



OPEN

Comparative analyses of *Mikania* (Asteraceae: Eupatorieae) plastomes and impact of data partitioning and inference methods on phylogenetic relationships

Verônica A. Thode¹, Caetano T. Oliveira², Benoît Loeuille³, Carolina M. Siniscalchi⁴✉ & José R. Pirani⁵

We assembled new plastomes of 19 species of *Mikania* and of *Ageratina fastigiata*, *Litothamnus nitidus*, and *Stevia collina*, all belonging to tribe Eupatorieae (Asteraceae). We analyzed the structure and content of the assembled plastomes and used the newly generated sequences to infer phylogenetic relationships and study the effects of different data partitions and inference methods on the topologies. Most phylogenetic studies with plastomes ignore that processes like recombination and biparental inheritance can occur in this organelle, using the whole genome as a single locus. Our study sought to compare this approach with multispecies coalescent methods that assume that different parts of the genome evolve at different rates. We found that the overall gene content, structure, and orientation are very conserved in all plastomes of the studied species. As observed in other Asteraceae, the 22 plastomes assembled here contain two nested inversions in the LSC region. The plastomes show similar length and the same gene content. The two most variable regions within *Mikania* are *rpl32-ndhF* and *rpl16-rps3*, while the three genes with the highest percentage of variable sites are *ycf1*, *rpoA*, and *psbT*. We generated six phylogenetic trees using concatenated maximum likelihood and multispecies coalescent methods and three data partitions: coding and non-coding sequences and both combined. All trees strongly support that the sampled *Mikania* species form a monophyletic group, which is further subdivided into three clades. The internal relationships within each clade are sensitive to the data partitioning and inference methods employed. The trees resulting from concatenated analysis are more similar among each other than to the correspondent tree generated with the same data partition but a different method. The multispecies coalescent analysis indicate a high level of incongruence between species and gene trees. The lack of resolution and congruence among trees can be explained by the sparse sampling (~ 0.45% of the currently accepted species) and by the low number of informative characters present in the sequences. Our study sheds light into the impact of data partitioning and methods over phylogenetic resolution and brings relevant information for the study of *Mikania* diversity and evolution, as well as for the Asteraceae family as a whole.

Chloroplasts are organelles linked to photosynthesis, which have many essential functions in plants, such as carbon fixation and biosynthesis of starch, fatty acids, amino acids, and pigments^{1,2}. The chloroplast genome (plastome) in angiosperms usually has a circular shape, with 120 to 180 kb in size, divided in four main regions:

¹Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Avenida Bento Gonçalves, Porto Alegre, Rio Grande do Sul 91501-970, Brazil. ²Sítio Roberto Burle Marx, Instituto do Patrimônio Histórico e Artístico Nacional, Estrada Roberto Burle Marx, 2019, Barra de Guaratiba, Rio de Janeiro, Rio de Janeiro 23020-240, Brazil. ³Departamento de Botânica, Centro de Biociências, Universidade Federal de Pernambuco, Avenida Professor Moraes Rego, 1235, Recife, Pernambuco 50670-901, Brazil. ⁴Department of Biological Sciences, Mississippi State University, 295 Lee Blvd, Mississippi State, Mississippi, MS 39762, USA. ⁵Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, Rua do Matão, Tv. 14, São Paulo, São Paulo 05508-090, Brazil. ✉email: carol.siniscalchi@gmail.com

two Inverted Repeat (IR) regions, one Large Single Copy (LSC), and one Small Single Copy (SSC) region³. Plastome gene composition and order are generally conserved among land plants^{4,5}, but recent studies have documented that variation at many levels can occur^{6,7}. It has been widely accepted that plastomes are uniparentally inherited and do not present recombination, with the whole genome frequently being interpreted as a single locus in phylogenetic analysis, implying in all genes evolving concertedly⁸. However, in the last decade, evidence has accumulated that this organelle can be biparentally inherited, copies with different sequences can occur⁹ and that different portions of the genome can evolve at different paces¹⁰. Recent studies recommend analyzing plastome genes individually in phylogenetic inferences, through methods like the multispecies coalescent, which also accounts for possible incongruence between gene trees and species trees^{10,11}.

Mikania Willd. is the most diverse genus within the tribe Eupatorieae and the largest genus of climbing plants in Asteraceae, with around 450 species^{12,13}. It has a pantropical distribution, mainly neotropical, with most of the diversity concentrated in South America^{12,13}. The large number of species makes carrying out taxonomic revisions and molecular studies for the genus difficult¹⁴. Nevertheless, *Mikania* is easily morphologically recognized by its four-flowered heads surrounded by four involucre bracts; its circumscription has been indisputable since its description in 1742¹⁵. A reevaluation of the current infrageneric classification of the genus is needed¹⁶, but the lack of broadly sampled phylogenies prevents the elaboration of a classification based on evolutionary relationships and monophyletic groups¹⁵. Several regional taxonomic studies of *Mikania* (e.g.^{17–19}) are available, but large-scale taxonomic treatments are still a challenge. Some taxa, such as *Mikania glomerata* Spreng. and *M. laevigata* Sch.Bip. ex Baker, have known pharmacological uses, especially in the treatment of respiratory diseases²⁰. Consequently, the genus is well represented in phytochemical studies, which have been conducted with approximately 12% of all *Mikania* species²⁰. The main chemical compounds linked to pharmacological activities, which are found in different parts of *Mikania* plants, are coumarins and derivatives, sesquiterpenes, sesquiterpenes lactones, diterpenes, phytosterols/terpenoids, and flavonoids²⁰. Some species are considered invasive, including some widespread weeds, such as *Mikania micrantha* Kunth¹⁶.

The first attempt to investigate phylogenetic relationships within *Mikania* with molecular data was based on AFLP markers, the intronic region of the plastid gene *rps16*, and ribosomal ITS and ETS, but included only representatives of generic sections proposed for Brazilian species of *Mikania*²¹. More molecular studies are needed to evaluate infrageneric limits within this genus, as well as to explore its morphological and chemical evolution, biogeographic history, and diversification. The only genomic resources reported in the literature for the genus are the complete plastome (NC031833.1²²) and the chromosome-scale genome²³ of *Mikania micrantha*, an invasive plant well known for causing significant damage to natural ecosystems and crops in several parts of the world. For the tribe Eupatorieae as a whole, which has ~ 180 genera and 2200 species²⁴, only three other plastomes were published to date, besides *M. micrantha*: i.e., *Ageratina adenophora* (Spreng.) R.M.King & H.Rob. (NC015621²⁵), *Praxelis clematidea* R.M.King & H.Rob. (NC023833²⁶), and *Ageratum conyzoides* L. (MK905238²⁷).

The family Asteraceae is one of the most species-rich families of flowering plants, including an impressive morphological and ecological diversity²⁸. While the backbone relationships within the family have recently become clearer, internal relationships at tribal and generic levels still need a lot of attention^{28,29}. Although many Asteraceae plastomes were recently published, most studies characterize the plastome of a single taxon and focus in comparative genomic analyses at higher taxonomic levels (e.g.^{22,25–27,30,31}). These studies are important to improve the understanding of plastome variation in the family as a whole^{27,31} and to provide more information on their phylogenetic utility. Yet, larger infrageneric samplings are essential to explore evolution and phylogenetic relationships in Asteraceae at lower taxonomic levels (e.g.^{32,33}).

In this study, we sequenced new complete plastomes of 19 species of *Mikania* representative of the morphological diversity of the genus. The sampled taxa are: *M. additicia* B.L.Rob., *M. brevifauca* W.C.Holmes & McDaniel, *M. burchellii* Baker, *M. decora* Poepp., *M. decumbens* Malme, *M. fasciculata* C.T.Oliveira & Pirani, *M. glomerata* Spreng., *M. haenkeana* DC., *M. lehmanii* Hieron., *M. neurocaula* DC., *M. oblongifolia* DC., *M. obtusata* DC., *M. parvifolia* Baker, *M. purpurascens* (Baker) R.M.King & H.Rob., *M. salviifolia* Gardner, *M. smaragdina* Dusén ex Malme, *M. sylvatica* Klatt., *M. ternata* (Vell.) B.L.Rob., and *M. triangularis* Baker (Table 1). The plastomes of three species from other genera of Eupatorieae, namely *Ageratina fastigiata* (Kunth) R.M.King & H.Rob., *Lithothamnus nitidus* (DC.) W.C.Holmes, and *Stevia collina* Gardner, were also sequenced and used as outgroups. The previously published plastome of *Mikania micrantha* (NC031833.1²²) was included in the analyses as well. This study aims to characterize and compare the plastomes within *Mikania* and among closely related genera within tribe Eupatorieae to improve our understanding about the evolution of this genome and investigate different methods of phylogenetic reconstruction with this dataset. More specifically, we: (i) sequenced, assembled, and characterized the overall plastome structure; (ii) performed comparative genomic analyses within *Mikania*, and among *Mikania* and other Eupatorieae genera; (iii) identified putative repeated regions; and (iv) investigated phylogenetic relationships using both concatenation and multispecies coalescent methods with different data partitions.

