

The first complete plastid genomes of Melastomataceae are highly structurally conserved

Marcelo Reginato¹, Kurt M. Neubig², Lucas C. Majure³ and Fabian A. Michelangeli¹

¹Institute of Systematic Botany, The New York Botanical Garden, Bronx, New York, United States

²Department of Plant Biology, Southern Illinois University of Carbondale, Carbondale, Illinois, United States

³Department of Research, Conservation and Collections, Desert Botanical Garden, Phoenix, Arizona, United States

ABSTRACT

Background: In the past three decades, several studies have predominantly relied on a small sample of the plastome to infer deep phylogenetic relationships in the species-rich Melastomataceae. Here, we report the first full plastid sequences of this family, compare general features of the sampled plastomes to other sequenced Myrtales, and survey the plastomes for highly informative regions for phylogenetics.

Methods: Genome skimming was performed for 16 species spread across the Melastomataceae. Plastomes were assembled, annotated and compared to eight sequenced plastids in the Myrtales. Phylogenetic inference was performed using Maximum Likelihood on six different data sets, where putative biases were taken into account. Summary statistics were generated for all introns and intergenic spacers with suitable size for polymerase chain reaction (PCR) amplification and used to rank the markers by phylogenetic information.

Results: The majority of the plastomes sampled are conserved in gene content and order, as well as in sequence length and GC content within plastid regions and sequence classes. Departures include the putative presence of *rps16* and *rpl2* pseudogenes in some plastomes. Phylogenetic analyses of the majority of the schemes analyzed resulted in the same topology with high values of bootstrap support. Although there is still uncertainty in some relationships, in the highest supported topologies only two nodes received bootstrap values lower than 95%.

Discussion: Melastomataceae plastomes are no exception for the general patterns observed in the genomic structure of land plant chloroplasts, being highly conserved and structurally similar to most other Myrtales. Despite the fact that the full plastome phylogeny shares most of the clades with the previously widely used and reduced data set, some changes are still observed and bootstrap support is higher. The plastome data set presented here is a step towards phylogenomic analyses in the Melastomataceae and will be a useful resource for future studies.

Submitted 20 September 2016

Accepted 24 October 2016

Published 29 November 2016

Corresponding author

Marcelo Reginato,
reginatobio@yahoo.com.br

Academic editor

David Tollervey

Additional Information and
Declarations can be found on
page 12

DOI 10.7717/peerj.2715

© Copyright

2016 Reginato et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Bioinformatics, Molecular Biology, Plant Science, Taxonomy

Keywords Myrtales, Chloroplast, Melastomataceae, Plastome, Genome skimming, Phylogenomics, NGS

INTRODUCTION

The Melastomataceae Juss. has over 5,000 species distributed predominantly across the tropical regions. The observed levels of diversity, endemism or abundance of its members across different habitats make the family an important ecological group, as well as an excellent model for a variety evolutionary studies. The Melastomataceae belong in the Myrtales, where it is sister to the small CAP clade (Crypteroniaceae, Alzateaceae and Penaeaceae), which all together form a clade sister to Myrtaceae + Vochysiaceae (Berger *et al.*, 2016). Plastid markers along with the nuclear ribosomal spacers (nrETS and nrITS) have been the major, and very often the exclusive, source of phylogenetic information in the family. Melastomataceae debut in molecular phylogenies was in a Myrtales-focused study, based on a partial amino acid sequence of the *rbcS* gene (Martin & Dowd, 1986). This study was followed by a more comprehensive nucleotide-based phylogeny, where the plastid *rbcL* gene was analyzed (Conti, Litt & Sytsma, 1996). The first Melastomataceae-wide phylogeny used a plastid data set including the *rbcL* and *ndhF* genes plus the *rpl16* intron (Clausing & Renner, 2001). This plastid data set is still the most employed source of information in studies focusing on generic relationships across the family (Fritsch *et al.*, 2004; Renner, 2004; Amorim, Goldenberg & Michelangeli, 2009; Michelangeli *et al.*, 2011; Goldenberg *et al.*, 2012; Michelangeli, Ulloa & Sosa, 2014; Goldenberg *et al.*, 2015; Zeng *et al.*, 2016). Phylogenetic studies within lower lineages of Melastomataceae have predominantly used the plastid spacers *accD-psaI*, *atpF-atpH*, *psbK-psbI*, and *trnS-trnG*, along with the ribosomal spacers nrETS and nrITS (Bécquer-Granados *et al.*, 2008; Reginato, Michelangeli & Goldenberg, 2010; Kriebel, Michelangeli & Kelly, 2015; Reginato & Michelangeli, 2016). Recently, the latter data set has also been used at deeper level studies (Michelangeli *et al.*, 2013; Rocha *et al.*, 2016).

Family-wide phylogenetic studies based on plastid markers have uncovered major relationships in the Melastomataceae, with several implications to the classification and evolutionary understanding in the family. Early studies have consolidated the sister relationship of Olisbeoideae and the remaining Melastomataceae, settling on the currently accepted family circumscription (Conti, Litt & Sytsma, 1996; Angiosperm Phylogeny Group, 1998; but see Clausing & Renner, 2001 for a different perspective). Latter studies focused in some tribal re-arrangements (Fritsch *et al.*, 2004; Penneys *et al.*, 2010; Michelangeli *et al.*, 2011), generic placement (Amorim, Goldenberg & Michelangeli, 2009; Goldenberg *et al.*, 2012; Michelangeli, Ulloa & Sosa, 2014; Goldenberg *et al.*, 2015; Kriebel, 2016; Rocha *et al.*, 2016; Zeng *et al.*, 2016), phylogenetic evaluation of higher species-rich lineages (Michelangeli *et al.*, 2004; Stone, 2006; Goldenberg *et al.*, 2008; Martin *et al.*, 2008; Michelangeli *et al.*, 2008; Michelangeli *et al.*, 2013), and lower taxon phylogenies (Bécquer-Granados *et al.*, 2008; Reginato, Michelangeli & Goldenberg, 2010; Penneys, 2013; Kriebel, Michelangeli & Kelly, 2015; Gamba-Moreno & Almeda, 2014; Majure *et al.*, 2015; Reginato & Michelangeli, 2016). Even in family-wide phylogenies, the level of variation across these few sampled plastid markers is unsatisfactory, as evidenced by low statistical support among many relationships in different published analyses. This issue becomes more prominent in phylogenetic analyses of lineages within

Melastomataceae, where the plastid phylogeny is overall weakly supported, and concatenated results tend to be dominated by the more variable nuclear ribosomal data (Reginato, Michelangeli & Goldenberg, 2010; Reginato & Michelangeli, 2016).

Phylogenomic studies are sparse in the Myrtales and absent in the Melastomataceae. Currently, there are 54 full plastids of Myrtales on the NCBI database, covering three out of the nine families in the order (Lythraceae, Myrtaceae and Onagraceae). Full plastomes can potentially improve hypotheses of phylogenetic relationships within the family, as well as in the Myrtales, and provide basic information for other aspects of molecular biology (e.g., DNA barcoding, plastome evolution, development of molecular markers). Here, we present the first complete plastid genomes in the Melastomataceae, covering 16 species spread across the family. The objectives of this study are to describe the structure of the sampled plastomes; compare main features of the plastomes within the family and to other available Myrtales plastomes; and survey the plastomes for highly informative phylogenetic markers for future use.

