenborniana MK170177
Roldana barba-johannis MK170178
Telanthophora grandifolia MK170179
Arnoglossum atriplicifolium MK170176
Ligularia fischeri NC 039352
Gynoxys mandonii MK056106
Gynoxys megacephala MN328892
Gynoxys asterotricha MK044798
Nordenstamia repanda MK086040
Nordenstamia kingii MT528249
Gynoxys longifolia MT528245
Gynoxys sp.MN328891
Gynoxys tomentosissima MN328890
Nordenstamia cajamarcensis MT528248
Gynoxys ignaciana MT528252
Paracalia jungioides MT528251
Paragynoxys martingrantii MT528250
Paragynoxys venezuelae MT528246
Aequatorium jamesonii MT528247
Gynoxys violacea MT528243
Gynoxys baccharoides MT528244
Paracalia pentamera MT942595
Roldana aschenborniana MK170177
Roldana barba-johannis MK170178
Arnoglossum atriplicifolium MK170176
Telanthophora grandifolia MK170179
Ligularia fischeri NC 039352
69
61
100
99
98
63
97
66
65
100
100
100
98
75
100

FIGURE 5 Comparison of phylogenetic trees of the Gynoxoid group inferred before and after alignment adjustment and under different datasets. The tree displayed on the left is held constant across both comparisons and constitutes the tree with the highest likelihood score inferred on the concatenated MSAs of all three plastid genome partitions upon the coding of indels and after alignment adjustment. Three clades (i.e., "A", "B", and "C") are highlighted on this tree by red letters located next to the most recent common ancestor of each clade. (A) Comparison of the best tree against its equivalent inferred before alignment adjustment; and (B) comparison of the best tree against its equivalent inferred on the concatenated MSAs of all coding regions only

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

dependent on the adjustment of the alignments and the exact plastid partitions employed.

By comparison, small, if any, differences in the results of our phylogenetic reconstructions were identified in relation to the coding of indels and between the two tree inference methods employed. First, the coding of indels had only a small impact on tree inference: several trees exhibited minor topological changes upon inclusion of an indel coding matrix in the phylogenetic reconstruction, but the overall relationships, as well as the branch support, exhibited few, if any, differences. For the concatenation of all plastid partitions upon alignment adjustment, the best ML tree before the coding of indels was topologically identical to the best ML tree after indels coding (Table 3; Appendix S11). Similarly, the phylogenetic trees reconstructed under the concatenation of all coding regions experienced no topological changes upon the coding of indels, whereas trees reconstructed under the concatenation of all introns and all intergenic spacers exhibited differences in one or occasionally multiple nodes (Appendices S8–S10). Second, the phylogenetic reconstructions conducted under different inference methods generated trees that were primarily different with respect to branch support. Branch support for reconstructions under BI was often higher than branch support for reconstructions under ML (Appendices S12 and S13). For the concatenation of all plastid partitions upon alignment adjustment and the coding of indels, the best ML tree was topologically identical to the best BI tree inferred (Table 3; Appendix S11).

Phylogenetic reconstruction on the concatenation of all plastid partitions resulted in trees that were consistent in topology and similar in branch support across different tree inference methods and the coding of indels after the adjustment of alignments (Appendix S11), but inconsistent in topology and branch support before these adjustments (Figure 5A; Appendices S12 and S13). Except for the most recent common ancestor of the two clades containing *Gynoxys* and the node of a clade comprising *Gynoxys asterotricha* and *G. megacephala*, all nodes of the best ML tree based on the concatenation of all plastid partitions exhibited BS support >50% upon alignment adjustment (Appendix S12). In the 50% majority-rule consensus tree of the posterior tree distribution of the BI, most nodes had PP values ≥ 0.9 (Appendix S13). By contrast, the best ML tree inferred before the adjustment of alignments exhibited lower BS support as well as topological incongruence regarding the positions of *Gynoxys mandonii*, *G. longifolia*, *G. megacephala*, and *Nordenstamia kingii*, and the paraphyly of *Roldana* with respect to *Telantophora* (Appendix S12). These differences in tree topology before and after the alignment adjustments were even more pronounced under BI and may be discordant given the high branch support received under this optimality criterion (Appendix S13).

Inference of relationships

The phylogenetic relationships recovered through our tree inferences on the concatenation of all plastid partitions resulted in high branch support for most clades (Appendices S12 and S13). The Gynoxoid group was recovered as a clade with maximum BS and PP support, comprising the genera *Aequatorium*, *Gynoxys*, *Nordenstamia*, *Paracalia*, and *Paragynoxys*. The non-Gynoxoid taxa of the Tussilagininae (i.e., *Roldana*, *Telanthophora*, and *Arnoglossum*) were recovered as sister to this Gynoxoid group in each reconstruction and with maximum support. *Paracalia pentamera*, which constitutes the type species in the genus, was recovered as the earliest-diverging lineage of the Gynoxoid group in each inference, again with high branch support. Except for *Paragynoxys*, all other genera of the Gynoxoid group that are represented by two or more species were found to be non-monophyletic. Specifically, *Paragynoxys martingrantii* and *P. venezuelae* were recovered as sister species with maximum branch support and identified to be sister to the specimen of *Aequatorium* included in this study (medium BS support, maximum PP support). *Gynoxys* and *Nordenstamia* were recovered in two separate subclades and found to be either para- or polyphyletic, depending on the reconstruction observed. One subclade comprised four species of *Gynoxys* and two species of *Nordenstamia* and exhibited maximum branch support under both ML and BI. Specifically, it comprised *Gynoxys asterotricha*, *G. longifolia*, *G. mandonii*, and *G. megacephala*, as well as *Nordenstamia kingii* and *N. repanda*, the last of which is the type for the genus. However, the most recent common ancestor of *Nordenstamia kingii* and *Gynoxys megacephala* was poorly supported in this subclade and part of a polytomy under BI upon coding indels. The other subclade comprised five species of *Gynoxys*, one species of *Nordenstamia*, and one species of *Paracalia* and had maximum PP but only medium BS support (BS > 67). This clade comprised *G. baccharoides*, which constitutes the type species for *Gynoxys*, *G. ignaciana*, *G. tomentosissima*, *G. violacea*, and an unidentified species of *Gynoxys* (i.e., *Gynoxys* sp. SEN301). Moreover, the clade comprised *Paracalia jungioides* and *Nordenstamia cajamarcensis*. The species *Paracalia jungioides*, *Gynoxys baccharoides*, and *G. violacea* formed a clade under both ML and BI, which exhibited maximum PP and high BS support (BS > 81). The other species of this subclade were also recovered as a monophyletic group with maximum branch support.