Results

Plastome assembly and characterization. Approximately 1–2 GB of data and 10,488,036–20,546,020 paired-end raw reads for each plastome were obtained. The 19 *Mikania* plastomes range in length from 151,773 (*M. glomerata*) to 152,229 bp (*M. salviifolia*) (Table 1, Supplementary Fig. S1). All assembled plastomes show the general structure found in most angiosperms, divided in four main regions, which in *Mikania* consists of a LSC (83,527–83,878 bp), a SSC (18,163–18,347 bp), and a pair of IR (24,877–25,014 bp) regions (Table 1, Supplementary Fig. S1). As observed in other Asteraceae taxa, the 22 plastomes assembled here contain two inversions in the LSC region³⁴: a large inversion (22,211–22,423 bp, *M. salviifolia* and *Ageratina fastigiata*, respectively) including 16 genes from *trnS^{GCU}-trnC^{GCA}* to *trnG^{UCC}-trnI^{GGU}* and a small inversion (3223–3304 bp, *Lithothamnus nitidus* and *A. fastigiata*, respectively) nested within the former, which includes six genes located between

Species	Voucher	GenBank (accession)	Plastome length (bp)	LSC length (bp)	IR length (bp)	SSC length (bp)	Inv1	Inv2
<i>Mikania additicia</i> B.L.Rob.	Oliveira 822 (SPF)	MT793849	151,983	83,827	24,927	18,302	22,358	3250
<i>Mikania brevifauca</i> W.C.Holmes & McDaniel	Oliveira 903 (SPF)	MT793850	152,161	83,868	24,980	18,333	22,343	3252
<i>Mikania burchellii</i> Baker	Oliveira 706 (SPF)	MT793851	151,829	83,767	24,877	18,308	22,362	3273
<i>Mikania decora</i> Poepp. & Endl.	Oliveira 986 (SPF)	MT793834	151,914	83,659	24,964	18,327	22,374	3276
<i>Mikania decumbens</i> Malme	Oliveira 783 (SPF)	MT793835	152,056	83,772	24,998	18,288	22,337	3266
<i>Mikania fasciculata</i> C.T.Oliveira & Pirani	Oliveira 977 (SPF)	MT793852	152,070	83,747	24,996	18,331	22,307	3249
<i>Mikania glomerata</i> Spreng.	Oliveira 917 (SPF)	MT793836	151,773	83,675	24,938	18,222	22,374	3250
<i>Mikania haenkeana</i> DC.	Oliveira 897 (SPF)	MT793837	151,865	83,717	24,911	18,326	22,340	3252
<i>Mikania lehmannii</i> Hieron.	Oliveira 891 (SPF)	MT793838	152,062	83,878	24,960	18,264	22,369	3250
<i>Mikania neurocaula</i> DC.	Fernandes 59 (BHCB)	MT793839	152,045	83,764	24,995	18,291	22,379	3266
<i>Mikania oblongifolia</i> DC.	Oliveira 966 (SPF)	MT793840	151,845	83,527	24,998	18,322	22,365	3267
<i>Mikania obtusata</i> DC.	Oliveira 809 (SPF)	MT793841	152,106	83,817	25,003	18,283	22,406	3266
<i>Mikania parvifolia</i> Baker	Fernandes 128 (BHCB)	MT793842	152,023	83,725	24,992	18,314	22,386	3261
<i>Mikania purpurascens</i> (Baker) R.M.King & H.Rob.	Oliveira 974 (SPF)	MT793853	152,037	83,746	25,005	18,281	22,337	3249
<i>Mikania salvifolia</i> Gardner	Oliveira 813 (SPF)	MT793843	152,229	83,870	25,014	18,331	22,211	3240
<i>Mikania smaragdina</i> Dusén ex Malme	Oliveira 798 (SPF)	MT793854	152,020	83,745	24,975	18,325	22,384	3251
<i>Mikania sylvatica</i> Klatt.	Oliveira 981 (SPF)	MT793855	152,099	83,808	24,972	18,347	22,363	3257
<i>Mikania ternata</i> (Vell.) B.L.Rob.	Oliveira 806 (SPF)	MT793844	151,957	83,791	24,981	18,204	22,324	3250
<i>Mikania triangularis</i> Baker	Oliveira 916 (SPF)	MT793845	151,956	83,789	25,002	18,163	22,368	3252
Outgroups								
<i>Ageratina fastigiata</i> (Kunth) R.M.King & H.Rob.	Oliveira 980 (SPF)	MT793847	152,359	84,009	25,022	18,306	22,423	3,304
<i>Litothamnus nitidus</i> (DC.) W.C.Holmes	Oliveira 971 (SPF)	MT793848	151,517	82,844	25,046	18,581	22,326	3223
<i>Stevia collina</i> Gardner	Oliveira 769 (SPF)	MT793846	151,248	83,386	24,791	18,280	22,303	3280

Table 1. Summary of the plastomes sequenced in this study.

trnS^{GCU}trnC^{GCA} and *trnE^{UUC}* (Table 1, Supplementary Fig. S1). All plastomes sequenced in this study encode 113 unique genes, including 79 protein-coding genes (CDS), 17 of which contain introns, 30 tRNA genes, and four rRNA genes (Table 2). The plastomes of *Mikania* and the other three Eupatorieae have identical structure and order. The boundaries between the four main plastome regions are very conserved within *Mikania* species and among the three Eupatorieae genera sampled here: the LSC/IRb border is within *rps19*, the IRb/SSC is within *ycf1*, the SSC/IRa is between *ndhF* and a partial *ycf1* (*ψycf1*), and the IRa/LSC is between a truncated *rps19* (*†rps19*) and *trnH^{GUG}* (Supplementary Figs. S1, S2).

Variable regions across plastomes. Pairwise comparison of divergent regions performed in mVISTA within three selected *Mikania* plastomes and between the Eupatorieae genera sequenced here reveals low intra-generic sequence divergence within *Mikania* and higher sequence variation among genera and in noncoding regions, except for the *ycf1* gene (Supplementary Fig. S3). The nucleotide variability (π) values within 800 bp across the plastomes range from 0 to 0.013, with a mean value of 0.0036. We identified only two regions with $\pi > 0.01$ (*rpl32-ndhF* and *rpl16-rps3*) and six regions with π values around 0.009 (*rbcl*, *ycf1*, *petN-psbM*, *rps16-trnQ^{UUG}*, *trnH^{GUG}-psbA*, *atpI-atpH*) (Fig. 1A).

In alignments of the 19 complete *Mikania* plastomes assembled here, the noncoding regions are more variable (i.e., 3.15% of the intergenic regions and 2.24% of the introns) than the coding regions (i.e., 1.34% of the protein-coding genes; Table 3). Among the 79 protein-coding genes, the ten genes with the highest percentage of variable sites are: *ycf1* (3.83%), *rpoA* (3.43%), *psbT* (2.94%), *ndhF* (2.43%), *rpl32* (2.42%), *ccsA* (2.37%), *rbcl* (2.33%), *matK* (2.24%), *rps3* (2.13%), and *rpl20* (2.1%) (Fig. 1B,C; Supplementary Tables S1, S2).

Analyses of SSR and tandem repeats. In the 19 *Mikania* plastomes, the total number of SSRs range from 34 to 44 SSRs, while 51, 49, and 38 SSRs are recovered in *A. fastigiata*, *L. nitidus*, and *S. collina*, respectively (Supplementary Fig. S4A–C). The most abundant SSRs are A or T mononucleotide repeats, which account for 54.3–69% of the total SSRs in *Mikania*, 70.6% in *A. fastigiata*, 53.1% in *L. nitidus*, and 52.6% *S. collina*; G or C repeats, on the other hand, are rare (Supplementary Table S3). Among the total number of SSR motifs in *Mikania*, 20–29 (57.5–69%) are mono-repeats, 4–6 (9.5–14.3%) are di-repeats, 2–5 (5.4–12.5%) are tri-repeats, 5–7 (12.2–17.6%) are tetra-repeats, 0–1 (0–2.3%) is penta-repeat, and 0–1 (0–2.9%) is hexa-repeat (Supplementary Fig. S4B, Supplementary Table S3). Furthermore, most of the SSRs in the *Mikania* species are located in the LSC

Gene function	Gene type	Gene
Self-replication	Ribosomal RNA genes	<i>rrn4.5^c, rrn5^c, rrn16^c, rrn23^c</i>
	Transfer RNA genes	<i>trnA-UGC^{a,c}, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnM-CAU, trnG-GCC, trnG-UCC^a, trnH-GUG, trnI-CAU^f, trnI-GAU^{a,c}, trnK-UUU^b, trnL-CAA^c, trnL-UAA^a, trnL-UAG, trnM-CAU, trnN-GUU^c, trnP-UGG, trnQ-UUG, trnR-ACG, trnR-UCU^f, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC^c, trnV-UAC^c, trnW-CCA, trnY-GUA</i>
	Small ribosomal subunit	<i>rps2, rps3, rps4, rps7^c, rps8, rps11, rps12^{b,c}, rps14, rps15, rps16^a, rps18, rps19^d</i>
	Large ribosomal subunit	<i>rpl2^{a,c}, rpl14, rpl16, rpl20, rpl22, rpl23^{a,c}, rpl32, rpl33, rpl36</i>
	RNA polymerase subunits	<i>rpoA, rpoB, rpoC1^a, rpoC2</i>
Photosynthesis	Photosystem I	<i>psaA, psaB, psaC, psaI, psaJ, ycf3^b, ycf4</i>
	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	NADH-dehydrogenase	<i>ndhA^a, ndhB^{a,c}, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	Cytochrome b6/f complex	<i>petA, petB^a, petD^b, petG, petL, petN</i>
	ATP synthase	<i>atpA, atpB, atpE, atpF^a, atpH, atpI</i>
	Rubisco	<i>rbcl</i>
Other genes	Translational initiator factor	<i>infA</i>
	Maturase	<i>matK</i>
	Protease	<i>clpP^b</i>
	Envelope membrane protein	<i>cemA</i>
	Subunit of Acetyl-CoA-carboxylase	<i>accD</i>
	c-type cytochrome synthesis	<i>ccsA</i>
Unknown function	Conserved open read frames	<i>ycf1^d, ycf2^c, ycf15^c</i>

Table 2. Genes encoded by the *Mikania* species, *Ageratina fastigiata*, *Litothamnus nitidus*, and *Stevia collina* plastomes. ^aGene with one intron. ^bGene with two introns. ^cGene duplicated. ^dGene partially duplicated.

region (70.3–83.8%), while the IR regions include between 0 and 12.2% of the SSRs, and the SSC region includes between 9.1 and 29.7%. Yet, the relative density of SSRs in the LSC is somewhat similar to that found in the SSC when considering the size of each region (Supplementary Fig. S4A–C, Supplementary Table S3). In the three other Eupatorieae genera sequenced here, 73.5–74.5% of the SSRs are located in the LSC, 2–11.8% in the IRs, and 13.7–24.5% in the SSC region (Supplementary Fig. S4A–C, Supplementary Table S3).