MATERIAL AND METHODS

Taxon sampling, DNA extraction and sequencing

Genome skimming was performed for 16 species of Melastomataceae. Sampling was based on previous family wide phylogenetic studies (Michelangeli, Ulloa & Sosa, 2014; Goldenberg *et al.*, 2015), where each sample belongs to a different major lineage of the family, either with a formal tribe status or not. Voucher information along with GenBank accession codes are presented in Table 1. Total genomic DNA was isolated from silica-dried tissue using the Qiagen DNAeasy plant mini-kit (Qiagen, Valencia, CA, USA) following the protocol suggested by Alexander *et al.* (2007) or used a modified CTAB extraction where the aqueous supernatant was silica-column purified (Neubig *et al.*, 2014). Total DNA samples were quantified using a NanoDrop Spectrophotometer (Thermo Scientific, Waltham, MA, USA) or Qubit 2.0 (Invitrogen, Carlsbad, CA, USA). Total genomic libraries and barcoding was performed at Cold Spring Harbor Laboratories or at Rapid Genomics (Gainesville, FL, USA) for sequencing on an Illumina HiSeq2000 platform (Illumina, Inc., San Diego, CA, USA).

Plastid genome assembly and annotation

Total reads number yielded was on average ca. 11.5 Gb per sample (s.d. = 6 Gb). Paired reads were imported into Geneious 7.1 (Biomatters Ltd., Auckland, New Zealand), trimmed by quality (at 0.05 probability) and de novo assembled (Geneious Assembler, “low sensitivity” option, default settings). Filtered assembled contigs (length > 1 kb) were blasted against the *Eucalyptus polybractea* plastome (NC022393). The identified plastid contigs were then reference assembled against the *E. polybractea* plastome in order to generate a single contig to construct the circular maps. Eventual short gaps were filled by iteratively mapping the total paired reads against the contig ends. Plastid annotation was performed in Geneious 7.1 with *Arabidopsis thaliana* (NC000932) and *Eucalyptus polybractea* (NC022393) as references. Graphical representations of the plastid circular

Table 1 Voucher information and GenBank accessions of the chloroplast sequenced in the Melastomataceae.

Species	Tribe/"clade"	GenBank	Voucher	Herbarium
<i>Allomaieta villosa</i> (Gleason) Lozano	Cyphostyleae	KX826819	David, H. 2188	HUA, NY
<i>Bertolonia acuminata</i> Gardner	Bertoloniaeae	KX826820	Goldenberg, R. 810	NY, UPCB
<i>Blakea schlimii</i> (Naudin) Triana	Blakeeae	KX826821	Michelangeli, F.A. 1227	NY
<i>Eriocnema fulva</i> Naudin	"Eriocnema"	KX826822	Almeda, F. 8416	CAS
<i>Graffenrieda moritziana</i> Triana	Merianieae	KX826823	Michelangeli, F.A. 832	NY
<i>Henriettea barkeri</i> (Urb. & Ekman) Alain	Henrietteae	KX826824	Ionta, G. 2029	FLAS
<i>Merianthera pulchra</i> Kuhlman	"Cambessedesia"	KX826825	Goldenberg, R. 1153	NY, UPCB
<i>Miconia dodecandra</i> Cogn.	Miconieae	KX826826	Michelangeli, F.A. 758	NY
<i>Nepsera aquatica</i> (Aubl.) Naudin	"Marctia"	KX826827	Michelangeli, F.A. 1998	NY
<i>Opisthocentra clidemioides</i> Hook. f.	Unplaced	KX826828	Caddah, M.K. 578	NY, UPCB
<i>Pterogastra divaricata</i> (Bonpl.) Naudin	Melastomeae	KX826829	Michelangeli, F.A. 540	NY
<i>Rhexia virginica</i> L.	Rhexieae	KX826830	Michelangeli, F.A. 1448	NY
<i>Rhynchanthera bracteata</i> Triana	Microlicieae	KX826831	Zenteno, F. 8801	NY
<i>Salpinga maranoniensis</i> Wurdack	Merianieae	KX826832	Clark, J.L. 13577	UNA
<i>Tibouchina longifolia</i> (Vahl) Baill.	Melastomeae	KX826833	Majure, L. 4277	FLAS
<i>Triolena amazonica</i> (Pilg.) Wurdack	"Triolena"	KX826834	Michelangeli, F.A. 1366	NY

Note:

Informal clades are quoted.

and linear maps were generated with OGDRAW (Lohse et al., 2013) and the R package genoPlotR (R Development Core Team, 2016; Guy, Kultima & Andersson, 2010).

Plastome structure, gene content, and general characteristics of the plastid genome were compared among the 16 Melastomataceae plastomes and to eight published plastomes of Myrtales, covering all families in this order available on the NCBI website. The Myrtales plastomes included one species in the Lythraceae (*Lagerstroemia fauriei*–NC029808), one Onagraceae (*Oenothera grandiflora*–NC029211) and six Myrtaceae (*Allosyncarpia ternata*–NC022413; *Angophora costata*–NC022412; *Corymbia gummifera*–NC022407; *Eucalyptus polybractea*–NC022393; *Eugenia uniflora*–NC027744; and *Stockwellia quadrifida*–NC022414).

Phylogenetic analyses

Three major data sets were generated for phylogenetic inference. The first included the non-coding regions (ncs data set), the second included 78 protein-coding genes (cds data set), and the third consisted of fully assembled plastomes (full data set). In all data sets one of the IR sequences was removed to reduce overrepresentation of duplicated sequences. Full plastids were aligned with MAFFT v. 7 using the FFT-NS-i \times 1,000 strategy (Katoh & Standley, 2013). Coding sequences were extracted from the full alignment, resulting in the cds and ncs data sets. Each gene in the cds data set was re-aligned using its translation under the same strategy of the full data set and then concatenated. Given that phylogenetic inference might be biased by poorly aligned regions with ambiguous homology, heterogeneous rates of substitution in the different codon positions, synonymous substitutions in Arginine, Leucine and Serine codons, among others

(Misof & Misof, 2009; Cox et al., 2014), we further divided the three major data sets into six different schemes where we attempted to circumvent those issues. Poorly aligned regions of the ncs data set were removed using aliscore.pl with the $-N$ and $-r$ options (Misof & Misof, 2009), and in the cds data set; all codons coding for Arginine, Leucine and Serine were ambiguated. Thus, the final six schemes included: 1. all ncs data set (ncs); 2. ncs data set without poorly aligned sites (ncs filtered); 3. all cds data set (cds); 4. cds with A, L and S codons ambiguated (cds ambiguated); 5. translated cds (protein); 6. ncs filtered plus all cds non-ambiguated (full). Additionally, in order to carry out a more objective comparison with previous phylogenetic hypotheses, we also analyzed a reduced data set that included only the three more commonly used markers for family wide phylogenies in the Melastomataceae (*ndhF* and *rbcl* genes along with the *rpl16* intron, concatenated).

Phylogenetic inference for all schemes was performed using Maximum Likelihood implemented in RAxML 8.2.4 (Stamatakis, 2014). The GTR+G model was employed for all nucleotide data and the PROT+G model for the protein sequences. Support was estimated through 1,000 bootstrap replicates. Protein-coding sequences were partitioned by codon position in all schemes, while no partitioning was employed for the non-coding regions.