DISCUSSION

Plastid genomes of Senecioneae

The present investigation is the first to compare plastid genomes from different genera of the Tussilagininae and to employ their sequences in a phylogenetic analysis. As of December 2020, 29 complete plastid genomes of the Senecioneae have been sequenced and are available through the NCBI Nucleotide database (https://www.ncbi.nlm.nih.gov/nucleotide/), representing five different genera and two of the four subtribes. Specifically, plastid genomes have been

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

sequenced for *Dendrosenecio* (Hauman ex Hedberg) B. Nord., *Jacobaea* Mill., *Pericallis* D. Don, and *Senecio* L. of the subtribe Senecioninae, and *Ligularia* Cass. of the subtribe Tussilagininae. Gichira et al. (2019), sequenced and compared 11 plastid genomes of *Dendrosenecio* (Hauman ex Hedberg) B. Nord. and *Senecio* to identify variable regions for phylogenetic analysis. Similarly, Chen et al. (2018) generated plastid genomes of six species of *Ligularia* to examine the utility of these genomes as barcodes for identifying individual species. The present investigation expands the list of sequenced plastid genomes of the Tussilagininae by eight genera. All plastid genomes sequenced in this study display a highly conserved genome structure and exhibit the two large inversions that separate the Asteraceae from other flowering plant families (Kim et al., 2005). The size range of these newly sequenced plastid genomes and their gene order and content is highly similar to other Senecioneae (Gichira et al., 2019; Chen et al., 2018).

Phylogenetic position of the Gynoxoid clade

The results of our phylogenetic reconstructions corroborate the previously supported relationships of the Tussilaginineae as inferred by Pelser et al. (2007). Specifically, our results support the phylogenetic relationships of the *Aequatorium–Arnoglossum* clade that were reported by Pelser et al. (2007) and confirm the Gynoxoid group as being monophyletic with high statistical support. While previous studies had already reported the Gynoxoid group as a clade (e.g., Pelser et al., 2010; Quedensley et al., 2018, both primarily based on ITS DNA sequences), their sampling of taxa and genomic regions was generally insufficient to infer the monophyly of the Gynoxoid group. Moreover, we recovered the Gynoxoid clade as sister to a clade of North and Central America taxa (i.e., *Arnoglossum*, *Telantophora*, and *Roldana*) that are distributed from the United States (Quedensley et al., 2018) to Panama (Funston, 2009; Clark and Pruski, 2015) and possibly Colombia (Calvo, 2016), and this sister relationship was previously illustrated (Pelser et al., 2010). Quedensley et al. (2018) reported that several genera in this sister clade to the Gynoxoid group were highly polyphyletic, and our results support this assessment, as we found *Roldana* non-monophyletic in several reconstructions (e.g., Appendix S12). Future studies on the phylogenetic relationships of the Tussilaginineae should further increase the taxon sampling.

Phylogenetic relationships in the Gynoxoid clade

Our phylogenetic reconstructions recovered several relationships within the Gynoxoid clade with high clade support. For example, our results suggest that genus *Paracalia* is not monophyletic and that its type species (i.e., *Paracalia pentamera*) is sister to the rest of the Gynoxiod group, whereas *Paragynoxys* is monophyletic with full support. The monophyly of *Paragynoxys* is also supported by morphological characters such as discoid capitula and deep-lobed white corollas (Cuatrecasas, 1955). The reconstructions based on the full plastid genome revealed that *Aequatorium* and *Paragynoxys* were sister genera (1.0 PP and 80% ML-BS), whereas reconstructions based on individual plastid regions could not resolve this relationship (e.g., Pelser et al., 2010). Moreover, these complete plastome reconstructions identified three highly supported subclades within the Gynoxoid clade irrespective of the tree inference method applied: subclade A, which comprises *G. ignaciana*, *G. tomentosissima*, *Gynoxys* sp., and *Nordenstamia cajamarcensis* and is also supported by morphological characters such as opposite leaves and granular hairs along the veins on the upper face of the leaves; subclade B, which comprises *G. baccharioides*, *G. violacea*, and *Paracalia jungioides*; and subclade C, which comprises *G. megacephala*, *G. mandonii*, *Nordenstamia repanda*, *G. asterotricha*, *Nordemstamia kingii*, and *G. longifolia*. All internal nodes in these subclades were found to be well supported, except for the position of *N. kingii* in subclade C (weakly supported in the ML and BI trees when ignoring indel coding and unsupported upon inclusion of the indel coding matrix). The type species of *Gynoxys* and *Nordenstamia* (i.e., *Gynoxys baccharoides* and *Nordenstamia repanda*, respectively) were recovered in a clade in which the species of both genera did not segregate, indicating the non-monophyly of both genera.

Importance of manual alignment adjustment

The present investigation highlights the importance of evaluating and, where necessary, adjusting software-generated MSAs before phylogenomic analysis. While different alignment algorithms have been implemented in the various software tools available for automatic DNA sequence alignment (e.g., Needleman-Wunsch algorithm in CLUSTAL W; Thompson et al., 1994), their mechanistic processes often depart from the actual biological processes that shape the molecular evolution of DNA sequences. For example, biological processes often comprise the instantaneous insertion, deletion, inversion, or translocation of multiple nucleotides, yet many alignment algorithms cannot replicate these mechanisms (Graham et al., 2000; Borsch and Quandt, 2009; Ochoterena, 2008). Molecular phylogenetic studies, thus, often adjust their DNA sequence alignments manually using motif-based approaches. Such approaches were conceptualized as motif alignments that follow defined rules (e.g., Kelchner, 2000; Loehne and Borsch, 2005; Morrison, 2006, 2015) and aim to improve the alignment of microstructural mutations while assessing positional homology (de Pinna, 1991). Numerous investigations have demonstrated the impact that the selection of alignment method can have on the reconstruction of phylogenetic trees (e.g., Morrison and Ellis, 1997; Simmons