We also used REPuter to identify tandem repeat sequences longer than 30 bp in the plastomes sequenced here. In all 22 plastomes, repeats with 30–33 bp are the most common. Most repeats are found in the LSC, a few in the IRs, and none in the SSC (Supplementary Fig. S4D,E, Supplementary Table S4). The total number of repeats in *Mikania* range between 17 and 45, with maximum sizes of 48 bp in all *Mikania* species (Supplementary Fig. S4D,E, Supplementary Table S4). The *Mikania* plastomes contain 8–23 forward repeats, 6–14 palindrome repeats, 0–13 reverse repeats, with complement repeats being rare, 0–2 (Supplementary Table S4). The total number of repeats in *A. fastigiata* is 31 bp, while in *L. nitidus* and *S. collina* is 17 bp. The maximum repeat size in *A. fastigiata* is 46 bp, in *L. nitidus* is 58 bp, while in *S. collina* is 48 bp (Supplementary Fig. S4D,E, Supplementary Table S4).

Phylogenetic relationships of twenty *Mikania* species. Phylogenetic analyses using two different methods, concatenated maximum likelihood and multispecies coalescent, and three datasets (only coding regions, only non-coding regions and both combined) generated six different topologies with different degrees of support (Fig. 2). *Ageratina fastigiata*, *Litothamnus nitidus*, *Stevia collina*, and *Helianthus annuus* (NC007977; Heliantheae) were used as outgroups and the trees were rooted using *H. annuus* (Fig. 3). *Mikania* emerges as monophyletic and all trees present three generally well-supported main clades (bootstrap support (BS) $\geq 90\%$, local posterior probabilities (LPP) ≥ 0.95) containing the same species: Clade I (*Mikania sylvatica* and *M. lehmannii*), Clade II (*M. brevifauca*, *M. salviifolia*, *M. fasciculata*, *M. purpurascens*, *M. ternata*, *M. micrantha*, *M. decumbens*) and Clade III (*M. parvifolia*, *M. smaragdina*, *M. triangularis*, *M. additicia*, *M. obtusata*, *M. neurocaula*, *M. burchellii*, *M. oblongifolia*, *M. glomerata*, *M. haenkeana*, *M. decora*). Within each clade, the relationships between some species pairs are stable, but the position of some taxa (e.g., *M. smaragdina*, *M. ternata*) consistently change, especially in Clade III (Fig. 2). The adjusted Robinson-Foulds distances fitted in a multidimensional scaling model show that topologies are considerably different among themselves, especially the three coalescent trees (Fig. 4, Supplementary Table S5). Although not directly comparable, support values are generally higher in the concatenated analyses (BS) than in the multispecies coalescent analyses (LPP), and in the total dataset in comparison with the coding-only or non-coding-only datasets (Fig. 2). Gene discordance analyses ran with the coalescent trees show a high level of incongruence between the species tree and the gene trees, especially in the dataset containing only coding regions (Fig. 5, Supplementary Fig. S5A).

Discussion

In this study, we assembled 19 complete plastomes of *Mikania* species and of three other species from tribe Eupatorieae (i.e., *Ageratina fastigiata*, *Litothamnus nitidus*, and *Stevia collina*), and conducted phylogenetic analyses with different datasets and inference methods. The organization of *Mikania* plastomes is similar across

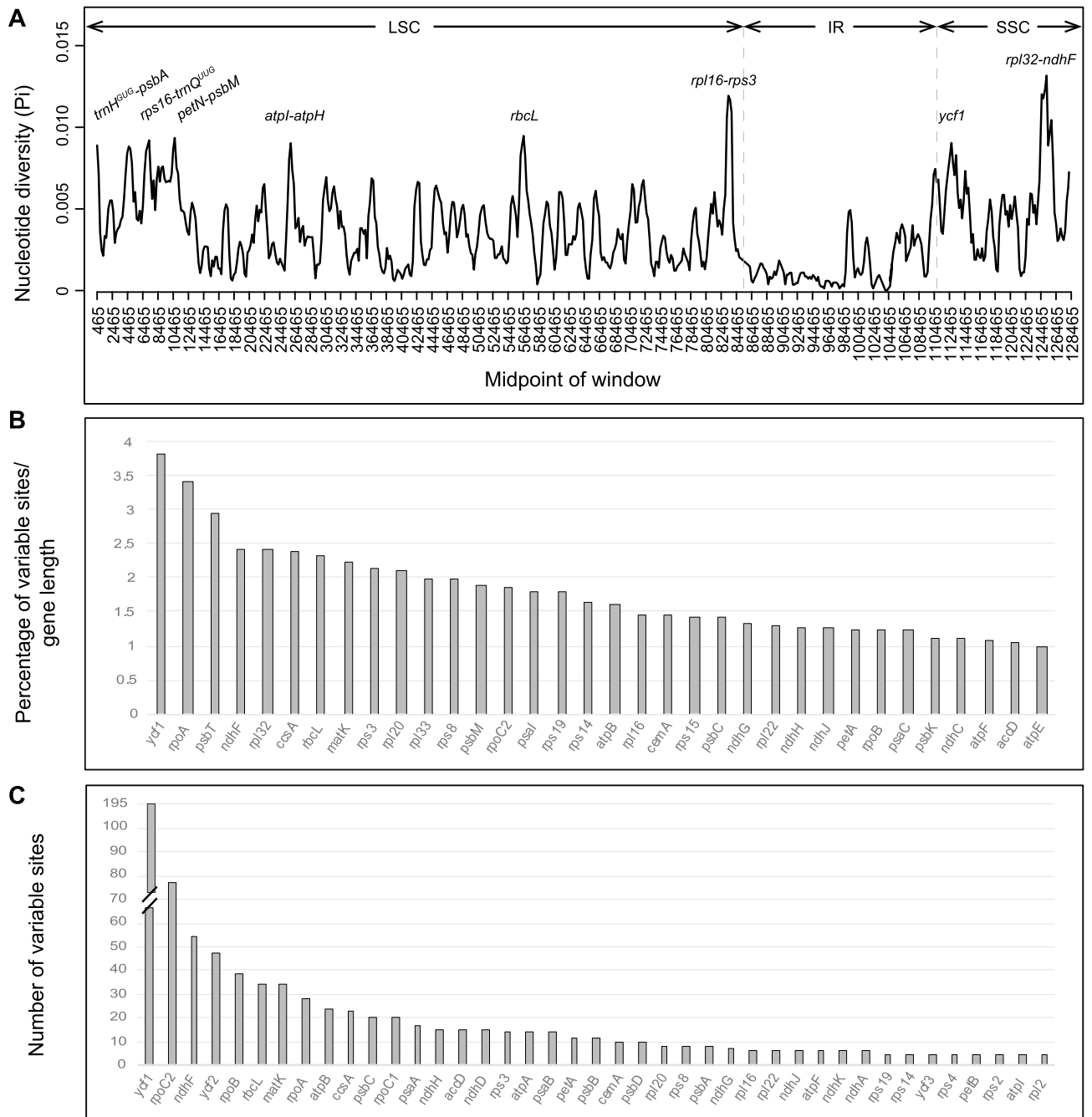


Figure 1. (A) Sliding window analysis of the chloroplast genomes of 20 *Mikania* plastomes (i.e., 19 sequenced here plus *M. micrantha*, NC031833.1) (window length: 800 bp, step size: 200 bp). X-axis, position of the midpoint of a window; Y-axis, nucleotide diversity (π) in each window. (B,C) Most variable protein-coding genes within the 20 *Mikania* plastomes. (B) Genes with up to 1% of variable sites. (C) Genes with up to five variable sites.

Dataset	Length (bp)	Var. sites	Var. sites %	Pi sites	GC%
Plastomes (LSC/IR/SSC)	128,486	2966	2.31	1000	36.5
79 genes	67,822	911	1.34	288	37.8
Intergenic regions	62,897	1981	3.15	559	35.2
Introns	13,637	306	2.24	88	34

Table 3. Summary of datasets including only the 19 *Mikania* plastomes including length, number of variable sites (Var. sites), percentage of variable sites (Var. sites %), parsimony informative sites (Pi sites), and percentage of GC content (GC%).

