J Biosci (2020)45:96
DOI: 10.1007/s12038-020-00061-2

Re-evaluation of the phylogenetic relationships and species delimitation of two closely related families (Lamiaceae and Verbenaceae) using two DNA barcode markers

O O Oyebanji[1]*, E C Chukwuma[2], K A Bolarinwa[3], O I Adejobi[4], S B Adeyemi[5] and A O Ayoola[6]

[1]*Department of Botany, University of Lagos, Akoka, Yaba, Lagos, Nigeria*

[2]*Forest Herbarium Ibadan (FHI), Forestry Research Institute of Nigeria, Ibadan, Nigeria*

[3]*Department of Education Science (Biology Unit), Distance Learning Institute, University of Lagos, Akoka, Lagos, Nigeria*

[4]*Landmark University, Omu-Aran, Kwara State, Nigeria*

[5]*Ethnobotany Unit, Department of Plant Biology, Faculty of Life Sciences, University of Ilorin, Ilorin, Nigeria*

[6]*Department of Ecotourism and Wildlife Management, Federal University of Technology, Akure, Ondo State, Nigeria*

**Corresponding author (Email, oyebanjioyetola@gmail.com)*

MS received 21 September 2019; accepted 27 May 2020

The families Lamiaceae and Verbenaceae comprise several closely related species that possess high morphological synapomorphic traits. Hence, there is a tendency of species misidentification using only the morphological characters. Herein, we evaluated the discriminatory power of the universal DNA barcodes (*matK* and *rbcL*) for 53 species spanning the two families. Using these markers, we inferred phylogenetic relationships and conducted species delimitation analysis using four delimitation methods: Automated Barcode Gap Discovery (ABGD), TaxonDNA, Bayesian Poisson Tree Processes (bPTP) and General Mixed Yule Coalescent (GMYC). The phylogenetic reconstruction based on the *matK* gene resolved the relationships between the families and further suggested the expansion of the Lamiaceae to include some core Verbanaceae genus, e.g., *Gmelina*. The *rbcL* marker using the TaxonDNA method displayed high species delimitation resolutions, while the ABGD, GMYC, and bPTP generated different number of Operational Taxonomic Units/genetic clusters. Our results underscored the efficiency of the *matK* and *rbcL* genes as reliable markers for resolving phylogenetic relationships and species delimitation of both families, respectively. The current study provides insights into the DNA barcode applications in these families, at the same time contributing to the current understanding of genetic divergence patterns in angiosperms.

Keywords. Lamiaceae; phylogenetic relationship; plastid marker; species identification; Verbenaceae

1. Introduction

Taxonomy of the families Lamiaceae Martinov and Verbenaceae J.St.-Hil *sensu lato* as core members of the order Lamiales have long been known (Cronquist 1981; Dahlgren 1983; Thorne 1983, APG 2009). Previous taxonomic studies revealed that the family Lamiaceae evolved from the Verbenaceae, indicating the possibility of close morphology affinities between the families (Abu-Asab and Cantino 1992; Cantino 1992; Scheen *et al*. 2010). Morphological characteristics such as opposite leaves, zygomorphic flowers, and

Electronic supplementary material: The online version of this article (https://doi.org/10.1007/s12038-020-00061-2) contains supplementary material, which is available to authorized users.

Published online: 09 July 2020

bicarpellate gynoecium, which develop into four uniovulate locules through the formation of false partitions are synapomorphic while the ovary structure is autapomorphic between these families (Ryding 1995; Wagstaff and Olmstead 1997). In the past, these morphological characters were employed as delimiting traits for taxa (Marx *et al.* 2010). Nevertheless, morphological-based species identification is complex between these families, hence resulted in ambiguous species discrimination and incorrect phylogenetic inferences.