Phylogenetic informative regions

In order to identify and rank highly phylogenetically informative regions in the Melastomataceae plastomes, all introns (19) and variable intergenic spacers with suitable size for PCR amplification (22) were selected and compared. Each individual marker was aligned with MAFFT (FFT-NS-i \times 1,000 strategy), and its Maximum likelihood tree inferred with RAxML (not partitioned, GTR+G model, 100 bootstrap replicates). For each marker, we report the number of variable sites, number of parsimony informative sites, mean sequence distance (under K80 model), alignment length, mean sequence length, mean bootstrap support and distance to the full scheme plastid tree (RF distance; Robinson & Foulds, 1981). The metrics were retrieved using functions of the R packages ape and phangorn (Paradis, Claude & Strimmer, 2004; Schliep, 2011). Markers were ranked by phylogenetic information using a weighted mean of relative values of the following metrics: number of variable sites (weight = 1), mean bootstrap (weight = 2) and distance to the full plastid tree (weight = 3). For the top 10 markers identified in the previous step, we designed primer pairs for PCR amplification. Primers flanking the target regions were designed with Primer3, using the default settings (Rozen & Skaletsky, 2000). All metrics reported, as well primer design, considered only the ingroup (the 16 Melastomataceae plastids).

RESULTS

Plastome structure

All plastomes have a quadripartite organization, with one large single copy region (LSC), one small single copy (SSC) and two inverted repeats (IRs). A circular map of the *Miconia dodecandra* plastome is presented in Fig. 1 and linear maps of all Melastomataceae

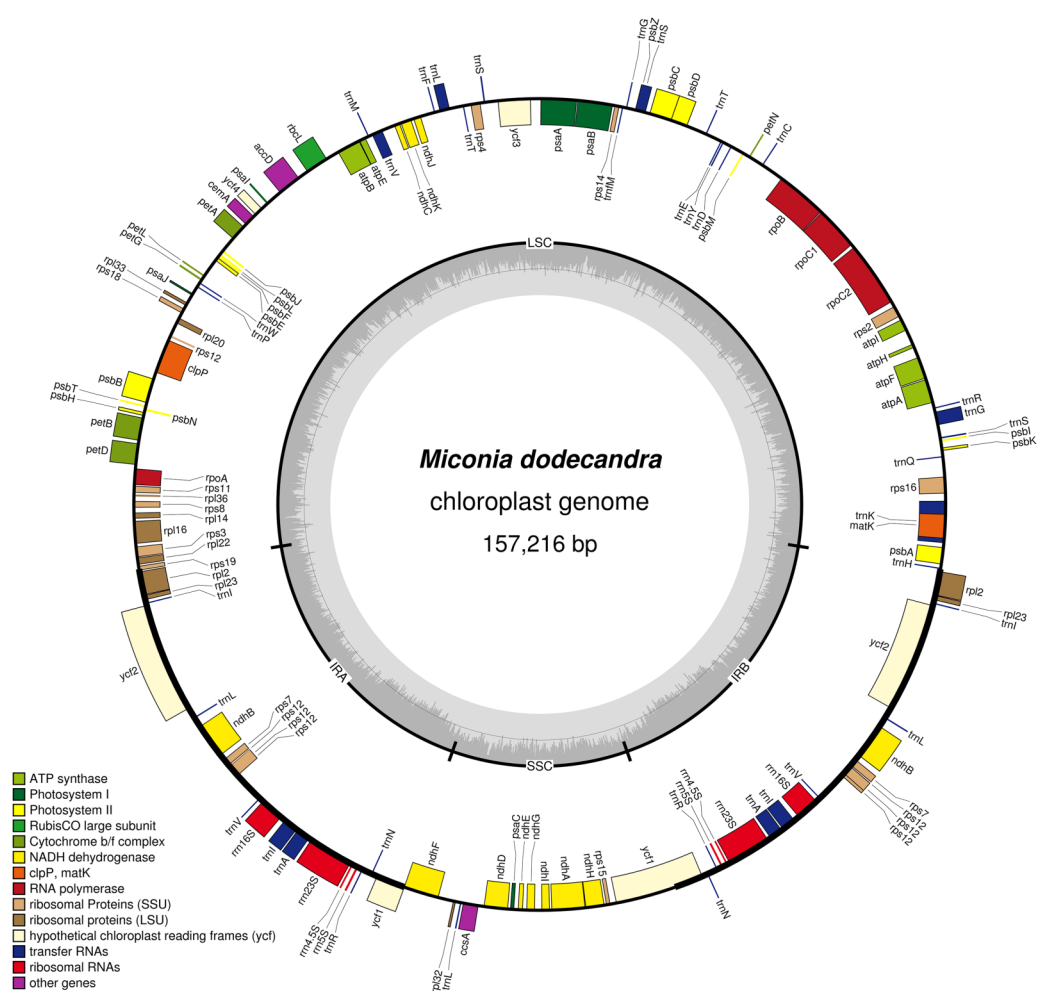


Figure 1 Map of the *Miconia dodecandra* plastid genome. Genes shown outside the outer circle are transcribed counterclockwise and genes inside the outer circle are transcribed clockwise. Genes in different functional groups are color coded following the legend. The shaded area inside the inner circle indicates the GC content, with dark shading indicating GC percent.

plastomes in Fig. 2. Sequence depth ranged from 42 to 705 (mean = 289) and plastome length from 153,311 to 157,216 bp (mean = 155,806 bp). Sequence length and GC content of the different regions across the Melastomataceae plastomes are presented in Table 2. Overall, GC content is similar across species within the same plastid region, while the LSC regions has the greatest standard deviation in sequence length (s.d. = 616 bp), followed by IR (s.d. = 250 bp) and the SSC (s.d. = 126 bp).

Most plastomes have 84 protein-coding genes (CDS), 37 transfer RNA (tRNA) and eight ribosomal (rRNA), totaling 129 genes (including duplicates and *ycf1*, *ycf2*, *ycf3* and *ycf4*). Among the duplicated genes in the IR, there are six CDS, seven tRNA, and four rRNA. As for the plastid regions, GC content is similar across different species within the same sequence class (CDS, tRNA, rRNA, intron and intergenic spacers), whereas the greatest variation in sequence length is observed across intergenic spacers (s.d. = 617 bp). A comparative summary of length and GC content in the different sequence classes across the Melastomataceae plastomes is given in Table 3. In the majority of the species sampled,