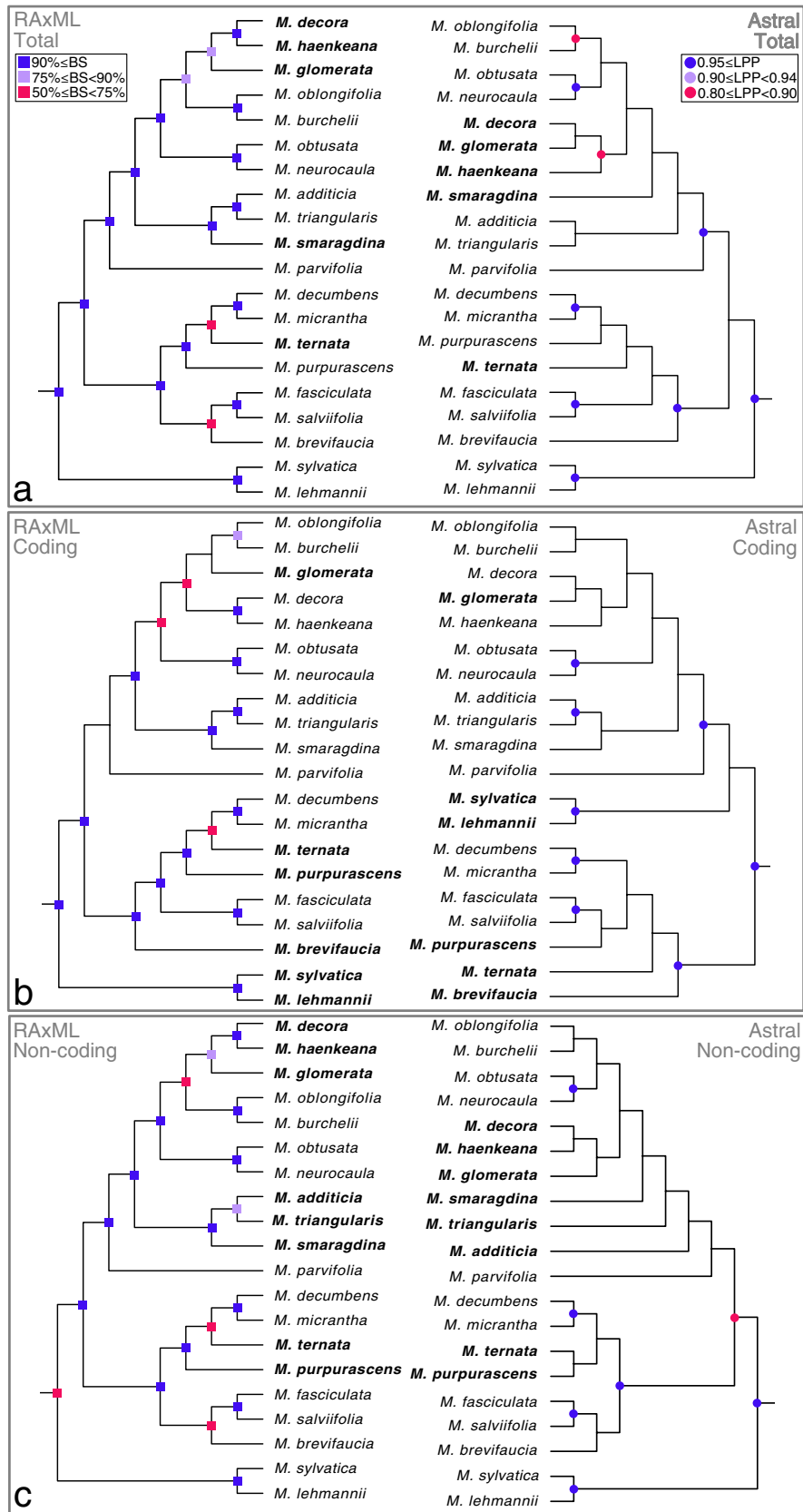


Figure 2. Phylogenetic analyses of *Mikania* using different data partitions and inference methods. RAxML refers to concatenate maximum likelihood analyses and Astral refers to multispecies coalescent inference. Support values are color-coded and nodes without symbols correspond to unsupported branches (BS < 50%, LPP < 0.8). (a) Analyses carried out with the whole plastome sequence with one of the IRs removed. (b) Analyses conducted only with the coding regions (CDS) of the plastome. (c) Analyses carried out only with non-coding regions of the plastome (intergenic regions and introns).

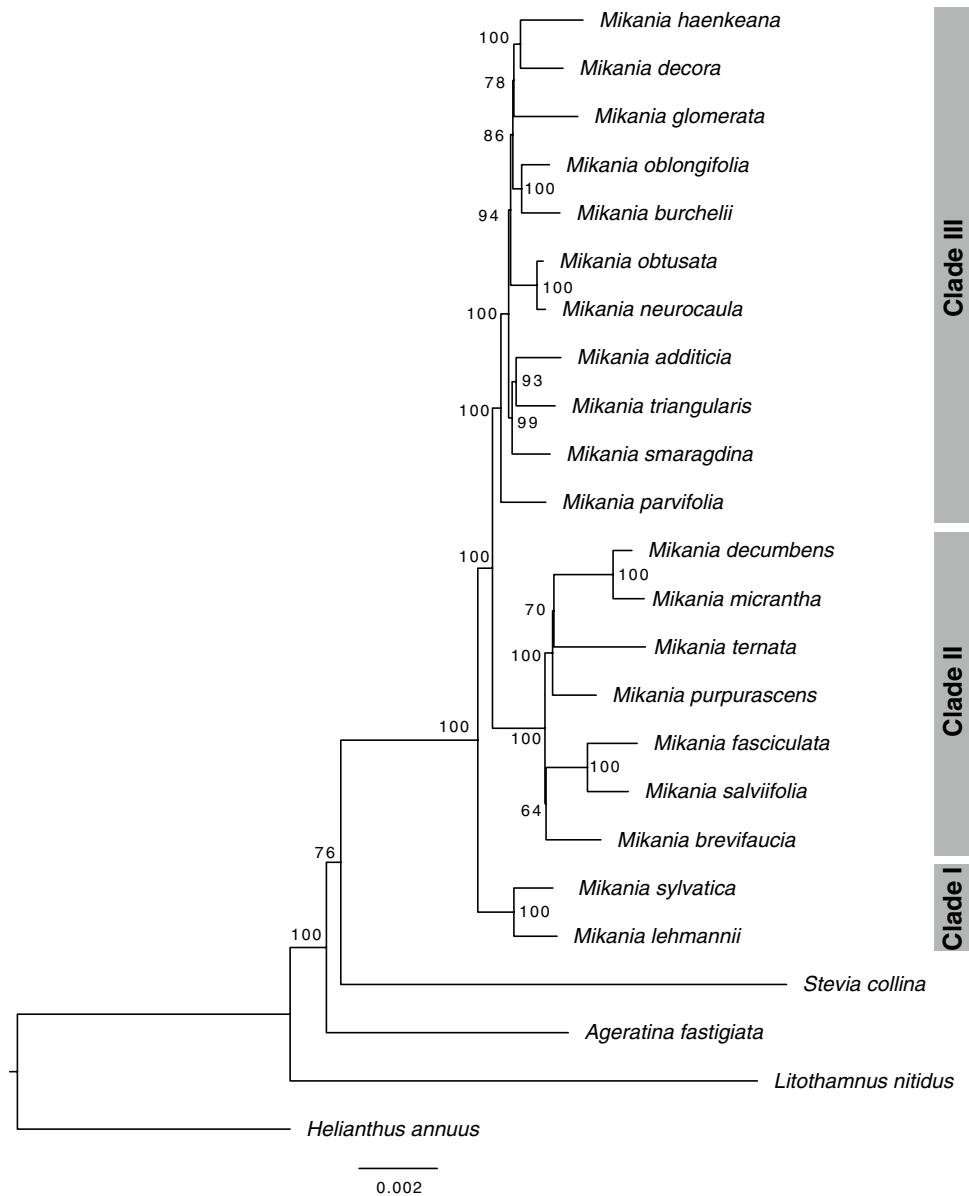


Figure 3. Representative topology of phylogenetic relationships in *Mikania*, showing three main clades. This tree represents the concatenated maximum likelihood analysis conducted with the whole plastome sequence with one of the IRs removed. Bootstrap support values shown in each node.

the studied species and to other Asteraceae plastomes. The overall genomic structure among *Mikania* plastomes is very conserved, including similar lengths, boundaries between the SC/IR regions, and number of duplicated genes in the IRs (Tables 1, 2, Supplementary Figs. S1–S3). All plastomes assembled here show the two inversions in the LSC present in most Asteraceae taxa, except for the early diverging tribe Barnadesieae^{30,35,36}. These inversions present a very conserved structure, including the same genes and similar sizes, in all of the 22 plastomes reported here (Table 1), as well as when comparing with more distantly related Asteraceae genomes, such as *Helianthus annuus* (NC007977³⁰) and *Lactuca sativa* (NC007578³⁰). On the other hand, in *Ageratina adenophora*²⁵ and *Praxelis clematidea*²⁶, these inversions do not start between the *trn*^{GCU} and *trn*^{GCA} genes, as in other Asteraceae taxa, but between the *trn*^{GCA} and *petN* genes^{25,26}. We also noticed an inversion within the *ycf1* gene in the SSC region in the *Ageratina adenophora*²⁵ and *Praxelis clematidea*²⁶ plastomes, which was not observed in the plastomes assembled here, nor in *M. micrantha* (NC031833.1²²) or *Ageratum conyzoides* (MK905238²⁷) (Supplementary Fig. S6).