Previous phylogenetic studies present unstable phylogenetic relationships between the Lamiaceae and Verbenaceae (Wagstaff and Olmstead 1997; Olmstead *et al.* 2000, 2001, Sanders 2001; Atkins 2004; Oxelman *et al*. 2005). For instance, the placement of genera (e.g., *Clerodendrum* L., *Gmelina* L., *Lamium* L., *Premna* L., *Tectona* L. f., *Verbena* L., and *Vitex* L.) are still controversial (Wagstaff *et al*. 1998; Bremer *et al*. 2002; GRIN 2004; Chen *et al*. 2014; Li *et al*. 2016). Both the plastid (coding and non-coding regions, single and multiple loci) and nuclear (ITS) markers have been utilized to resolve the evolutionary relationships between these families (Chen *et al*. 2014; Li *et al*. 2016). The unresolved relationship may be fundamental, i.e., linked with indeterminate taxonomic identification using morphological-based assessment. To address this, a molecular-based identification method for the species of Lamiaceae and Verbenaceae is imperative. It is fascinating that efforts have been channeled on the selection of accurate DNA barcodes for plant species identification (Fazekas *et al*. 2008; Liu *et al*. 2011).

The advent of DNA barcoding improves rapid species identification and enables the resolution of enigmatic phylogenetic relationships (Fazekas *et al*. 2008; Kress and Erickson 2007). Molecular markers are less prone to environmental influences; thus, they have a vast application (Lahaye *et al.* 2008; Surya *et al.* 2014). Successful implementation of plastid (*matK*, *rbcL*, *rpoB*, *rpoC1*, *atpF–atpH*, *trnH-psbA*, and *psbK–psbI*) and nuclear ribosomal (internal transcribed spacers [ITS]) DNA markers in species identification at different taxonomic ranks have been reported in previous studies (Liu *et al*. 2011; Kress and Erickson 2007; Yan *et al*. 2015; Baldwin *et al*. 1995; Pennisi 2007; Kress *et al*. 2010). Moreover, the *matK* and *rbcL* markers have been accepted as universal DNA barcode for plants (CBOL 2009).

DNA barcode approach, a proposed method of molecular taxonomy or systematics, identifies species based on the fragments of DNA nucleotide sequences (Xiang *et al*. 2011; Li *et* al. 2012; Chen *et al*. 2015). Many automatic, fast, and user-friendly approaches for species delimitation process have been recommended with zenith outcomes (Chen *et al*. 2015; Saddhe *et al*. 2017). The most common methods are Automatic Barcode Gap Discovery (ABGD, Tang *et al*. 2014), Bayesian Poisson Tree Processes (bPTP, Dumas *et al*. 2015), Generalized Mixed Yule Coalescent (GMYC, Sahu *et al*. 2016), and TaxonDNA (Meier *et al*. 2006). Studies have supported the taxonomic accuracy, resolution efficiency, and rapidity of these methods (Rozas 2009; Zou *et al.* 2016). Expectedly, the DNA barcode through barcode gap estimation and more priors will perform precise species identification and deliver a novel perspective to the species diversity assessment of the families Lamiaceae and Verbenaceae.

Previous studies have only unraveled the phylogenetic relationships between the families Lamiaceae and Verbenaceae. Best to our knowledge, no comprehensive molecular-based identification studies have been conducted for both families. Nevertheless, studies have shown individual genes, e.g. *matK* and *rbcL*, to be useful in resolving long-standing phylogenetic relationships among angiosperms (Doyle *et al*. 1997; Lavin *et al*. 2005; Selvaraj *et al*. 2008). In the present study, we retrieved 53 sequences, each for two plastid markers (*matK* and *rbcL*) spanning the Lamiaceae and Verbenaceae from the GenBank. We inferred the phylogenetic relationships and evaluated the species delimitations resolution power of the individual gene loci. Also, this study will provide an adequate knowledge for non-taxonomists, researchers, and biodiversity managers on the utilization of molecular methods for accurate identification.