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

et al., 2010a, b; Wong et al., 2008). Moreover, several simulation studies have shown that alignment accuracy is highly dependent on the frequency of genomic insertions and deletions (e.g., Pervez et al., 2014). Hence, the manual adjustment of DNA sequence alignments following a motif-based approach has become a common practice among molecular phylogenetic studies, particularly those based on single genetic markers. Phylogenomic studies, by contrast, often dismiss the practice of manual alignment adjustment, as the amount of sequence data under analysis is considered to outweigh the ability of any researcher to correct software-induced alignment errors for all but the smallest data sets (Wu et al., 2012) or to lead to a lack of repeatability (Edwards et al., 2016). Instead, most phylogenomic studies often exclude those regions of a MSA that are deemed unreliable via positional filtering processes (e.g., Ali et al., 2019). While such a practice may eliminate some of the erroneous statements of positional homology among aligned nucleotides, they often fail to recognize inversions (Figure 2) or eliminate positions with sequence gaps although the respective indel motifs allow clear homologous positions. It is, therefore, not surprising that methods for automated alignment filtering, in which a threshold of shared gaps is employed to determine the removal of nucleotides, may actually counteract accurate tree inference (e.g., Tan et al., 2015). The inadvertent analysis of DNA matrices that harbor sequence inversions is particularly problematic, as these inversions can be highly homoplastic (Kelchner and Wendel, 1996) and often lead to spurious phylogenetic results (Joly et al., 2010). Unsurprisingly, the manual examination of the plastid phylogenomic data set of this investigation for alignment errors has led to the identification of numerous sequence inversions that were not recognized at the stage of the software-driven sequence alignment (Figure 2; Appendix S1).

Process of manual alignment adjustment

To alleviate the problem of missing positional homology upon automatic sequence alignment, we conducted visual examinations and manual corrections of the software-derived MSAs using a motif approach. Specifically, we removed the nucleotide positions from the final sequence matrix for which positional homology could not be established. For example, we removed or truncated length-variable poly-A/T microsatellites that originated through repeated and independent insertions of single or multiple nucleotides and did not form recognizable sequence motifs. In addition to improving the positional homology for individual MSAs, this strategy allowed us to (1) identify and re-invert naturally occurring sequence inversions that cannot be automatically aligned and would introduce erroneous nucleotide polymorphisms if left unedited (Chen et al., 2016), and (2) mask and, thus, exclude those nucleotide positions for which positional homology could not be reasonably identified. Time-efficient work was facilitated through automatically partitioning of annotated genomic regions into individual data sets, which could then be edited manually in PhyDE without affecting the overall alignment. Upon alignment adjustment, individual MSAs were then concatenated automatically using the pipeline of Gruenstaeudl et al. (2018). The homoplasy indices improved in value upon alignment adjustment (Appendix S7), and this change was also seen in the inferred tree topologies when comparing tree inference before and after alignment adjustments (Table 2B; Figure 5A). Adjusting software-derived alignments may have the effect of avoiding erroneous phylogenetic signal from both substitutions and coded microstructural mutations (i.e., insertions, deletions, and inversions) within the alignable sections of DNA sequences and, additionally, of avoiding spurious signal from the unalignable sections. Moreover, the evaluation and adjustment of sequence alignments may also assist in identifying cases of natural gene rearrangements among plastid genomes, which would lead to MSAs comprising non-homologous sequence elements in noncoding sequence partitions, as such rearrangements are typically not identified through software-derived sequence alignment (Fonseca and Lohmann, 2017). Plastid phylogenomic reconstructions should, thus, consider the evaluation and, where necessary, correction of automatic alignment results.

Mosaic-like evolution of plastid genomes

The large majority of angiosperm plastid genomes display a highly conserved structure, uniparental inheritance, and a general absence of recombination between chromosomes (Marechal and Brisson, 2010, but see Ruhlman et al., 2017; Li et al., 2020). Hence, many researchers assume that the plastid genome evolves as a single linkage unit (Bock, 2007), with different genomic regions sharing the same evolutionary history (e.g., Lu et al., 2018). However, recent studies have reported widespread phylogenetic incongruence between different regions of the plastid genome (e.g., Goncalves et al., 2019; Gruenstaeudl, 2019; Walker et al., 2019), which indicates that the plastid genome may not represent a homogeneous genetic locus. This phylogenetic incongruence may partially be the result of the different mutation rates and selective constraints across the plastid genome, which is illustrated by the co-existence of the relatively slowly evolving gene *rbcL*, the more rapidly evolving gene *matK* with nearly equal site rates in all three codon positions, and the even faster evolving noncoding markers *trnL* intron, *trnT–L* intergenic spacer, and *trnL–F* intergenic spacer in the same genome (Müller et al., 2006). Examples for selective constraints are directed selection on certain nucleotide positions and compensatory base changes (Kelchner, 2002; Borsch et al., 2003), which can cause patterns of homoplasy and lead to a spurious phylogenetic signal, especially when the number of variable nucleotides is limited. This observed phylogenetic incongruence may also be the result of the different structural constraints in the

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

plastid genome, which comprises a succession of conserved and variable elements that can display secondary DNA structure. For example, the highly variable and AT-rich stem loops of noncoding plastid DNA often exhibit accelerated, lineage-specific nucleotide substitution rates and a high frequency of microstructural mutations (Korotkova et al., 2014). The presence of phylogenetic incongruence due to systematic error, such as the selection of incorrect nucleotide substitution models or incorrect homology statements during MSA (e.g., Zhang et al., 2020), represents another possible explanation, and its extent is under active investigation (Goncalves et al., 2020). In summary, the plastid genome seems to exhibit a mosaic-like pattern of molecular evolution, and if that pattern is not adequately modeled during phylogenetic reconstruction, the resulting inferences may be highly supported yet spurious (see discussion by Walker et al., 2019; Thode et al., 2020; Zhang et al., 2020). Plastid phylogenomics may, thus, be facing a new trend in which researchers aim to better account for the differential phylogenetic signal of particular genome regions (Goncalves et al., 2020).