The gene content found in the 22 plastomes assembled here resembles that found in other Asteraceae genomes^{25,31,33}. They encode 113 unique genes, including 79 protein-coding genes (CDS), 30 tRNA genes, and four rRNA genes. All plastomes include 17 intron-containing genes (14 contain one, while three contain two introns; Table 3). Within Eupatorieae, a duplication of the *trn*^{F_{GAA}} gene was detected in *Ageratina adenophora*²⁵ and *Praxelis clematidea*²⁶, which was previously reported for other Asteraceae subfamilies (Carduoideae,

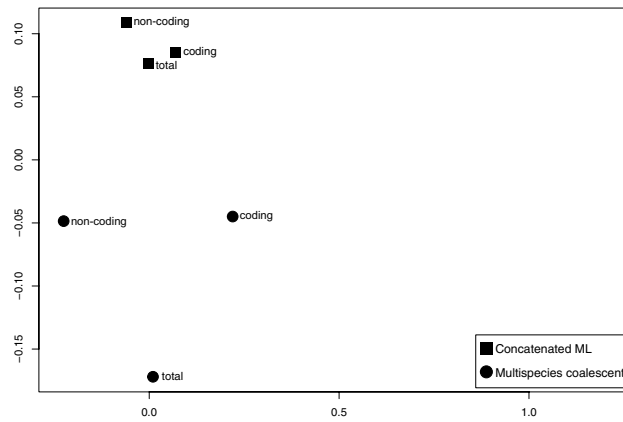


Figure 4. Multidimensional scaling of adjusted Robinson-Foulds values calculated from the pairwise comparison of all rooted trees. Coding: analyses conducted only with the coding regions (CDS). Non-coding: analyses conducted with non-coding regions (intergenic regions and introns). Total: whole plastome sequence with one IR removed.

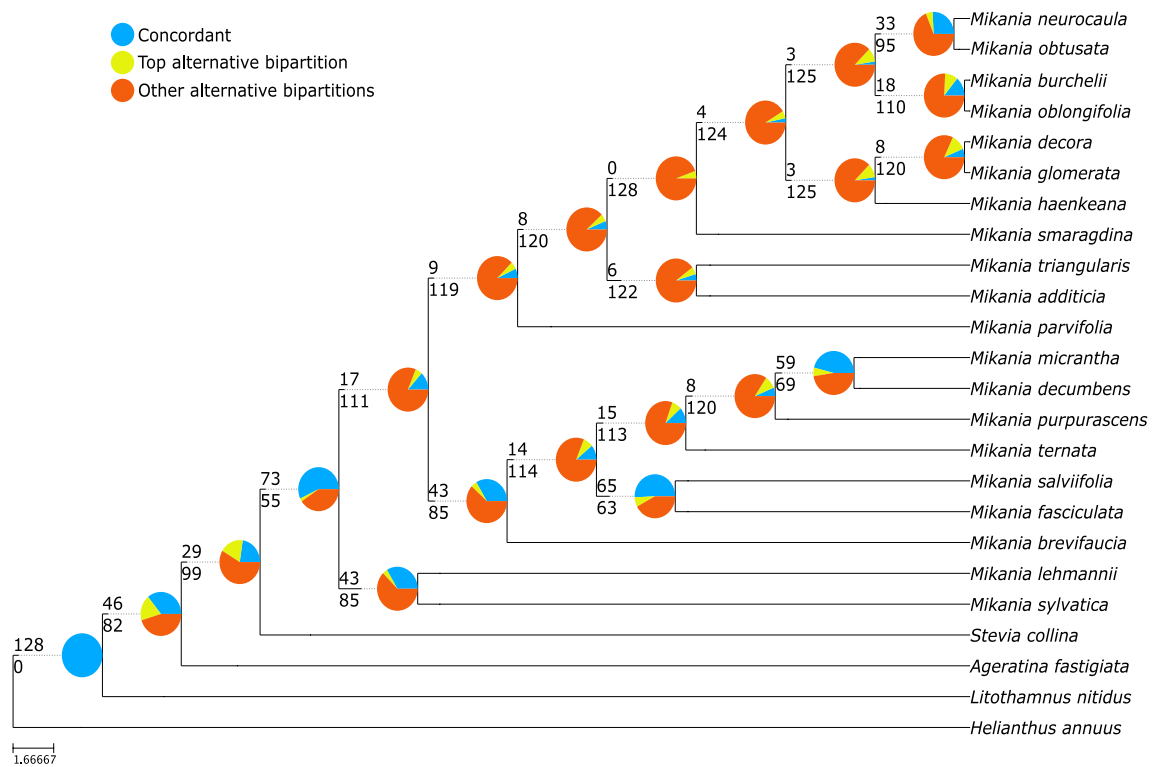


Figure 5. Gene tree discordance analysis conducted with the total dataset and multispecies coalescent inference (Astral total). The number above the branch indicates the number of concordant gene trees and the one below the number of conflicting gene trees. Pie charts indicate the proportion of gene trees supporting that clade (blue), the proportion that supports the main alternative topology for that clade (yellow), and the proportion that supports all other topologies (orange).

Cichorioideae, Asteroideae, and Heliantheae alliance)³⁷. The *rpoC1* gene in *Ageratina adenophora*²⁵, *Ageratum conyzoides*²⁷, and *Praxelis clematidea*²⁶ contains two introns, while in all plastomes assembled here it contains only one intron, similarly to other Asteraceae plastomes sequenced to date^{25,33}. Previous plastome comparative studies within Heliantheae detected a ~450 bp deletion in the *ycf2* gene for some taxa^{30,33}, which was not observed in the previously published Eupatorieae plastomes or those newly sequenced here.

The nucleotide variability is relatively low within *Mikania* plastome sequences (mean π value of 0.0036). Yet, another comparative plastome study in Asteraceae, with 36 species of *Aldama* (Heliantheae), found an even lower mean π value, 0.00118³³. The *rpl32-ndhF* and *rpl16-rps3* intergenic regions are the most variable loci

found within *Mikania* plastomes, making them candidate markers for phylogenetic studies at the species level within the genus. Other regions with higher nucleotide variability within *Mikania* are: *rbcl*, *ycf1*, *petN-psbM*, *rps16-trnQ^{UUG}*, *trnH^{GUG}-psbA*, and *atpI-atpH* (Fig. 4A). The noncoding regions are more variable than the coding regions, as expected³⁸ (Supplementary Fig. S3, Supplementary Table S2). Some of the noncoding regions that are variable within *Mikania* have been reported to be likely useful for molecular studies at lower taxonomic levels^{39,40}. Considering only the coding regions and the percentage of variable sites, the *ycf1* gene is the most variable (3.83%) followed by *rpoA* (3.43%) (Fig. 4B,C, Supplementary Table S1). The *ycf1* gene is well known as a variable coding region at lower taxonomic levels, including within Asteraceae, and has been used in phylogenetic studies within distinct plant groups^{33,41}. In addition, the *rpo* genes (*rpoA*, *rpoB*, *rpoC1*, and *rpoC2*) have been previously reported to be relatively rapidly evolving^{42,43} and divergent within Asteraceae^{25,33,41}.

The number of Single Sequence Repeats (SSRs), 34 to 44, identified within *Mikania* plastomes is similar to that reported for other Asteraceae, such as within Heliantheae, where 38–57 SSRs were found in a study with 15 species³³. In all plastomes assembled here, most SSRs found are mononucleotide repeats (59–61% within *Mikania* and 22–37% in the other genera), followed by tetranucleotide repeats (5–7% within *Mikania* and 57–72% in the other genera). The A or T motifs are the most common SSR repeat, in agreement with other studies^{31,44} (Supplementary Fig. S4A–C, Supplementary Table S3). Dispersed repeats are considered to have important influence in genome structure, size, recombination, and rearrangements²⁵. The number of repeats ≥ 30 bp found in the plastomes sequenced in this study range from 17 to 45. The maximum repeat size found within all *Mikania* species was 48 bp (Supplementary Fig. S4D,E, Supplementary Table S4). In *Myriopholis dioica* (Pertyae) 58 repeats ≥ 20 bp were found and the maximum repeat length was the same found for *Mikania* (48 bp)³¹. In *Lactuca sativa* and *Helianthus annuus*, 15 and 33 repeats ≥ 23 bp were found, respectively, of which most were smaller than 40 bp, with only two larger than 90 bp³⁰. In *Ageratina denophora*, 59 repeats ≥ 15 bp were found, most ranging between 15 and 50 bp, but repeats > 100 bp were also present²⁵.

The phylogenetic analyses performed here sampled only 20 out of ~450 *Mikania* species (Fig. 3). Yet, the relationships within this genus were never investigated using complete plastomes and represent an advance in our knowledge of infrageneric evolutionary relationships. The only phylogenetic study of the genus to date²¹ was focused on species delimitation of a few highly variable taxa, such as *M. micrantha* and *M. cordifolia*, based on AFLPs and two nuclear ribosomal markers. The divergent sampling, with only four species overlapping between both studies, precludes a proper comparison between the topologies from the previous study and the ones found here. The differences in the genomic compartment used by both studies further hinders a proper comparison, given the frequent occurrence of discrepancies between nuclear and plastidial phylogenies. The comparison among trees obtained with different reconstruction methods and datasets show a scenario of incongruence among topologies, especially in the higher nested clades (Figs. 2, 4). The backbones of most trees show Clade I as sister to a clade formed by Clade II and Clade III, except for the Astral coding tree, which shows Clade II as sister to Clade I and Clade III (Fig. 2B). Clade II presents the same relationships in all three concatenated analyses, while in the coalescent trees the relative positions of some taxa, such as *Mikania ternata* and *M. purpurascens*, change in all trees, but especially when comparing the coding tree with the non-coding and total trees (Fig. 2). The relationships within Clade III are responsible for most of the incongruence among all six trees, as they change in each topology. Some species emerge as sisters in most topologies, such as *M. decora* + *M. haenkeana*, *M. oblongifolia* + *M. burchelii*, *M. obtusata* + *M. neurocaula* and *M. additicia* + *M. triangularis*. Similarly, *M. parvifolia* emerges as sister to all other species in Clade III in all analyses, but the relative positions of *M. smaragdina* and *M. glomerata* are variable across all topologies, usually with *M. smaragdina* being close to *M. additicia* and *M. triangularis*, and *M. glomerata* close to *M. decora* and *M. haenkeana* (Fig. 2).