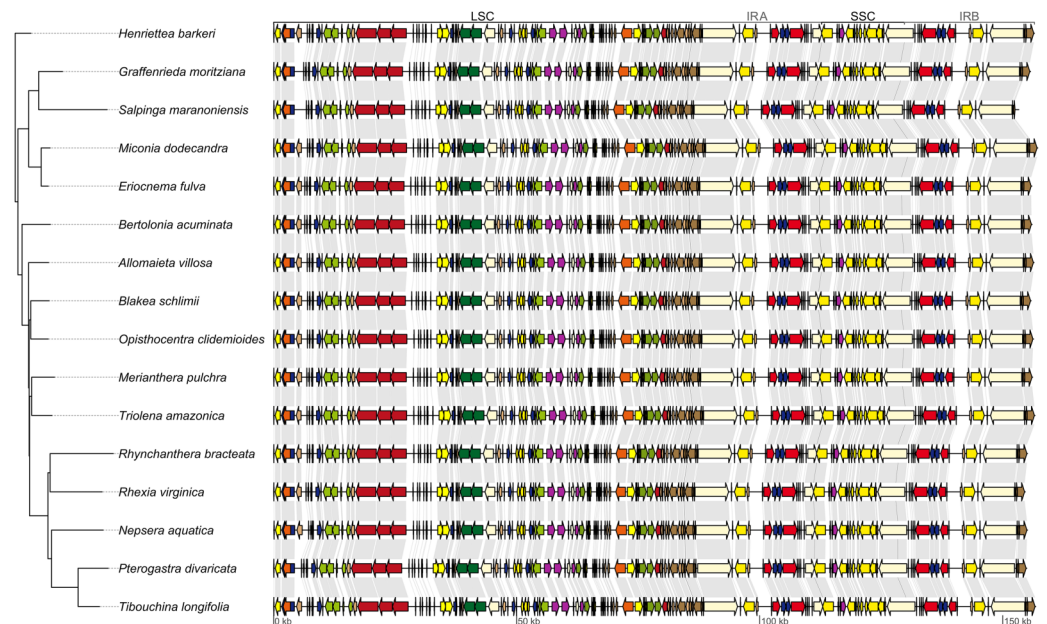


Figure 2 Maximum likelihood tree recovered with the full data set (left). On the right, linear plastid maps of the 16 Melastomataceae species. All genes are depicted as arrows (indicating transcription direction) and color coded following the legend of Fig. 1. Gray lines link the same genes on contiguous maps. LSC, long single copy region; SSC, small single copy region; IRA, inverted repeat A; IRB, inverted repeat B.

Table 2 Comparison of plastid genome size and GC content across different regions in the 16 Melastomataceae species.

Species	Coverage (mean)	LSC		SSC		IR		Full	
		bp	GC	bp	GC	bp	GC	bp	GC
<i>Allomaieta villosa</i>	278	85,915	0.347	16,975	0.306	26,781	0.425	156,452	0.369
<i>Bertolonia acuminata</i>	189	85,571	0.347	17,008	0.308	26,733	0.425	156,045	0.370
<i>Blakea schlimii</i>	170	85,370	0.349	16,998	0.308	26,747	0.425	155,862	0.370
<i>Eriocnema fulva</i>	42	85,431	0.348	16,953	0.308	26,805	0.425	155,994	0.370
<i>Graffenrieda moritziana</i>	683	85,341	0.347	16,924	0.309	26,734	0.425	155,733	0.370
<i>Henriettea barkeri</i>	130	85,991	0.347	17,036	0.306	26,750	0.425	156,527	0.369
<i>Merianthera pulchra</i>	56	85,621	0.348	17,001	0.307	26,773	0.424	156,168	0.370
<i>Miconia dodecandra</i>	318	86,609	0.348	16,999	0.310	26,804	0.425	157,216	0.370
<i>Nepsera aquatica</i>	705	84,644	0.348	17,066	0.310	26,700	0.426	155,110	0.371
<i>Opisthocentra clidemioides</i>	100	85,866	0.348	16,942	0.309	26,772	0.425	156,352	0.370
<i>Pterogastra divaricata</i>	184	84,718	0.351	17,156	0.312	26,537	0.425	154,948	0.372
<i>Rhexia virginica</i>	683	84,459	0.351	16,924	0.311	26,626	0.425	154,635	0.372
<i>Rhynchanthera bracteata</i>	304	85,093	0.347	16,729	0.307	26,643	0.426	155,108	0.370
<i>Salpinga maranoniensis</i>	537	85,128	0.353	16,653	0.317	25,765	0.428	153,311	0.374
<i>Tibouchina longifolia</i>	195	86,297	0.349	17,124	0.311	26,684	0.425	156,789	0.371
<i>Triolena amazonica</i>	48	86,200	0.347	16,970	0.307	26,741	0.425	156,652	0.369

Notes:

Length, length in bp; GC, GC content %.

LSC, long single copy region; SSC, small single copy region; IR, inverted repeat; Full, full plastome.

Table 3 Comparison of length and GC content across different sequence classes in the plastome of the 16 Melastomataceae species.

Species	Protein-coding		tRNA		rRNA		Intron		Intergenic	
	bp	GC	bp	GC	bp	GC	bp	GC	bp	GC
<i>Allomaieta villosa</i>	80,826	0.374	3,348	0.497	9,050	0.425	20,553	0.347	42,675	0.316
<i>Bertolonia acuminata</i>	80,670	0.375	3,356	0.497	9,050	0.425	20,437	0.347	42,532	0.316
<i>Blakea schlimii</i>	80,742	0.375	3,348	0.498	9,050	0.425	20,541	0.347	42,181	0.319
<i>Eriocnema fulva</i>	80,628	0.375	3,354	0.497	9,050	0.425	20,540	0.347	42,422	0.318
<i>Graffenrieda moritziana</i>	80,286	0.375	3,349	0.497	9,050	0.425	19,691	0.347	43,357	0.317
<i>Henriettea barkeri</i>	80,781	0.374	3,363	0.495	9,050	0.425	20,571	0.347	42,762	0.315
<i>Merianthera pulchra</i>	80,751	0.375	3,364	0.498	9,050	0.425	20,478	0.347	42,525	0.318
<i>Miconia dodecandra</i>	80,586	0.376	3,354	0.498	9,050	0.425	20,548	0.347	43,678	0.317
<i>Nepsera aquatica</i>	80,646	0.375	3,370	0.496	9,050	0.425	20,619	0.347	41,425	0.318
<i>Opisthocentra clidemioides</i>	80,643	0.376	3,360	0.496	9,050	0.425	20,641	0.347	42,658	0.317
<i>Pterogastra divaricata</i>	80,427	0.377	3,339	0.498	9,050	0.425	19,911	0.347	42,221	0.318
<i>Rhexia virginica</i>	80,466	0.377	3,353	0.496	9,050	0.425	20,260	0.347	41,506	0.319
<i>Rhynchanthera bracteata</i>	80,415	0.375	3,241	0.502	9,048	0.425	20,538	0.347	41,866	0.317
<i>Salpinga maranoniensis</i>	79,326	0.376	3,349	0.500	9,050	0.425	18,991	0.347	42,595	0.326
<i>Tibouchina longifolia</i>	80,682	0.377	3,348	0.497	9,050	0.425	20,666	0.347	43,043	0.317
<i>Triolena amazonica</i>	80,619	0.375	3,337	0.496	9,050	0.425	20,476	0.347	43,170	0.316

Note:

Length, length in bp; GC, GC content %.

gene content and order is similar to other Myrtales plastids, such as *Lagerstroemia fauriei* (NC029808) and *Eucalyptus polybractea* (NC022393). The exceptions are *rps16* and *rpl2*, which are putative pseudogenes in some plastids. The former seems to have been pseudogenized in *Graffenrieda moritziana* and *Pterogastra divaricata* (where the first exon is absent) and in *Salpinga margaritacea* (with several insertions changing the reading frame in the second exon); the second copy of *rpl2* gene (in the IRB) is likely a pseudogene in *Salpinga margaritacea* due to a shift in the IRB-LSC boundary in that plastid, which resulted in the loss of the second exon. Additionally, some variation is observed in all region boundaries across the Melastomataceae plastomes. The LSC-IRA boundary is located in the *rps19* gene in most species, except in *S. margaritacea* where it is located in the intron of the *rpl2* gene; the IRA-SSC boundary is located in the overlapping $\psi ycf1$ and *ndhF*; the SSC-IRB in the *ycf1*; and the IRB-LSC in the *rpl2-trnH* spacer or in the *trnH* gene. Introns are found in 17 genes in all Melastomataceae plastomes, including six tRNA genes and 11 CDS, from which three have two introns (*clpP*, *rps12* and *ycf3*). A comparison of the number of genes, regions and plastome length of one Melastomataceae (*M. dodecandra*) and eight Myrtales plastids is presented in Table 4. The sequence length of the full plastome and its regions in the Melastomataceae sampled here are in the range observed for other Myrtales.