Phylogenetic utility of different plastome regions

The different regions of the plastid genome exhibit different molecular constraints and, thus, a different utility for phylogenetic reconstructions. The coding regions of the plastid genome have traditionally been used for the reconstruction of deep-level phylogenetic relationships (Graham and Olmstead, 2000; Xi et al., 2012; Walker et al., 2019), but these may be biased when genes with incongruent phylogenetic signals coexist in the genome (e.g., Gruenstaeudl, 2019; Goncalves et al., 2019; Walker et al., 2019). Accordingly, Goncalves et al. (2020) suggested to infer phylogenetic relationships on individual genes and to compare the resulting topologies to the reconstructions based on the concatenation of all coding regions. In this regard, it is important to note that gene tree incongruence may have different causes: incongruent signal may be (1) truly phylogenetic in origin (i.e., genes reflect different evolutionary histories), or (2) tree-like and originate from homoplastic nucleotide substitutions, which are often associated with highly unequal substitution rates (Morton and Clegg, 1995). Incongruent, yet tree-like signal may additionally be compounded by inadequate model fit, poor sample size, and other systemic errors during phylogenetic tree inference (Kelchner and Thomas, 2006). Given that the plastid genome as a whole shares a common evolutionary history within our study group (as it does in most plastid phylogenomic investigations on land plant lineages), it is likely that an incongruent, yet tree-like signal is also present in the data sets analyzed here, as evidenced by the incongruent phylogenetic trees inferred from different plastid partitions. Another potential explanation for gene tree incongruence rests with the different quantities of phylogenetic information in genes, introns, and intergenic spacers: the noncoding regions of the plastid genome experience only limited selective pressures and are known to evolve at faster rates than the coding regions (Clegg et al., 1994; Kim et al., 1999), leading to a higher frequency of potentially informative sites in these partitions. On the other hand, different regions of the plastid genome can also differ in the quality of the phylogenetic signal they contain (Barniske et al., 2012), underscoring the different molecular evolutionary patterns acting within them.

Most molecular phylogenetic studies preferentially employ noncoding genome regions to reconstruct the relationships of taxa that exhibit shallow levels of sequence divergence (Kelchner, 2002; Shaw et al., 2007; Androsiuk et al., 2020). Such levels are often found among recently radiated plant lineages, and the noncoding portion of the plastid genome is particularly useful in acquiring phylogenetic resolution in such lineages, as demonstrated here for the Gynoxoid group. However, noncoding regions also exhibit a higher frequency of microstructural mutations and require greater attention when generating MSAs. In this study, the intron partition contributed fewer potentially informative characters than the other partitions, which corresponds with the lower average variability of introns compared to intergenic spacers. This lower variability is, however, often compounded with very specific mutational dynamics associated with secondary DNA structures and the alternation of highly conserved and highly variable sequence elements (Kelchner, 2002). Nonetheless, many phylogenetic studies have demonstrated that introns can contain high levels of hierarchical phylogenetic signal at various taxonomic levels (Creer, 2007). The intron of *rpl16*, for instance, is one of the fastest-evolving introns in the plastid genome of land plants and has been used extensively for reconstructing phylogenetic relationships at the species level (Kelchner, 2002) and often constitutes the plastome partition with the most useful phylogenetic signal (Korotkova et al., 2011). However, with the extremely low genetic distances present in the Gynoxoid clade, the potentially higher signal quality in introns seems to have been outcompeted by the more numerous variable sites among the intergenic spacers.

Harnessing all regions of the plastid genome

The results of this investigation suggest that the analysis of all regions of the plastid genome can be highly beneficial for acquiring a well-supported phylogenetic reconstruction of a recently diverged plant lineage. This conclusion can likely be generalized to many plastid phylogenomic studies at the species level, as the phylogenetic signal of any plastid partition individually (i.e., genes, introns, and intergenic spacers alone) is likely insufficient to retrieve fully resolved and well-supported trees. Several studies have demonstrated that phylogenetic tree inference is affected by the selection of plastid genomic regions employed (e.g., Lu et al., 2018;

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Wikström et al., 2020). Thode et al. (2020), for example, recovered congruent phylogenetic trees from coding and noncoding plastid regions of neotropical lianas, but the nodes exhibited stronger support values in reconstructions using noncoding sequences. Similarly, Koehler et al. (2020) discovered that only a subset of the plastid genome regions was particularly useful in the phylogenetic reconstruction of the subfamily Opuntioideae (Cactaceae). Furthermore, Zhang et al. (2020) encountered considerable phylogenetic incongruence in a phylogenomic analysis of 36 tribes of the Fabaceae, including incongruence among strongly supported nodes and in relation to coding versus noncoding sections of the plastid genome. Despite these findings, many plastid phylogenomic investigations ignore the noncoding sections of the plastid genome during phylogenetic tree inference (e.g., Ma et al., 2014; Ross et al., 2015), often due to challenges in the alignment of these regions (e.g., Zhang et al., 2017). This preferential selection of coding over noncoding regions in plastid phylogenomic studies may have led to reduced power in many reconstructions, and few, if any, biological reasons exist to exclude such genome regions from phylogenetic analysis. Our study contributes to the ongoing discussion of how the phylogenetic signal of the complete plastid genome can be partitioned and employed more effectively for phylogenetic inference (Koehler et al., 2020; Thode et al., 2020). More research is needed to address this question and should include different lineages of land plants and different levels of genetic distance among taxa and sequences.

CONCLUSIONS

In this plastid phylogenomic investigation, we analyzed the phylogenetic relationships of the Gynoxoid group, an Andean lineage of the Asteraceae with low genetic distances between taxa. Our results indicated that at least two, and possibly three, of the five genera are polyphyletic. Moreover, our results demonstrated that the inclusion of all plastid genome partitions was needed to infer well-supported phylogenetic trees of the Gynoxoid group and that manual correction of sequence alignments had a considerable effect on tree inference. Furthermore, our results indicated that the adjustment of software-derived DNA sequence alignments may constitute an important step toward improved phylogenetic analyses in plastid phylogenomic studies. Specifically, the same standards of DNA sequence alignment and matrix construction that have been applied in studies of individual genomic regions should also be applied to plastid phylogenomic data sets. The impact of incorrect positional homology in a sequence matrix may be particularly severe among plastid data sets with low genetic distances, as the misaligned regions may contain a high proportion of the potentially informative sites. Consequently, species-level investigations that require the analysis of complete plastid genomes to resolve phylogenetic relationships should apply the utmost rigor in the motif-based alignment of nucleotide sequences and consider excluding areas of uncertain homology from their alignments.