The gene tree discordance analyses (Fig. 5, Supplementary Fig. S5) show strong discordance across the three datasets (total, coding, non-coding), with few gene trees agreeing with the relationships shown in the species tree. The multispecies coalescent has been extensively used in the context of multi-locus phylogenies obtained from target capture data⁴⁵, but few studies have applied it to plastid data, due to the widespread misconceptions about the lack of biparental inheritance and recombination in this organelle¹¹. Among recent studies that used the multispecies coalescent in plastid data, three of them refer to higher-level phylogenies, i.e., among Angiosperms⁴⁶, among Rosids¹⁰ and among tribes of Asteraceae⁴⁷, while one deals with a single genus⁴⁸. Most of these studies found incongruences between concatenated and coalescent analyses, but only two of them presented information about gene tree/species tree discordance^{46,47}, both showing wide discordance between the inferred species tree and the gene trees.

Walker et al.⁴⁶ proposes that uninformative genes are one of the reasons for discordance, and this is likely one of the issues in our trees. In the coding dataset, the number of variable sites in each gene across species of *Mikania* varies from 0 to 3.83%, while the non-coding dataset presents a little more variation, from 0.5% to 8.8% (Supplementary Tables S1, S2). The large variation in gene tree topologies, summarized by the gene tree discordance analysis (Fig. 5, Supplementary Fig. S5), leads to weakly resolved gene trees and consequently to poorly supported species trees, as the calculation of local posterior probabilities (LPP) depends on the concordance among the three possible topologies for a determined quartet of branches⁴⁹. The length of each individual locus alignment also influences on the degree of conflict, as shorter loci tend to have less informative sites, contributing to the lack of resolution in gene trees⁴⁶. In our dataset, these two factors seem to be correlated, with most of the shorter alignments presenting very few variable sites (Supplementary Fig. S7, Supplementary Table S1, S2).

Recent studies have shown that in concatenated analysis, a few outlier genes can drive topology inference^{45,50}. Our concatenated analyses present topologies more similar to each other than the coalescent topologies, regardless of the dataset (Fig. 4). The concatenated coding tree is more similar to the concatenated total topology (Figs. 2, 4), possibly indicating that one or more specific genes are responsible for defining most of the topology, whereas the three coalescent topologies are all different from each other, due to lack of resolution in individual gene trees. Incomplete lineage sorting (ILS) is usually considered a source of conflict in the multispecies

coalescent⁴⁵, and one metric that can be used to assess its occurrence is the normalized quartet score of a coalescent tree, which measures the percentage of quartet trees found in the species tree from all calculated quartet trees⁵¹. The normalized quartet scores calculated for the three Astral topologies are considered very high (total: ~ 52%, coding: ~ 49%, non-coding: ~ 57%), but considering the sparse sampling in our study, which included ca. 0.45% of *Mikania* species, it is difficult to delimit the occurrence of ILS in relation to the lack of actual sampling. Further studies with more complete sampling could help untangle cases of ILS and lack of resolution due to uninformative genes, by also increasing the likelihood of sequence variation.

Although applying the multispecies coalescent methods to chloroplast sequences makes sense biologically, due to the possibility of evolutionary process that could lead to different parts of the genome evolving in different rates, in practice the results tend to be confounding, as seen here and in previous studies^{10,46}. The causes of plastome conflict are still poorly understood⁴⁶, and in lower-level phylogenies, as the case presented here, it might be hard to untangle sources of conflict inherent to plastome biology from lack of sequence variability due to rapid radiations over short evolutionary times. In *Mikania*, where the phylogenetic relationships are poorly known, especially in relation to the nuclear genome, it is difficult to map out other potential root causes for conflict, such as hybridization, plastome capture, or incomplete lineage sorting. An expanded sampling, both in terms of species and genome compartment (e.g., adding nuclear markers), could bring a clearer picture of the evolutionary relationships in the genus and of other biological factors that might impact phylogenetic reconstructions in *Mikania*.

Material and methods

Sampling, DNA preparation, sequencing, plastome assembly, and annotation. Whole genomic DNA extraction, Illumina libraries preparation, and NGS sequencing of the 19 *Mikania* and three outgroup species from other Eupatorieae genera (Table 1) follow³³. Sequencing was performed using Illumina HiSeq 2500 Genome Analyzer (Illumina, San Diego, California, USA) in paired-end mode. We assembled all plastomes using Fast-Plast 1.2.8⁵², with the following software: (i) Trimmomatic 0.32⁵³ to remove adaptors and trim low quality reads using the parameters SLIDINGWINDOW:10:20 and MINLEN:40; (ii) Bowtie2 2.1.0⁵⁴ with default parameters to select only chloroplast-like reads using the plastome of *Mikania micrantha* (NC031833.1²² as reference); (iii) SPAdes 3.1.0⁵⁵ to assemble the selected reads into contigs with k-mers of 57 and 87, using the “only-assembler” option; (iv) *afin* (<http://bitbucket.org/benine/afin/>) to assemble the contigs from the previous step with the complete reads dataset with the following parameters -l 150,50,50, -f 0.1, -d 100,—× 100, -p 20,15,10, and -i 2,1,1. We evaluated plastome coverage in Jellyfish 2.1.3⁵⁶. We annotated the sequences using Geneious 9.1.5⁵⁷, DOGMA⁵⁸, and BLAST^{59,60}, with start and stop codons checked manually. We used OGDRAW⁶¹ to prepare the graphical representation for the resulting plastome. Finally, we analyzed the boundaries between the plastome regions (i.e., LSC/IRb/SSC/IRa) using Geneious and IRscope⁶² (<https://irscope.shinyapps.io/irapp/>).

Comparative analyses of the assembled *Mikania* plastomes. We conducted comparative analyses within 20 *Mikania* plastomes (i.e., 19 sequenced here plus *M. micrantha*, NC031833.1) and among *Mikania* and the three outgroup taxa plastomes assembled in this study (i.e., *Ageratina fastigiata*, *Lithothamnus nitidus*, and *Stevia collina*). We used MAFFT 7⁶³ with the FFT-NS-2 method⁶⁴ to perform the alignment of the complete plastome sequences, with one copy of the IRs manually excluded to avoid data duplication. To search variable regions, we used mVISTA⁶⁵ with Shuffle-LAGAN⁶⁶ with the previously annotated *M. decora* plastome as reference, plus two *Mikania* species and the three outgroup taxa sequenced here. Based on the phylogeny recovered in this study, we selected one species from each of the three main recovered clades of *Mikania* (i.e., *M. decora*, *M. decumbens*, and *M. sylvatica*). We calculated nucleotide variability values (π) within 20 *Mikania* plastomes. We used DnaSP 6.10⁶⁷ to conduct a sliding window analysis with a 200 bp step size and 800 bp window length. The resulting π values were plotted using R⁶⁸. We analyzed the variable sites using MEGA 7⁶⁹ in the alignments of the 20 *Mikania* complete plastomes and of 79 protein-coding genes (Supplementary Table S1) extracted from these genomes. Each gene was extracted from the complete plastome alignment and separately re-aligned in Geneious with the ClustalW plugin⁷⁰ considering codon positions.

Analyses of repeated regions. We searched for microsatellites or Simple Sequence Repeats (SSRs; i.e., tandemly arranged repeats of short DNA motifs of 1–6 bp in length) and repeated elements using MISA⁷¹ and REPuter⁷², respectively, in the plastomes of the 19 *Mikania* species and three other Eupatorieae representatives sequenced here. We analyzed SSRs with motifs between 1 and 6 nucleotides and a minimum number of repetition units as follows: 10 for mono-, 5 for di-, and 4 for trinucleotide, and 3 for tetra-, penta-, and hexanucleotide SSRs. We identified repeated elements ≥ 30 bp (forward, palindrome, reverse, and complement) using $\geq 90\%$ of sequence identity and hamming distance = 3.