Phylogenetic analyses

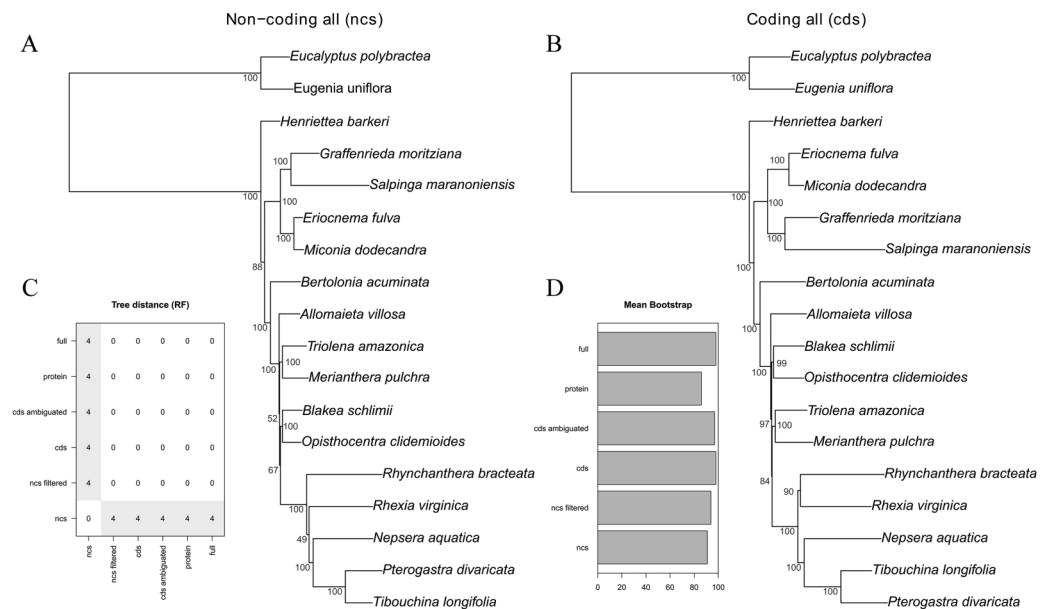
The majority of the six analytical schemes recovered the same topology (Figs. 2 and 3B). The only exception was the “all non-coding” scheme (i.e., the full non-coding regions

Table 4 Comparison of plastid genome size of one Melastomataceae species (*Miconia dodecandra*) with eight other Myrtales.

Family	Species	Coding	tRNA	rRNA	LSC	SSC	IR	Full
Melastomataceae	<i>Miconia dodecandra</i>	84	37	8	86,609	16,999	26,804	157,216
Myrtaceae	<i>Allosyncarpia ternata</i>	84	37	8	88,218	18,571	26,402	159,563
Myrtaceae	<i>Angophora costata</i>	84	37	8	88,769	18,773	26,392	160,326
Myrtaceae	<i>Corymbia gummifera</i>	84	37	8	88,310	17,197	27,603	160,713
Myrtaceae	<i>Eucalyptus polybractea</i>	84	37	8	88,944	18,530	26,397	160,268
Myrtaceae	<i>Eugenia uniflora</i>	84	37	8	87,459	18,318	26,334	158,445
Lythraceae	<i>Lagerstroemia fauriei</i>	84	37	8	83,923	16,933	25,792	152,440
Onagraceae	<i>Oenothera grandiflora</i>	84	38	8	89,862	19,035	28,824	166,545
Myrtaceae	<i>Stockwellia quadrifida</i>	84	37	8	88,247	18,544	26,385	159,561

Notes:

Protein-coding, tRNA and rRNA (number of genes); LSC, long single copy region, length in bp; SSC, small single copy region, length in bp; IR, inverted repeat, length in bp and Full (length in bp).

**Figure 3** Maximum likelihood trees of the all non-coding–ncs (A) and all coding genes–cds (B) data sets. Bootstrap support is given adjacent to the nodes. (C) Tree distance (RF) pairwise matrix between all six schemes analyzed. (D) Mean bootstrap support of all six schemes analyzed.

without filtering of dubiously aligned base pairs), where *Blakea* + *Opistocentra*, *Triolena* + *Merianthera* and *Rhynchanthera* assume a different position (Fig. 3A). Pairwise tree distances among all schemes are depicted in Fig. 3C, and all Maximum Likelihood trees with bootstrap support values are given in the Fig. S1. Bootstrap support is highest in the “full” and “cds” schemes and lower in the “protein” and “all non-coding” schemes (Fig. 3D). In the highest supported topologies, there are only two nodes with bootstrap values lower than 95, and those involve the relationship disagreements between the two alternate topologies (Figs. 3A and 3B). While filtering the non-coding poorly aligned sites improved bootstrap support and also changed the topology (“ncs” vs. “ncs filtered,”

Table 5 Summary statistics for the top 10 introns and intergenic spacers with suitable size for PCR amplification. Markers are ranked by phylogenetic information based on a weighed mean of relative values of number of variable sites (weight = 1), mean bootstrap (weight = 2) and distance to the full plastid tree (weight = 3).

Marker	Bases	Aligned (bp)	Variable sites	PIS	DNA distance (mean)	Tree distance	Bootstrap (mean)
1. <i>trnS-trnG</i> spacer	780 [628, 884]	1,125	438 (38.9%)	128 (11.4%)	0.104	4	82
2. <i>ndhF-rpl32</i> spacer	898 [849, 965]	1,266	507 (40%)	171 (13.5%)	0.114	6	71
3. <i>trnG</i> intron	762 [743, 790]	846	236 (27.9%)	76 (9%)	0.059	4	75
4. <i>ndhC-trnV</i> spacer	734 [504, 821]	991	330 (33.3%)	98 (9.9%)	0.081	4	63
5. <i>ndhA</i> intron	1,016 [939, 1,045]	1,127	250 (22.2%)	74 (6.6%)	0.046	4	64
6. <i>trnG-atpA</i> spacer	641 [550, 750]	895	353 (39.4%)	136 (15.2%)	0.114	6	65
7. <i>atpH-atpI</i> spacer	898 [638, 980]	1,178	323 (27.4%)	92 (7.8%)	0.062	8	76
8. <i>psbE-petL</i> spacer	1,058 [570, 1,165]	1,396	381 (27.3%)	132 (9.5%)	0.068	8	70
9. <i>petA-psbJ</i> spacer	736 [420, 944]	1,062	285 (26.8%)	90 (8.5%)	0.076	8	76
10. <i>trnE-trnT</i> spacer	842 [478, 1,029]	1,345	406 (30.2%)	121 (9%)	0.089	8	63

Note:

PIS, parsimony informative sites; Tree distance, RF distance.