ACKNOWLEDGMENTS

The authors thank Norbert Kilian, Tilo Henning, Daniel Montesinos, Stephan Beck, Carla Maldonado, and Huber Villca for assistance during sample collection, Sabine Scheel for assistance with DNA isolations, and Robert K. Jansen (University of Texas at Austin) for guidance in the sequencing and assembly of several of the plastid genomes. The authors also thank two anonymous reviewers and Associate Editor Mark P. Simmons for critically reading the manuscript and providing helpful suggestions for improvement. Funding for fieldwork was provided to B.E. by the Julia Krieg Forschungsfonds of the BGBM in the context of the collaboration with the Herbario Nacional de Bolivia and by the association Friends of the Botanical Garden of the BGBM. Funding for fieldwork was provided to T.S.Q. by the University of Texas at Austin Plant Biology Graduate Program. The authors acknowledge the high-performance computing service of the ZEDAT of the Freie Universität Berlin for providing allocations of computing time. Open Access funding enabled and organized by Projekt DEAL.

AUTHOR CONTRIBUTIONS

B.E. and T.S.Q. conducted the fieldwork, generated herbarium vouchers, and extracted DNA. B.E. and M.G. conducted the lab work and assembled and annotated the plastid genomes. B.E. and T.B. conducted the visual inspection and adjustment of sequence alignments. B.E. and M.G. conducted the phylogenetic analyses and generated all figures and tables. B.E. and M.G. led the writing of the manuscript, with additional contributions by T.B. and T.S.Q. All authors approved the final version of the manuscript.

DATA AVAILABILITY STATEMENT

All data sets analyzed during the present investigation are available from Zenodo at https://zenodo.org/record/4428211#.YYlLmVNMFKM.

REFERENCES

Ali, R., M. Bogusz, and S. Whelan. 2019. Identifying clusters of high confidence homologies in multiple sequence alignments. *Molecular Biology and Evolution* 36: 2340–2351.

Androsiuk, P., J. Jastrzebski, L. Paukszto, K. Makowczenko, A. Okorski, A. Pszczolkowska, K. Chwedorzewska, et al. 2020. Evolutionary dynamics of the chloroplast genome sequences of six *Colobanthus* species. *Scientific Reports* 10: 11522.

Bakker, F., D. Lei, J. Yu, S. Mohammadin, Z. Wei, S. van de Kerke, B. Gravendeel, et al. 2015. Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an iterative organelle genome assembly pipeline. *Biological Journal of the Linnean Society* 117: 33–43.

Barniske, A., T. Borsch, K. Müller, M. Krug, A. Worberg, C. Neinhuis, and D. Quandt. 2012. Phylogenetics of early branching eudicots: comparing phylogenetic signal across plastid introns, spacers, and genes. *Journal of Systematics and Evolution* 50: 85–108.

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Barriel, V. 1994. Molecular phylogenies and nucleotide insertion–deletion. *Comptes rendus de l'Academie des sciences, III, Sciences de la vie* 317: 693–701.

Beck, S., and D. Ibáñez. 2014. Asteraceae. *In* P. M. Jorgensen, M. H. Nee, and S. G. Beck [eds.], Catálogo de las plantas vasculares de Bolivia. *Monographs in Systematic Botany from the Missouri Botanical Garden* 127: 290–382.

Bellot, S., T. Mitchell, and H. Schaefer. 2020. Phylogenetic informativeness analyses to clarify past diversification processes in Cucurbitaceae. *Scientific Reports* 10: 1–13.

Beltran, H., and J. Campos de la Cruz. 2009. *Nordenstamia magnifolia* (Asteraceae: Senecioneae), una especie nueva del norte de Peru. *Arnolda* 16: 37–40.

Beltran, H., A. Granda, B. León, A. Sagástegui, I. Sanchez, and M. Zapata. 2006. Asteraceas endémicas de Perú. *Revista Peruana Botanica* 13: 64–164.

Bock, R. 2007. Structure, function, and inheritance of plastid genomes. *In* R. Bock [ed.], Cell and molecular biology of plastids, 29–63. Springer Verlag, Heidelberg, Germany.

Borsch, T., K. Hilu, D. Quandt, V. Wilde, C. Neinhuis, and W. Barthlott. 2003. Noncoding plastid *trnT-trnF* sequences reveal a well resolved phylogeny of basal angiosperms. *Journal of Evolutionary Biology* 16: 558–576.

Borsch, T., and D. Quandt. 2009. Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. *Plant Systematics and Evolution* 282: 169–199.

Brudno, M., C. Do, G. Cooper, M. Kim, E. Davydov, NISC Comparative Sequencing Program, E. Green, et al. 2003. LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA. *Genome Research* 13: 721–731.

Calvo, J. 2016. A new combination in *Roldana* (Compositae, Senecioneae). *Phytotaxa* 272: 225–227.

Chen, R., Y. Lau, Y. Zhang, and W. Yang. 2016. SRinversion: a tool for detecting short inversions by splitting and re-aligning poorly mapped and unmapped sequencing reads. *Bioinformatics* 32: btw516.

Chen, X., J. Zhou, Y. Cui, Y. Wang, B. Duan, and H. Yao. 2018. Identification of *Ligularia* herbs using the complete chloroplast genome as a super-barcode. *Frontiers in Pharmacology* 9: 695–706.

Clark, B., and J. Pruski. 2015. Telanthophora H. Rob. et Brettell. *In* J. Pruski and H. Robinson [eds.], Flora mesoamericana, 427–469. Missouri Botanical Garden Press, St. Louis, MO, USA.

Clegg, M., B. Gaut, G. Learn, and B. Morton. 1994. Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences, USA* 91: 6795–6801.

Creer, S. 2007. Choosing and using introns in molecular phylogenetics. *Evolutionary Bioinformatics Online* 3: 99–108.

Cuatrecasas, J. 1951. Contributions to the flora of South America: studies on Andean Compositae, II. Studies in South American plants, III. *Fieldiana* 27: 1–74.

Cuatrecasas, J. 1955. A new genus and other novelties in Compositae. *Brittonia* 8: 151–163.

de Pinna, M. 1991. Concepts and tests of homology in the cladistic paradigm. *Cladistics* 7: 367–394.

Dierckxsens, N., P. Mardulyn, and G. Smits. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* 45: e18.

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

Du, Y., S. Wu, S. Edwards, and L. Liu. 2019. The effect of alignment uncertainty, substitution models and priors in building and dating the mammal tree of life. *BMC Evolutionary Biology* 191: 203.

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

Farris, J. 1989. The retention index and the rescaled consistency index. *Cladistics* 5: 417–419.