Phylogenetic reconstruction. We reconstructed phylogenetic relationships among 20 *Mikania* plastomes (i.e., 19 sequenced here plus *M. micrantha*; NC031833.1) and three species from other Eupatorieae genera assembled in this study plus *Helianthus annuus* (NC007977; Heliantheae) as outgroup. Three concatenated matrices were assembled: one containing the whole plastome sequence with one IR removed (total), one containing only the CDS regions of all 79 protein-coding genes (coding) and one containing all intergenic regions and introns (non-coding). All matrices were aligned using MAFFT 7⁶³ using the FFT-NS-2 method⁶⁴. Maximum likelihood reconstructions were carried out in RAxML 8.2.9⁷³ using the GTR + G model with node support assessed by rapid bootstrap (-f a) using 1000 non-parametric bootstrap pseudo-replicates. The multispecies pseudocoalescent model from Astral III⁵¹ was used to obtain species trees from individual gene trees. Three datasets were used in these analyses: one containing only each individual CDS region from all 79 protein-coding genes (coding),

one containing intergenic regions longer than 300 bp (non-coding), and one combining both datasets (total). Character evolution models for each gene matrix were calculated with PartitionFinder v.1.1.0^{74–76}, evaluating the GTR+G and GTR+G+I models in the RAxML version with rcluster search option and unlinked branch lengths, using the corrected Akaike Information Criterion to choose between models. Unrooted gene trees were obtained in RAxML 8.2.9, using the rapid bootstrap mode and 100 pseudo-replicates. Branch support was calculated using local posterior probabilities (LPP)⁵¹.

Gene tree discordance. Discordance between the species tree and gene trees, expressed as the proportion of gene trees presenting each of the clades found in the species tree, was calculated using phyparts with the thorough conflict analysis options (-a 1)⁷⁷. All species and gene trees were rooted using *Helianthus annuus* as outgroup using the function pxrr in the package phyx⁷⁸. The proportion of gene trees in agreement with the species tree in each node, as well as the proportion of uninformative gene trees or those supporting alternative topologies, were plotted as pie charts at each node of the tree using the phypartspiecharts.py script⁷⁹.

Topological comparisons. The adjusted Robinson-Foulds (RF) distance was used to calculate the distance among the six topologies. The RF distance was calculated between all pairs of rooted trees using PAUP* v4.0a⁸⁰ and adjusted by the number of nodes in the trees (RFadj = RF/(2n - 6)), resulting in values ranging from 0 to 1. A multidimensional scaling approach was used to observe the level of similarity among the topologies, using the “cmdscale” command in the R package “stats”, and subsequently plotted.

Data archiving statement. The complete plastome sequence data of the 19 *Mikania* plastomes and that of *Ageratina fastigiata*, *Lithothamnium nitidus*, and *Stevia collina* are available in GenBank (NCBI) with the accession numbers MT793834–MT793855.

Received: 30 September 2020; Accepted: 15 June 2021

Published online: 24 June 2021

References

- Jansen, R. K. & Ruhlman, T. A. Plastid genomes of seed plants. In *Genomics of Chloroplasts and Mitochondria: Advances in Photosynthesis and Respiration (Including Bioenergy and Related Processes)* (eds Bock, R. & Knoop, V.) 103–126 (Springer, 2012).
- Daniell, H., Lin, C. S., Yu, M. & Chang, W. J. Chloroplast genomes: Diversity, evolution and applications in genetic engineering. *Genome Biol.* **17**, 134. <https://doi.org/10.1186/s13059-016-1004-2> (2016).
- Green, B. R. Chloroplast genomes of photosynthetic eukaryotes. *Plant J.* **66**, 34–44 (2011).
- Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F. & Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **76**, 273–297. <https://doi.org/10.1007/s11103-011-9762-4> (2011).
- Reginato, M., Neubig, K. M., Majure, L. C. & Michelangeli, F. The first complete plastid genomes of Melastomataceae are highly structurally conserved. *PeerJ* **4**, e2715. <https://doi.org/10.7717/peerj.2715> (2016).
- Chumley, T. W. *et al.* The complete chloroplast genome sequence of *Pelargonium x hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol. Biol. Evol.* **23**, 2175–2190. <https://doi.org/10.1093/molbev/msl089> (2006).
- Zhu, A., Guo, W., Gupta, S., Fan, W. & Mower, J. P. Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction & loss on substitution rates. *New Phytol.* **209**, 1747–1756. <https://doi.org/10.1111/nph.13743> (2016).
- Doyle, J. J. Gene trees and species trees: Molecular systematics as one-character taxonomy. *Syst. Bot.* **17**, 144–163 (1992).
- Ramsey, A. J. & Mandel, J. R. When one genome is not enough: Organellar heteroplasmy in plants. *Annu. Plant Rev.* **2**, 1–40. <https://doi.org/10.1002/9781119312994.apr0616> (2019).
- Gonçalves, D. J. P., Simpson, B. B., Ortiz, E. M., Shimizu, G. H. & Jansen, R. K. Incongruence between gene trees and species trees and phylogenetic signal. *Mol. Phylogenet. Evol.* **138**, 219–232. <https://doi.org/10.1016/j.ympev.2019.05.022> (2019).
- Gonçalves, D. J. P., Jansen, R. K., Ruhlman, T. A. & Mandel, J. R. Under the rug: Abandoning persistent misconceptions that obfuscate organelle evolution. *Mol. Phylogenet. Evol.* **151**, 106903. <https://doi.org/10.1016/j.ympev.2020.106903> (2020).
- King, R. M. & Robinson, H. The Genera of the Eupatorieae (Asteraceae). *Monogr. Syst. Bot. Missouri Bot. Garden* **22**, 1–581 (1987).
- Robinson, H., Schilling, E. & Panero, J. L. Eupatorieae. In *Systematics, Evolution, and Biogeography of the Compositae* (eds Funk, V. A. *et al.*) 171–189 (International Association for Plant Taxonomy, 2009).
- Funk, V. A., Susanna, A., Stuessy, T. F. & Robinson, H. Classification of compositae. In *Systematics, Evolution, and Biogeography of the Compositae* (eds Funk, V. A. *et al.*) 171–189 (International Association for Plant Taxonomy, 2009).
- Honório, A. C., Quaresma, A. S. & Oliveira, C. T. Flora do Ceará, Brasil: *Mikania* (Asteraceae: Eupatorieae). *Rodriguésia* **70**, e02952017. <https://doi.org/10.1590/2175-7860201970003> (2019).
- Holmes, W. C. A proposed sectional classification for *Mikania* (Eupatorieae). In *Compositae: Systematics. Proceedings of the International Compositae Conference, Kew, 1994* (eds Hind, D. J. N. & Beentje, H. J.) 621–626 (Royal Botanic Gardens, 1996).
- Holmes, W. C. Revision of the old world *Mikania* (Compositae). *Bot. Jahrbücher für Syst.* **103**, 211–246 (1982).
- Ritter, M. R. & Miotto, S. T. S. Taxonomia de *Mikania* Willd. (Asteraceae) no Rio Grande do Sul, Brasil. *Hoehnea* **32**, 309–359 (2005).
- Robinson, H. & Holmes, W. C. Compositae-Eupatorieae 190 (3). In *Flora of Ecuador* Vol. 83 (eds Harling, G. & Andersson, L.) 1–347 (Botanical Institute, Goteborg University, 2008).
- Rufatto, L. C., Gower, A., Schwambach, J. & Moura, S. Genus *Mikania*: Chemical composition and phytotherapeutical activity. *Rev. Bras. de Farmacogn.* **22**, 1384–1403. <https://doi.org/10.1590/S0102-695X2012005000099> (2012).
- Godoy, S. M. *et al.* Phylogenetic relationships of Brazilian *Mikania* species (Asteraceae, Eupatorieae) based on multilocus DNA markers. *Bot. J. Linn. Soc.* **184**, 326–346. <https://doi.org/10.1093/botlinnean/box030> (2017).
- Huang, L., Wang, Z., Wang, T. & Su, Y.-J. The complete chloroplast genome sequence of *Mikania micrantha* (Asteraceae), a noxious invasive weed to South China. *Mitochondrial DNA B* **1**, 603–604. <https://doi.org/10.1080/23802359.2016.1209090> (2016).
- Liu, B. *et al.* *Mikania micrantha* genome provides insights into the molecular mechanism of rapid growth. *Nat. Commun.* **11**, 340. <https://doi.org/10.1038/s41467-019-13926-4> (2020).