Table 6 Primer pair sequences for the identified top five highly informative markers across the 16 plastomes of Melastomataceae.

Marker	Primer forward (5'–3')	Primer reverse (5'–3')	T _a (°C)
1. <i>trnS-trnG</i> spacer	CACTCAGCCATCTCTCCCAA	ACCCGCTACAATGCCATTATTG	55
2. <i>ndhF-rpl32</i> spacer	AGGAAAGGACCACATACGTCG	TCCTTGCTCATTTGATTTTGATCCA	55
3. <i>trnG</i> intron	GGTCCCTCGGATTTGCTTCA	GAACCCGCATCGTTAGCTTG	55
4. <i>ndhC-trnV</i> spacer	AGATGAACCTCCTAGGGAATGTGA	CCGAGAAGGTCTACGGTTTCG	55
5. <i>ndhA</i> intron	CGCTAGTCCAGAACCGTACA	ACCCCATGATTGGTTGATTAGTGA	55

Fig. 3), ambiguating common amino acids in the coding sequences did not have any apparent effect in the topology or support values (“cds” vs. “cds ambiguated;” Fig. 3D).

The commonly used plastid data set in previous family-wide studies (*rbcL*, *ndhF* and *rpl16* intron) also resulted in a different topology from the “full” scheme, although with most clades in common (Fig. S2). Disagreements involved the position of *Allomaieta*, *Trioleta* + *Merianthera*, *Blakea* + *Opisthocentra*, and *Rhynchanthera*; these disagreements manifest in nodes of low bootstrap support where, in the reduced data set, they range from 24 to 100 (mean = 73).

Phylogenetically informative regions

Summary statistics for all intron and intergenic spacers with suitable size for PCR amplification are presented in Table S1. A list of the top 10 markers ranked by phylogenetic information, taking into account topological distance to the tree based on the “full” scheme (Fig. 2), mean bootstrap support and number of variable sites is given in Table 5, and the full list is available in Table S1. All single marker phylogenies presented some disagreement to the tree based on the “full” scheme (RF tree distance ranging from 4 to 22). Bootstrap support ranged from 26 to 82 (mean = 63) and number of variable sites from 12 to 507 (mean = 224). Primer pair sequences for PCR amplification are provided for the top 5 markers in Table 6.

DISCUSSION

Plastid genomes of higher plants are of relatively small size, ranging from 115 to 165 kb in most groups, with an average of 90 CDS across most land plants (Ravi *et al.*, 2008; Wicke *et al.*, 2011). In general, the quadripartite organization, gene content and order are conserved, and GC content is usually stable within plastid regions and sequence classes (Ravi *et al.*, 2008; Wicke *et al.*, 2011). Melastomataceae plastomes are no exception for these patterns, being highly conserved and structurally similar to most other Myrtales, as well as to an ordinary angiosperm plastome. Melastomataceae plastomes' mean length (156 kb) is closer to the upper bound observed across most plants (165 kb), while the number of genes and GC content are around the average (90 genes, GC = 37%; Ravi *et al.*, 2008). High conservation in genomic structure of plastomes among the Myrtales has been previously suggested (Gu *et al.*, 2016) and is extended here to include Melastomataceae. The greatest variation in sequence length among different region classes in Melastomataceae are observed in the intergenic spacers, which is also another general pattern in plastomes (Ravi *et al.*, 2008; Gu *et al.*, 2016). Additionally, the boundaries of the IRs vary, as observed in some Myrtales and other groups (Bayly *et al.*, 2013).

Conservation in gene order, content and virtual lack of recombination make the plastome a useful tool for plant phylogenetic studies (Ravi *et al.*, 2008). An updated comprehensive phylogenetic hypothesis for the entire Melastomataceae is overdue, and full plastid sequences would contribute greatly to such an endeavor. Additionally, as sampling increases in the Myrtales, full plastids also might help to narrow down phylogenetic uncertainty in the Myrtales (e.g., Combretaceae position, Berger *et al.*, 2016). Despite the fact that the full plastome phylogeny recovered here shares most of the clades with the widely used “*rbcL* + *ndhF* + *rpl16*” tree, some changes are still observed and bootstrap support is higher. A more conclusive account on the extent of such changes will require more taxa to be sampled.

Here, we provide a list of potentially highly informative plastid markers for Melastomataceae. We acknowledge that the information descriptors employed are very sensitive to the taxa under analysis. Nonetheless, this ranked list can be used as guidance for sampling design of future studies, whereas the new family specific primers will increase the plastid options for Sanger sequencing-based phylogenies. There has been some debate as to whether the availability of full plastome sequencing (and other NGS tools) would render Sanger sequencing obsolete (Hert, Fredlake & Barron, 2008). Here we show that a full plastome phylogeny is an improvement on single or few plastid loci phylogenies, especially on the level of statistical support. However, considering scalability, computational complexity and budget limitations, a comprehensive NGS-based phylogeny for the mega-diverse Melastomataceae might not be achieved in the short term. Nonetheless, an expanded full plastome data set along with the more abundant Sanger-based sequences available, could be coupled in future studies. A hybrid NGS and Sanger sequencing approach has been employed for other groups (Xi *et al.*, 2012; Leaché *et al.*, 2014; Gardner *et al.*, 2016), and could help clarifying the backbone of a

comprehensive Melastomataceae phylogeny. Recalcitrant phylogenetic backbones are a widespread and challenging phenomenon in angiosperms (*Xi et al., 2012; Straub et al., 2014*), and their resolution is critical to increase the confidence of ancestral state reconstructions, historical biogeographical scenarios and other evolutionary hypotheses. Although full plastomes, or an expanded sample of plastid markers, may help to improve the confidence of phylogenetic relationships within the Melastomataceae, we also recognize the need of parallel sampling of additional independent genealogies (i.e., nuclear and mitochondrial genomes) for further refinement in the Melastomataceae tree.

ACKNOWLEDGEMENTS

We thank all collaborators who kindly provided tissue samples for this study. We also thank Kenneth Karol for helpful insights on assembling procedures. We thank Rapid Genomics (Gainesville, FL) for sequencing services.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This study was supported by the National Science Foundation (DEB-0818399 and DEB-1343612). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
National Science Foundation: DEB-0818399 and DEB-1343612.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Marcelo Reginato conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables.
- Kurt M. Neubig conceived and designed the experiments, performed the experiments, contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Lucas C. Majure conceived and designed the experiments, performed the experiments, contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Fabian A. Michelangeli conceived and designed the experiments, performed the experiments, contributed reagents/materials/analysis tools, reviewed drafts of the paper.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

GenBank: [KX826819](#); [KX826820](#); [KX826821](#); [KX826822](#); [KX826823](#); [KX826824](#); [KX826825](#); [KX826826](#); [KX826827](#); [KX826828](#); [KX826829](#); [KX826830](#); [KX826831](#); [KX826832](#); [KX826833](#); [KX826834](#)

Data Deposition

The following information was supplied regarding data availability:

The raw data has been supplied as [Supplemental Dataset Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.2715#supplemental-information>.