Fonseca, L., and L. Lohmann. 2017. Plastome rearrangements in the *Adenocalymma-Neojobertia* clade (Bignonieae, Bignoniaceae) and its phylogenetic implications. *Frontiers in Plant Science* 8: 1875.

Frazer, K., L. Pachter, A. Poliakov, E. Rubin, and I. Dubchak. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Research* 32: W273–W279.

Funston, A. 2009. Taxonomic revision of *Roldana* (Asteraceae: Senecioneae), a genus of the southwestern U.S.A., Mexico, and Central America. *Annals of the Missouri Botanical Garden* 95: 282–337.

Gernandt, D., X. Aguirre-Dugua, A. Vazquez-Lobo, A. Willyard, A. Moreno-Letelier, J. Perez de la Rosa, D. Pinero, and A. Liston. 2018. Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection *Australes*. *American Journal of Botany* 105: 711–725.

Gichira, A., S. Avoga, and Z. Li. 2019. Comparative genomics of 11 complete chloroplast genomes of Senecioneae (Asteraceae) species: DNA barcodes and phylogenetics. *Botanical Studies* 60: 1–17.

Givnish, T., A. Zuluaga, D. Spalink, M. Soto, V. Lam, J. Saarela, C. Sass, et al. 2018. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. *American Journal of Botany* 105: 1888–1910.

Glanz, S., and U. Kueck. 2009. Trans-splicing of organelle introns—a detour to continuous RNAs. *Bioessays* 31: 921–934.

Goncalves, D., R. Jansen, T. Ruhlman, and J. Mandel. 2020. Under the rug: abandoning persistent misconceptions that obfuscate organelle evolution. *Molecular Phylogenetics and Evolution* 151: 106903.

Goncalves, D., B. Simpson, E. Ortiz, G. Shimizu, and R. Jansen. 2019. Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. *Molecular Phylogenetics and Evolution* 138: 219–232.

Graham, S., and R. Olmstead. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* 87: 1712–1730.

Graham, S., P. Reeves, A. Burns, and R. Olmstead. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Science* 161: 83–96.

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 Research* 47: 59–64.

Greiner, S., Sobanski, J., and Bock, R. 2015. Why are most organelle genomes transmitted maternally? *BioEssays* 37: 80–94.

Gruenstaeudl, M. 2019. Why the monophyly of Nymphaeaceae currently remains indeterminate: an assessment based on gene-wise plastid phylogenomics. *Plant Systematics and Evolution* 305: 827–836.

Gruenstaeudl, M., N. Gerschler, and T. Borsch. 2018. Bioinformatic workflows for generating complete plastid genome sequences: an example from *Cabomba* (Cabombaceae) in the context of the phylogenomic analysis of the water-lily clade. *Life* 8: 1–17.

Hind, N. 2007. An annotated preliminary checklist of the Compositae of Bolivia, version 2. Website: https://www.kew.org/sites/default/files/2019-01/Bolivian%20compositae%20checklist.pdf [accessed 05 November 2021].

Joly, S., B. Pfeil, B. Oxelman, T. Mclenachan, and P. Lockhart. 2010. Erratum: A statistical approach for distinguishing hybridization and incomplete lineage sorting. *American Naturalist* 174: 621–622.

Kadereit, J., and C. Jeffrey. 1996. A preliminary analysis of cpDNA variation in the tribe Senecioneae (Compositae). *In* D. Hind and H. Beentje [eds.], Compositae: Systematics. Proceedings of the International Compositae Conference, vol. 1, 349–360. Royal Botanic Gardens, Kew, UK.

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

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

Kelchner, S. 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* 87: 482–498.

Kelchner, S. 2002. Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. *American Journal of Botany* 89: 1651–1669.

Kelchner, S., and M. Thomas. 2006. Model use in phylogenetics: nine key questions. *Trends in Ecology and Evolution* 22: 87–94.

Kelchner, S., and J. Wendel. 1996. Hairpins create minute inversions in non-coding regions of chloroplast DNA. *Current Genetics* 30: 259–262.

Kim, K., K. Choi, and R. Jansen. 2005. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Molecular Biology and Evolution* 22: 1783–1792.

Kim, S., D. Crawford, R. Jansen, and A. Santos-Guerra. 1999. The use of a non-coding region of chloroplast DNA in phylogenetic studies of the subtribe Sonchinae (Asteraceae: Lactuceae). *Plant Systematics and Evolution* 215: 85–99.

Kluge, A., and J. Farris. 1969. Quantitative phyletics and the evolution of anurans. *Systematic Zoology* 18: 1–32.

Knope, M., M. Bellinger, E. Datlof, T. Gallaher, and M. Johnson. 2020. Insights into the evolutionary history of the Hawaiian *Bidens* (Asteraceae) adaptive radiation revealed through phylogenomics. *Journal of Heredity* 111: 1–19.

Koehler, M., M. Reginato, T. Chies, and L. Majure. 2020. Insights into chloroplast genome evolution across Opuntioideae (Cactaceae) reveals robust yet sometimes conflicting phylogenetic topologies. *Frontiers in Plant Science* 11: 1–20.

Korotkova, N., T. Borsch, D. Quandt, N. Taylor, K. Mueller, and W. Barthlott. 2011. What does it take to resolve relationships and to identify species with molecular markers? An example from the epiphytic Rhipsalideae (Cactaceae). *American Journal of Botany* 98: 1549–1572.

Korotkova, N., L. Nauheimer, H. Ter-Voskanyan, M. Allgaier, and T. Borsch. 2014. Variability among the most rapidly evolving plastid genomic regions is lineage-specific: implications of pairwise genome comparisons in *Pyrus* (Rosaceae) and other angiosperms for marker choice. *PLoS One* 9: 1–16.

Langmead, B., and S. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357.

Leebens-Mack, J., L. Raubeson, L. Cui, J. Kuehl, M. Fourcade, T. Chumley, J. Boore, et al. 2005. Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. *Molecular Biology and Evolution* 22: 1948–1963.

Lewis, P. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50: 913–925.

Li, C., R. Wang, and D. Li. 2020. Comparative analysis of plastid genomes within the Campanulaceae and phylogenetic implications. *PLoS One* 15: 23.

Loehne, C., and T. Borsch. 2005. Molecular evolution and phylogenetic utility of the *petD* group II intron: A case study in basal angiosperms. *Molecular Biology and Evolution* 22: 317–332.