24. Hind, D. J. N. & Robinson, H. Tribe Eupatorieae Cass. In *The Families and Genera of Vascular Plants* Vol. 7 (ed. Kubitzki, K.) 510–574 (Springer, 2007).
25. Nie, X. *et al.* Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PLoS ONE* **7**, e36869. <https://doi.org/10.1371/journal.pone.0036869> (2012).
26. Zhang, Y., Li, L., Yan, T. L. & Liu, Q. Complete chloroplast genome sequences of *Praxelis* (*Eupatorium catarium* Veldkamp), an important invasive species. *Gene* **549**, 58–69. <https://doi.org/10.1016/j.gene.2014.07.041> (2014).
27. Qiao, Z.-P., Chen, Z.-X. & Wang, Q.-Z. The complete chloroplast genome of *Ageratum conyzoides* (Asteraceae). *Mitochondrial DNA B* **4**, 3342–3343. <https://doi.org/10.1080/23802359.2019.1673241> (2019).
28. Mandel, J. R. *et al.* A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 14083–14088. <https://doi.org/10.1073/pnas.1903871116> (2019).
29. Panero, J. L. *et al.* Resolution of deep nodes yields an improved backbone phylogeny and a new basal lineage to study early evolution of Asteraceae. *Mol. Phylogenet. Evol.* **80**, 43–53. <https://doi.org/10.1016/j.ympev.2014.07.012> (2014).
30. Timme, R. E., Kuehl, J. V., Boore, J. L. & Jansen, R. K. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: Identification of divergent regions and categorization of shared repeats. *Am. J. Bot.* **94**, 302–312. <https://doi.org/10.3732/ajb.94.3.302> (2007).
31. Lin, N. *et al.* Plastome sequencing of *Myriopholis dioica* and comparison within Asteraceae. *Plant Divers.* **41**, 315–322. <https://doi.org/10.1016/j.pld.2019.07.003> (2019).
32. Kim, G. *et al.* Comparative chloroplast genome analysis of *Artemisia* (Asteraceae) in East Asia: Insights into evolutionary divergence and phylogenomic implications. *BMC Genomics* **21**, 415. <https://doi.org/10.1186/s12864-020-06812-7> (2020).
33. Loeuille, B. *et al.* Extremely low nucleotide diversity among thirty-six new chloroplast genome sequences from *Aldama* (Heliantheae, Asteraceae) and comparative chloroplast genomics analyses with closely related genera. *PeerJ* **9**, e10886. <https://doi.org/10.7717/peerj.10886> (2021).
34. Kim, K. J. & Lee, H. L. Widespread occurrence of small inversions in the chloroplast genomes of land plants. *Mol. Cells* **19**, 104–113 (2005).
35. Jansen, R. K. & Palmer, J. D. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5818–5822. <https://doi.org/10.1073/pnas.84.16.5818> (1987).
36. Jansen, R. K. & Palmer, J. D. Chloroplast DNA from lettuce and *Barnadesia* (Asteraceae): Structure, gene localization, and characterization of a large inversion. *Curr. Genet.* **11**, 553–564. <https://doi.org/10.1007/BF00384619> (1987).
37. Salih, R. H. M., Majesky, L., Schwarbacher, T., Gornall, R. & Heslop-Harrison, P. Complete chloroplast genomes from apomictic *Taraxacum* (Asteraceae): Identity and variation between three microspecies. *PLoS ONE* **12**, e0168008. <https://doi.org/10.1371/journal.pone.0168008> (2017).
38. Lahaye, R. *et al.* DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2923–2928. <https://doi.org/10.1073/pnas.0709936105> (2008).
39. Shaw, J. *et al.* The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* **92**, 142–166. <https://doi.org/10.3732/ajb.92.1.142> (2005).
40. Shaw, J., Lickey, E. B., Schilling, E. E. & Small, R. L. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *Am. J. Bot.* **94**, 275–288. <https://doi.org/10.3732/ajb.94.3.275> (2007).
41. Zhang, X. *et al.* Plastome phylogenomics of *Saussurea* (Asteraceae: Cardueae). *BMC Plant Biol.* **19**, 290. <https://doi.org/10.1186/s12870-019-1896-6> (2019).
42. Little, M. C. & Hallick, R. B. Chloroplast *rpoA*, *rpoB* & *rpoC* genes specify at least three components of a chloroplast DNA-dependent RNA polymerase active in tRNA and mRNA transcription. *J. Biol. Chem.* **263**, 14302–14307 (1988).
43. Krawczyk, K. & Sawicki, J. The uneven rate of the molecular evolution of gene sequences of DNA-dependent RNA polymerase I of the genus *Lamium* L.. *Int. J. Mol. Sci.* **14**, 11376–11391. <https://doi.org/10.3390/ijms140611376> (2013).
44. Park, S. *et al.* Contrasting patterns of nucleotide substitution rates provide insight into dynamic evolution of plastid and mitochondrial genomes of *Geranium*. *Genome Biol. Evol.* **9**, 1766–1780. <https://doi.org/10.1093/gbe/evx124> (2017).
45. Smith, S., Walker-Hale, N., Walker, J. F. & Brown, J. W. Phylogenetic conflicts, combinability, and deep phylogenomics in plants. *Syst. Biol.* **69**, 579–592. <https://doi.org/10.1093/sysbio/syz078> (2020).
46. Walker-Hale, N., Vargas, O. M., Larson, D. A. & Stull, G. W. Characterizing gene tree conflict in plastome-inferred phylogenies. *PeerJ* **7**, e7747. <https://doi.org/10.7717/peerj.7747> (2019).
47. Watson, L. E., Siniscalchi, C. M. & Mandel, J. Phylogenomics of the hyperdiverse daisy tribes: Anthemideae, Astereae, Calenduleae, Gnaphalieae, and Senecioneae. *J. Syst. Evol.* **58**, 841–852. <https://doi.org/10.1111/jse.12698> (2020).
48. Thode, V. A., Lohmann, L. G. & Sanmartín, I. Evaluating character partitioning and molecular models in plastid phylogenomics at low taxonomic levels: A case study using *Amphilophium* (Bignoniaceae, Bignoniaceae). *J. Syst. Evol.* **58**, 1071–1089. <https://doi.org/10.1111/jse.12579> (2020).
49. Sayyari, E. & Mirarab, S. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* **33**, 1654–1668. <https://doi.org/10.1093/molbev/msw079> (2016).
50. Walker, J. F., Brown, J. W. & Smith, S. Analyzing contentious relationships and outlier genes in phylogenomics. *Syst. Biol.* **67**, 916–924. <https://doi.org/10.1093/sysbio/syy043> (2018).
51. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* **19**, 153. <https://doi.org/10.1186/s12859-018-2129-y> (2018).
52. McKain, M. & *Afinit Mrmckain/Fast-Plast: Fast-Plast 1.2.6*. Zenodo (2017). <https://doi.org/10.5281/zenodo.597709>.
53. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> (2014).
54. Langmead, B. & Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. <https://doi.org/10.1038/nmeth.1923> (2012).
55. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comp. Biol.* **19**, 455–477. <https://doi.org/10.1089/cmb.2012.0021> (2012).
56. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770. <https://doi.org/10.1093/bioinformatics/btr011> (2011).
57. Kearse, M. *et al.* Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199> (2012).
58. Wyman, S. K., Jansen, R. K. & Boore, J. L. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20**, 3252–3255. <https://doi.org/10.1093/bioinformatics/bth352> (2004).
59. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
60. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
61. Lohse, M., Drechsel, O., Kahlau, S. & Bock, R. OrganellarGenomeDRAW—A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* **41**, W575–W581. <https://doi.org/10.1093/nar/gkt289> (2013).

62. Amiryousefi, A., Hyvönen, J. & Poczai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **34**, 3030–3031. <https://doi.org/10.1093/bioinformatics/bty220> (2018).
63. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. <https://doi.org/10.1093/molbev/mst010> (2013).
64. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066. <https://doi.org/10.1093/nar/gkf436> (2002).
65. Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **32**, W273–W279. <https://doi.org/10.1093/nar/gkh458> (2004).
66. Brudno, M. *et al.* Glocal alignment: Finding rearrangements during alignment. *Bioinformatics* **19**, i54–i62. <https://doi.org/10.1093/bioinformatics/btg1005> (2003).
67. Rozas, J. *et al.* DnaSP v6: DNA sequence polymorphism analysis of large datasets. *Mol. Biol. Evol.* **34**, 3299–3302. <https://doi.org/10.1093/molbev/msx248> (2017).
68. R Development Core Team. *R: A Language and Environment for Statistical Computing* (2017). <http://www.r-project.org> (Accessed 2 January 2020)
69. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874. <https://doi.org/10.1093/molbev/msw054> (2016).
70. Larkin, M. A. *et al.* Clustal W and clustal X version 2.0. *Bioinformatics* **23**, 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404> (2007).
71. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **33**, 2583–2585. <https://doi.org/10.1093/bioinformatics/btx198> (2017).
72. Kurtz, S. *et al.* REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **29**, 4633–4642. <https://doi.org/10.1093/nar/29.22.4633> (2001).
73. Stamatakis, A. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033> (2014).
74. Stamatakis, A. RAxML-VI-HPC, maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690. <https://doi.org/10.1093/bioinformatics/btl446> (2006).
75. Lanfear, R., Calcott, B., Ho, S. Y. W. & Guindon, S. PartitionFinder, combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **29**, 1695–1701. <https://doi.org/10.1093/molbev/mss020> (2012).
76. Lanfear, R., Calcott, B., Kainer, D., Mayer, C. & Stamatakis, A. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol. Biol.* **14**, 82. <https://doi.org/10.1186/1471-2148-14-82> (2014).
77. Smith, S. A., Moore, M. J., Brown, J. W. & Yang, Y. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evol. Biol.* **15**, 150. <https://doi.org/10.1186/s12862-015-0423-0> (2015).
78. Brown, J. W., Walker, J. F. & Smith, S. A. Phyx: Phylogenetic tools for unix. *Bioinformatics* **33**, 1886–1888. <https://doi.org/10.1093/bioinformatics/btx063> (2017).
79. Johnson, M. (Accessed 25 March 2020); <https://github.com/mossmatters/phyloscripts/tree/master/phyupartspiecharts> (2017).
80. Swofford, D. L. *PAUP*: Phylogenetic Analysis Using Parsimony, version 4.0 b10* (2003).

Acknowledgements

We thank Gustavo Heiden for help during field work. This research was supported by a post-doctoral fellowship from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (PNPD-CAPES) to VAT from the Programa de Pós-Graduação em Botânica da Universidade Federal do Rio Grande do Sul (UFRGS). Additional funds were provided by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Grant ID 305776/2010-0) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Grant IDs 2011/18385-0 and 2012/12325-9).

Author contributions

V.A.T., C.T.O., B.L. and C.M.S. designed the study, defined sampling, obtained samples, interpreted the results, and co-wrote the manuscript. V.A.T. assembled Illumina sequences, annotated plastomes, and performed comparative and phylogenetic analyses. C.M.S. conducted phylogenetic analyses. J.R.P. co-wrote the manuscript, supervised, and obtained funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-92727-6>.

Correspondence and requests for materials should be addressed to C.M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021