REFERENCES

- Alexander PJ, Rajanikanth G, Bacon CD, Bailey CD. 2007. Recovery of plant DNA using a reciprocating saw and silica-based columns. *Molecular Ecology Notes* 7(1):5–9 DOI 10.1111/j.1471-8286.2006.01549.x.
- Amorim AM, Goldenberg R, Michelangeli FA. 2009. A new species of *Physeterostemon* (Melastomataceae) from Bahia, Brazil, with notes on the phylogeny of the genus. *Systematic Botany* 34(2):324–329 DOI 10.1600/036364409788606389.
- Angiosperm Phylogeny Group (APG). 1998. An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* 85(4):531–553 DOI 10.2307/2992015.
- Bayly MJ, Rigault P, Spokevicius A, Ladiges PY, Ades PK, Anderson C, Bossinger G, Merchant A, Udovicic F, Woodrow IE, Tibbits J. 2013. Chloroplast genome analysis of Australian eucalypts—*Eucalyptus*, *Corymbia*, *Angophora*, *Allosyncarpia* and *Stockwellia* (Myrtaceae). *Molecular Phylogenetics and Evolution* 69(3):704–716 DOI 10.1016/j.ympev.2013.07.006.
- Bécquer-Granados ER, Neubig KM, Judd WS, Michelangeli FA, Abbott JR, Penneys DS. 2008. Preliminary molecular phylogenetic studies in *Pachyanthus* (Miconieae, Melastomataceae). *The Botanical Review* 74(1):37–52 DOI 10.1007/s12229-008-9010-z.
- Berger BA, Kriebel R, Spalink D, Sytsma KJ. 2016. Divergence times, historical biogeography, and shifts in speciation rates of Myrtales. *Molecular Phylogenetics and Evolution* 95:116–136 DOI 10.1016/j.ympev.2015.10.001.
- Clausing G, Renner SS. 2001. Molecular phylogenetics of Melastomataceae and Memecylaceae: implications for character evolution. *American Journal of Botany* 88(3):486–498 DOI 10.2307/2657114.
- Conti E, Litt A, Sytsma KJ. 1996. Circumscription of Myrtales and their relationships to other rosids: evidence from *rbcl* sequence data. *American Journal of Botany* 83(2):221–233 DOI 10.2307/2445942.
- Cox CJ, Li B, Foster PG, Embley TM, Civián P. 2014. Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. *Systematic Biology* 63(2):272–279 DOI 10.1093/sysbio/syt109.
- Fritsch PW, Almeda F, Renner SS, Martins AB, Cruz BC. 2004. Phylogeny and circumscription of the near-endemic Brazilian tribe Microlicieae (Melastomataceae). *American Journal of Botany* 91(7):1105–1114 DOI 10.3732/ajb.91.7.1105.
- Gamba-Moreno DL, Almeda F. 2014. Systematics of the Octopleura clade of *Miconia* (Melastomataceae: Miconieae) in tropical America. *Phytotaxa* 179(1):1–174 DOI 10.11646/phytotaxa.179.1.
- Gardner AG, Sessa EB, Michener P, Johnson E, Shepherd KA, Howarth DG, Jabaily RS. 2016. Utilizing next-generation sequencing to resolve the backbone of the Core Goodeniaceae and

- inform future taxonomic and floral form studies. *Molecular Phylogenetics and Evolution* **94**:605–617 DOI [10.1016/j.ympev.2015.10.003](https://doi.org/10.1016/j.ympev.2015.10.003).
- Goldenberg R, Almeda F, Sosa K, Ribeiro RC, Michelangeli FA. 2015.** *Rupestrea*: a new Brazilian genus of Melastomataceae, with anomalous seeds and dry indehiscent fruits. *Systematic Botany* **40**(2):561–571 DOI [10.1600/036364415X688862](https://doi.org/10.1600/036364415X688862).
- Goldenberg R, de Fraga CN, Fontana AP, Nicolas AN, Michelangeli FA. 2012.** Taxonomy and phylogeny of *Merianthera* (Melastomataceae). *Taxon* **61**(5):1040–1056.
- Goldenberg R, Penneys DS, Almeda F, Judd WS, Michelangeli FA. 2008.** Phylogeny of *Miconia* (Melastomataceae): patterns of stamen diversification in a megadiverse neotropical genus. *International Journal of Plant Sciences* **169**(7):963–979 DOI [10.1086/589697](https://doi.org/10.1086/589697).
- Gu C, Tembrock LR, Johnson NG, Simmons MP, Wu Z. 2016.** The complete plastid genome of *Lagerstroemia fauriei* and loss of *rpl2* intron from *Lagerstroemia* (Lythraceae). *PLoS ONE* **11**(3):e0150752 DOI [10.1371/journal.pone.0150752](https://doi.org/10.1371/journal.pone.0150752).
- Guy L, Kultima JR, Andersson SGE. 2010.** genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26**(18):2334–2335 DOI [10.1093/bioinformatics/btq413](https://doi.org/10.1093/bioinformatics/btq413).
- Hert DG, Fredlake CP, Barron AE. 2008.** Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis* **29**(23):4618–4626 DOI [10.1002/elps.200800456](https://doi.org/10.1002/elps.200800456).
- Katoh K, Standley DM. 2013.** MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**(4):772–780 DOI [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).
- Kriebel R. 2016.** Phylogenetic placement of the monotypic genus *Schwackaea* (Melastomeae: Melastomataceae) and the evolution of its unique fruit. *International Journal of Plant Sciences* **177**(5):440–448 DOI [10.1086/685689](https://doi.org/10.1086/685689).
- Kriebel R, Michelangeli FA, Kelly LM. 2015.** Discovery of unusual anatomical and continuous characters in the evolutionary history of *Conostegia* (Miconieae: Melastomataceae). *Molecular Phylogenetics and Evolution* **82**(Part A):289–313 DOI [10.1016/j.ympev.2014.09.021](https://doi.org/10.1016/j.ympev.2014.09.021).
- Leaché AD, Wagner P, Linkem CW, Böhme W, Papenfuss TJ, Chong RA, Lavin BR, Bauer AM, Nielsen SV, Greenbaum E, Rödel M-O, Schmitz A, LeBreton M, Ineich I, Chirio L, Ofori-Boateng C, Eniang EA, El Din SB, Lemmon AR, Burbrink FT. 2014.** A hybrid phylogenetic–phylogenomic approach for species tree estimation in African *Agama* lizards with applications to biogeography, character evolution, and diversification. *Molecular Phylogenetics and Evolution* **79**:215–230 DOI [10.1016/j.ympev.2014.06.013](https://doi.org/10.1016/j.ympev.2014.06.013).
- Lohse M, Drechsel O, Kahlau S, Bock R. 2013.** OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research* **41**(W1):W575–W581 DOI [10.1093/nar/gkt289](https://doi.org/10.1093/nar/gkt289).
- Majure LC, Neubig KM, Skean JD, Bécquer ER, Judd WS. 2015.** Evolution of the Sandpaper clade (Miconieae, Melastomataceae). *International Journal of Plant Sciences* **176**(7):607–626 DOI [10.1086/682148](https://doi.org/10.1086/682148).
- Martin CV, Little DP, Goldenberg R, Michelangeli FA. 2008.** A phylogenetic evaluation of *Leandra* (Miconieae, Melastomataceae): a polyphyletic genus where the seeds tell the story, not the petals. *Cladistics* **24**(3):315–327 DOI [10.1111/j.1096-0031.2007.00185.x](https://doi.org/10.1111/j.1096-0031.2007.00185.x).
- Martin PG, Dowd JM. 1986.** Phylogenetic studies using protein sequences within the order Myrtales. *Annals of the Missouri Botanical Garden* **73**(2):442–448 DOI [10.2307/2399122](https://doi.org/10.2307/2399122).
- Michelangeli FA, Guimaraes PJE, Penneys DS, Almeda F, Kriebel R. 2013.** Phylogenetic relationships and distribution of new world Melastomeae (Melastomataceae). *Botanical Journal of the Linnean Society* **171**(1):38–60 DOI [10.1111/j.1095-8339.2012.01295.x](https://doi.org/10.1111/j.1095-8339.2012.01295.x).