Lu, L., C. Cox, S. Mathews, W. Wang, J. Wen, and Z. Chen. 2018. Optimal data partitioning, multispecies coalescent and Bayesian concordance analyses resolve early divergences of the grape family. *Cladistics* 34: 57–77.

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

Lundin, R. 2006. *Nordenstamia* Lundin (Compositae-Senecioneae), a new genus from the Andes of South America. *Compositae Newsletter* 44: 14–23.

Ma, P., Y. Zhang, C. Zeng, Z. Guo, and D. Li. 2014. Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae). *Systematic Biology* 63: 933–950.

Marechal, A., and N. Brisson. 2010. Recombination and the maintenance of plant organelle genome stability. *New Phytologist* 186: 299–317.

Moore, M., P. Soltis, C. Bell, J. Burleigh, and D. Soltis. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of Eudicots. *Proceedings of the National Academy of Sciences, USA* 107: 4623–4628.

Morillo, G., and B. Briceno. 2000. Distribución de las Asteraceas de los páramos venezolanos. *Acta Botánica Venezuélica* 23: 47–67.

Morrison, D. 2006. Multiple sequence alignment for phylogenetic purposes. *Australian Systematic Botany* 19: 479–539.

Morrison, D. 2015. Is sequence alignment an art or a science? *Systematic Botany* 40: 14–26.

Morrison, D., and J. Ellis. 1997. Effects of nucleotide sequence alignment on phylogeny estimation: a case study of 18 S rDNAs of *Apicomplexa*. *Molecular Biology and Evolution* 14: 428–441.

Morton, B., and M. Clegg. 1995. Neighboring base composition is strongly correlated with base substitution bias in a region of the chloroplast genome. *Journal of Molecular Evolution* 41: 597–603.

Mower, J., and T. Vickrey. 2018. Structural diversity among plastid genomes of land plants. *Advances in Botanical Research* 85: 263–292.

Müller, J., K. Müller, C. Neinhuis, and D. Quandt. 2010. PhyDE: Phylogenetic Data Editor. Website: http://www.phyde.de/ [accessed 19 August 2020].

Müller, K. 2005. SeqState: primer design and sequence statistics for phylogenetic DNA datasets. *Applied Bioinformatics* 4: 65–69.

Müller, K., T. Borsch, and K. Hilu. 2006. Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: contrasting *matK*, *trnT-F*, and *rbcL* in basal angiosperms. *Molecular Phylogenetics and Evolution* 41: 99–117.

Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, NY, NY, USA.

Nei, M., and W. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, USA* 76: 5269–5273.

Nordenstam, B. 2007. XII. Tribe Senecioneae. *In* J. Kadereit and C. Jeffrey [eds.], Flowering plants, eudicots: Asterales, 208–241. Springer, NY, NY, USA.

Nordenstam, B., P. Pelser, J. Kadereit, and L. Watson. 2009. Senecioneae. *In* V. Funk, A. Susanna, T. Stuessy, and R. Bayer [eds.], Systematics, evolution, and biogeography of Compositae, 503–525. International Association for Plant Taxonomy, Vienna, Austria.

Ochoterena, H. 2008. Homology in coding and non-coding DNA sequences: a parsimony perspective. *Plant Systematic and Evolution* 282: 151–168.

Paradis, E., and K. Schliep. 2018. Ape 5.0: AN environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526–528.

Pelser, P., A. Kennedy, E. Tepe, J. Shidler, B. Nordenstam, J. Kadereit, and L. Watson. 2010. Patterns and causes of incongruence between plastid and nuclear Senecioneae (Asteraceae) phylogenies. *American Journal of Botany* 97: 856–873.

Pelser, P., B. Nordenstam, J. Kadereit, and L. Watson. 2007. An ITS phylogeny of tribe Senecioneae (Asteraceae) and a new delimitation of *Senecio* L. *Taxon* 56: 1077–1104.

Pervez, M., M. Babar, A. Nadeem, M. Aslam, A. Awan, N. Aslam, T. Hussain, et al. 2014. Evaluating the accuracy and efficiency of multiple sequence alignment methods. *Evolutionary Bioinformatics* 10: 205–217.

Pouchon, C., A. Fernandez, J. Nassar, F. Boyer, S. Aubert, S. Lavergne, and J. Mavarez. 2018. Phylogenomic analysis of the explosive adaptive radiation of the *Espeletia* complex (Asteraceae) in the tropical Andes. *Systematic Biology* 67: 1041–1060.

Quedensley, T., M. Gruenstaeudl, and R. Jansen. 2018. Phylogenetic relationships of the Mexican tussilaginoid genera (Asteraceae: Senecioneae). *Journal of the Botanical Research Institute of Texas* 12: 481–498.

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Rambaut, A., A. Drummond, D. Xie, G. Baele, and M. Suchard. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 67: 901–904.

Ronquist, F., and J. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

Ross, T., C. Barrett, M. Soto Gomez, V. Lam, C. Henriquez, D. Les, J. Davis, et al. 2015. Plastid phylogenomics and molecular evolution of Alismatales. *Cladistics* 32: 160–178.

Rozas, J., A. Ferrer-Mata, J. Sánchez-DelBarrio, S. Guirao-Rico, P. Librado, S. Ramos-Onsins, and A. Sánchez-Gracia. 2017. DnaSP 6: DNA sequence polymorphism analysis of large datasets. *Molecular Biology and Evolution* 34: 3299–3302.

Ruhlman, T., J. Zhang, J. Blazier, J. Sabir, and R. Jansen. 2017. Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. *American Journal of Botany* 104: 559–572.

Schliep, K. 2011. Phangorn: phylogenetic analysis in R. *Bioinformatics* 27: 592–593.

Shaw, J., E. Lickey, E. Schilling, and R. Small. 2007. Comparison of whole chloroplast genome sequence to choose non-coding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* 94: 275–288.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* 51: 492–508.

Shimodaira, H., and M. Hasegawa. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 517: 1246–1247.

Simmons, M., K. Mueller, and A. Norton. 2010a. Alignment of, and phylogenetic inference from, random sequences: the susceptibility of alternative alignment methods to creating artifactual resolution and support. *Molecular Phylogenetics and Evolution* 57: 1004–1016.

Simmons, M., K. Mueller, and C. Webb. 2010b. The deterministic effects of alignment bias in phylogenetic inference. *Cladistics* 27: 402–416.

Simmons, M., and H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.

Smith, D. 2009. Unparalleled GC content in the plastid DNA of *Selaginella*. *Plant Molecular Biology* 71: 627–639.

Smith, D., and P. Keeling. 2015. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proceedings of the National Academy of Sciences, USA* 112: 10177–10184.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Stamatakis, A., J. Hoover, and J. Rougemint. 2008. A rapid boot-strap algorithm for the RAxML web servers. *Systematic Biology* 57: 758–771.

Tan, G., M. Muffato, C. Ledergerber, J. Herrero, N. Goldman, M. Gil, and C. Dessimoz. 2015. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Systematic Biology* 64: 778–791.

Team, R. D. C. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website: http://www.r-project.org/

Tesfaye, K., T. Borsch, K. Govers, and E. Bekele. 2007. Characterization of *Coffea* chloroplast microsatellites and evidence for the recent divergence of *C. arabica* and *C. eugenioides* chloroplast genomes. *Genome* 50: 1112–1129.

Thode, V., L. Lohmann, and I. Sanmartin. 2020. Evaluating character partitioning and molecular models in plastid phylogenomics at low taxonomic levels: a case study using *Amphilophium* (Bignonieae, Bignoniaceae). *Journal of Systematics and Evolution* 58: 1071–1089.

Thompson, J., D. Higgins, and T. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673–4680.

Tinoco, B., P. Astudillo, S. Latta, D. Strubbe, and C. Graham. 2013. Influence of patch factors and connectivity on the avifauna of fragmented *Polylepis* forest in the Ecuadorian Andes. *Biotropica* 45: 602–611.

Vargas, O., and S. Madrinan. 2012. Preliminary phylogeny of *Diplostephium* (Asteraceae): speciation rate and character evolution. *Lundellia* 15: 1–15.

Vargas, O., E. Ortiz, and B. Simpson. 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). *New Phytologist* 214: 1736–1750.

Vision, T., and M. Dillon. 1996. Sinopsis de *Senecio* (Senecioneae, Asteraceae) para el Peru. *Arnolda* 4: 23–46.

Walker, J., N. Walker-Hale, O. Vargas, D. Larson, and G. Stull. 2019. Characterizing gene tree conflict in plastome-inferred phylogenies. *PeerJ* 7: 1–31.

Wikström, N., B. Bremer, and C. Rydin. 2020. Conflicting phylogenetic signals in genomic data of the coffee family (Rubiaceae). *Journal of Systematics and Evolution* 58: 440–460.

Wong, K., M. Suchard, and J. Huelsenbeck. 2008. Alignment uncertainty and genomic analysis. *Science* 319: 473–476.

Wu, M., S. Chatterji, and J. Eisen. 2012. Accounting for alignment uncertainty in phylogenomics. *PLoS One* 7: 1–10.

Wyman, S., R. Jansen, and J. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255.

Xi, Z., B. Ruhfel, H. Schaefer, A. Amorim, S. Manickam, K. Wurdack, P. Endress, et al. 2012. Phylogenomics and a posteriori data portioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proceedings of the National Academy of Sciences, USA* 109: 17519–17524.

Yao, G., J. Jin, H. Li, J. Yang, V. Mandala, M. Croley, R. Mostow, et al. 2019. Plastid phylogenomic insights into the evolution of Caryophyllales. *Molecular Phylogenetics and Evolution* 134: 74–86.

Zhang, R., Y. Wang, J. Jin, G. Stull, A. Bruneau, D. Cardoso, L. Queiroz, et al. 2020. Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. *Systematic Biology* 69: 613–622.

Zhang, S., J. Jin, S. Chen, M. Chase, D. Soltis, H. Li, J. Yang, et al. 2017. Diversification of Rosaceae since the late cretaceous based on plastid phylogenomics. *New Phytologist* 214: 1355–1367.

Zhang, X., T. Deng, M. Moore, Y. Ji, N. Lin, Z. Huajie, A. Meng, et al. 2019. Plastome phylogenomics of *Saussurea* (Asteraceae: Cardueae). *BMC Plant Biology* 19: 1–10.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Appendix S1. Summary of the position and length of the microsatellites and small sequence inversions in the multiple sequence alignments (MSAs) that were masked during alignment adjustment.

Appendix S2. Overview of the total and regional lengths of the plastid genomes of the Gynoxoid group.

Appendix S3. Plastome map of *Gynoxys tomentosissima*, the longest of the sequenced plastid genomes of the Gynoxoid group.

Appendix S4. Alignment metrics and homoplasy indices of the MSAs of the coding regions before and after alignment adjustment.

Appendix S5. Alignment metrics and homoplasy indices of the MSAs of the intergenic spacers before and after alignment adjustment.

Appendix S6. Alignment metrics and homoplasy indices of the MSAs of the introns before and after alignment adjustment.

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Appendix S7. Consistency index (CI) and rescaled consistency index (RC) values across each MSA under study before and after the alignment adjustment.

Appendix S8. Results of phylogenetic tree inference via ML on the concatenated MSAs of all coding regions before and after alignment adjustment as well as with and without the coding of indels.

Appendix S9. Results of phylogenetic tree inference via ML on the concatenated MSAs of all intergenic spacers before and after alignment adjustment as well as with and without the coding of indels.

Appendix S10. Results of phylogenetic tree inference via ML on the concatenated MSAs of all introns before and after alignment adjustment as well as with and without the coding of indels.

Appendix S11. Comparison of phylogenetic trees of the Gynoxoid group inferred before and after the coding of indels and under different tree inference methods.

Appendix S12. Results of phylogenetic tree inference via ML on the concatenated MSAs of all three plastid genome partitions before and after alignment adjustment as well as with and without the coding of indels.

Appendix S13. Results of phylogenetic tree inference via BI on the concatenated MSAs of all three plastid genome partitions before and after alignment adjustment as well as with and without the coding of indels.

How to cite this article: Escobari, B., T. Borsch, T. S. Quedensley, and M. Gruenstaeudl. 2021. Plastid phylogenomics of the Gynoxoid group (Senecioneae, Asteraceae) highlights the importance of motif-based sequence alignment amid low genetic distances. *American Journal of Botany* 108(11): 2235–2256. https://doi.org/10.1002/ajb2.1775

15372197, 2021, 11, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1775 by Texas Christian University Tcu, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License