- Michelangeli FA, Judd WS, Penneys DS, Skean JD, Bécquer-Granados ER, Goldenberg R, Martin CV. 2008. Multiple events of dispersal and radiation of the tribe Miconieae (Melastomataceae) in the Caribbean. *The Botanical Review* 74(1):53–77
DOI 10.1007/s12229-008-9004-x.
- Michelangeli FA, Nicolas A, Morales-Puentes ME, David H. 2011. Phylogenetic relationships of *Allomaieta*, *Alloneuron*, *Cyphostyla*, and *Wurdastom* (Melastomataceae) and the resurrection of the tribe Cyphostyleae. *International Journal of Plant Sciences* 172(9):1165–1178
DOI 10.1086/662032.
- Michelangeli FA, Penneys DS, Giza J, Soltis D, Hils MH, Skean JD. 2004. A preliminary phylogeny of the tribe Miconieae (Melastomataceae) based on nrITS sequence data and its implications on inflorescence position. *Taxon* 53(2):279–290 DOI 10.2307/4135608.
- Michelangeli FA, Ulloa CU, Sosa K. 2014. *Quipuanthus*, a new genus of Melastomataceae from the foothills of the Andes in Ecuador and Peru. *Systematic Botany* 39(2):533–540
DOI 10.1600/036364414X680924.
- Misof B, Misof K. 2009. A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Systematic Biology* 58(1):21–34
DOI 10.1093/sysbio/syp006.
- Neubig KM, Whitten WM, Abbott JR, Elliott S, Soltis DE, Soltis PS. 2014. Variables affecting DNA preservation in archival DNA specimens. In: Applequist WL, Campbell LM, eds. *DNA Banking in the 21st Century: Proceedings of the U.S. Workshop on DNA Banking*. St. Louis: The William L. Brown Center at the Missouri Botanical Garden.
- Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20(2):289–290 DOI 10.1093/bioinformatics/btg412.
- Penneys DS. 2013. Preliminary phylogeny of the Astronieae (Melastomataceae) based on nuclear and plastid DNA sequence data, with comments on the Philippine endemic genus, *Astrocalyx*. *Philippine Journal of Science* 142:159–168.
- Penneys DS, Michelangeli FA, Judd WS, Almeda F. 2010. Henrietteae (Melastomataceae): a new neotropical berry-fruited tribe. *Systematic Botany* 35(4):783–800
DOI 10.1600/036364410X539862.
- R Development Core Team. 2016. *R: a language and environment for statistical computing*. Vienna: The R Foundation for Statistical Computing. Available at <http://www.R-project.org/>.
- Ravi V, Khurana JP, Tyagi AK, Khurana P. 2008. An update on chloroplast genomes. *Plant Systematics and Evolution* 271(1–2):101–122 DOI 10.1007/s00606-007-0608-0.
- Reginato M, Michelangeli FA. 2016. Untangling the phylogeny of *Leandra* s. str. (Melastomataceae, Miconieae). *Molecular Phylogenetics and Evolution* 96:17–32
DOI 10.1016/j.ympev.2015.11.015.
- Reginato M, Michelangeli FA, Goldenberg R. 2010. Phylogeny of *Pleiochiton* (Melastomataceae, Miconieae): total evidence. *Botanical Journal of the Linnean Society* 162(3):423–434
DOI 10.1111/j.1095-8339.2009.01022.x.
- Renner SS. 2004. Bayesian analysis of combined chloroplast loci, using multiple calibrations, supports the recent arrival of Melastomataceae in Africa and Madagascar. *American Journal of Botany* 91(9):1427–1435 DOI 10.3732/ajb.91.9.1427.
- Robinson DF, Foulds LR. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* 53(1–2):131–147 DOI 10.1016/0025-5564(81)90043-2.
- Rocha MJR, Guimarães PJE, Michelangeli FA, Romero R. 2016. Phylogenetic placement and a new circumscription of *Poteranthera* (Microlicieae; Melastomataceae). *Phytotaxa* 263(3):219–232 DOI 10.11646/phytotaxa.263.3.

- Rozen S, Skaletsky H. 2000.** Primer3 on the WWW for general users and for biologist programmers. In: Misener S, Krawetz SA, eds. In: *Bioinformatics Methods and Protocols*. Vol. 132. Totowa: Humana Press, 365–386.
- Schliep KP. 2011.** phangorn: phylogenetic analysis in R. *Bioinformatics* **27(4)**:592–593 DOI [10.1093/bioinformatics/btq706](https://doi.org/10.1093/bioinformatics/btq706).
- Stamatakis A. 2014.** RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30(9)**:1312–1313 DOI [10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033).
- Stone RD. 2006.** Phylogeny of major lineages in Melastomataceae, subfamily Olisbeoideae: utility of nuclear glyceraldehyde 3-phosphate dehydrogenase (*GapC*) gene sequences. *Systematic Botany* **31(1)**:107–121 DOI [10.1600/036364406775971741](https://doi.org/10.1600/036364406775971741).
- Straub SCK, Moore MJ, Soltis PS, Soltis DE, Liston A, Livshultz T. 2014.** Phylogenetic signal detection from an ancient rapid radiation: Effects of noise reduction, long-branch attraction, and model selection in crown clade Apocynaceae. *Molecular Phylogenetics and Evolution* **80**:169–185 DOI [10.1016/j.ympev.2014.07.020](https://doi.org/10.1016/j.ympev.2014.07.020).
- Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. 2011.** The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* **76(3–5)**:273–297 DOI [10.1007/s11103-011-9762-4](https://doi.org/10.1007/s11103-011-9762-4).
- Xi Z, Ruhfel BR, Schaefer H, Amorim AM, Sugumaran M, Wurdack KJ, Endress PK, Matthews ML, Stevens PF, Mathews S, Davis CC. 2012.** Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proceedings of the National Academy of Sciences of the United States of America* **109(43)**:17519–17524 DOI [10.1073/pnas.1205818109](https://doi.org/10.1073/pnas.1205818109).
- Zeng S-J, Zou L-H, Wang P, Hong W-J, Zhang G-Q, Chen L-J, Zhuang X-Y. 2016.** Preliminary phylogeny of *Fordiophyton* (Melastomataceae), with the description of two new species. *Phytotaxa* **247(1)**:45–61 DOI [10.11646/phytotaxa.247.1](https://doi.org/10.11646/phytotaxa.247.1).