2. Materials and methods

2.1 *Data retrieval and sequence alignment*

Our main objective was to appraise the species delimitation potentials and phylogenetic relationships signals of the *matK* and *rbcL* genes between the two families. We explored the available *matK* and *rbcL* sequences in the NCBI GenBank database (*http://www.ncbi.nlm.nih.gov/genebank*) and downloaded 300 and 354 sequences, respectively (data not shown). All the downloaded sequences were viewed, trimmed, and aligned in MEGA version 7 (Kumar *et al*. 2016). The hypervariable sites were excluded from the data matrices using the Gblocks version 0.91b (Castresana 2000; Talavera and Castresana, 2007) with the default

parameters. Afterward, we obtained the final lengths for the two chloroplast DNA genes (*matK* = 374 bp and *rbcL* = 154 bp). Following this, we constructed an initial phylogenetic tree using the aligned sequences for both genes. To avoid overrepresentation of species and evolutionary tree instability, we performed systematic taxon reduction by selecting representative species in a lineage from the initial tree constructed (data not shown). However, preference was given to species with unclear phylogenetic position (*Clerodendrum*, *Gmelina*, *Premna*, *Tectona*, and *Vitex*, previously known as core Verbenaceae (Hutchinson *et al*. 1963). This study included a total of 56 species (53 ingroups and three outgroups; supplementary file 1). We used three outgroups (two from the family Asteraceae and one from the family Anacardiaceae). These selected sequences (in group - 53 species) were used for phylogenetic and species delimitation analysis.

2.2 *Sequence characteristics*

At first, we performed sequence analysis to determine the nucleotide frequencies for each locus using MEGA7 (Kumar *et al*. 2016) while other information such as parsimony informative characters (PIC), consistency index (CI), retention index (RI), rescaled consistency (RC) and tree length were performed using PAUP 4.0b10 (Swofford 2002). Further, we conducted substitution homogeneity test patterns for individual gene marker using the Disparity Index Test (DIT) function incorporated in MEGA7 (Kumar *et al*. 2016; Kumar and Gadagkar 2001). The DIT presents species pairs with the probability that sequences have evolved with the same substitution pattern based on the differences in base composition biases between sequences. The analysis used a Monte Carlo test (500 replicates) to estimate the *P*-values, and the values < 0.05 were considered significant (supplementary file 2).

2.3 *Phylogenetic analysis*

Phylogenetic analysis was performed using a maximum likelihood (ML) and Bayesian Inference (BI) based on individual genes (Supplementary Figure 1). The ML trees were inferred in RAxML version 7.2.8 (Stamatakis 2006), using a gamma distribution evolutionary model. Branch supports were estimated through 1000 bootstrap (BS) replications. The Bayesian analyses were estimated in MrBayes version 3.2.1 (Ronquist *et al*. 2012). Best-fit nucleotide substitution model

Table 1. Evolutionary models

Tree search	Region	Model (-ln*L*)	AIC
BI	*matK*	GTR+G (2582.69)	5183.39
	rbcL	TVMef+I+G (610.37)	1232.76
ML	*matK*	GTRGAMMA (− 2577.51)	5183.39
	rbcL	GTRGAMMA (− 609.90)	1232.76

AIC Akaike information criterion

table 1 for each dataset were specified using the Akaike information criterion (AIC) implemented in MrModelTest version 3.7 (Posada and Buckley 2004). Two independent determinations of the Markov Chain Monte Carlo (MCMC) were performed for 10,000,000 generations initiated from a random tree, and the trees were saved each at 10,000th generation. Bayesian major rule consensus and posterior probability (PP) scored after the first 25% burn-in was discarded. The trees generated were viewed and edited in FigTree version 1.4.2 (Rambaut 2014). All phylogenetic analyses were conducted in the CIPRES Science Gateway version 3.3 web service (*http://www.phylo.org*; Miller *et al.* 2010). The bootstrap support values were categorized accordingly (i.e. 50–74% = weak support, 75–84% = moderate support, and 85–100% = strong support).

2.4 *Species delimitation analysis*

The species delimitation analysis was carried out using four methods, and the resulting clusters/matches were referred to as operational taxonomic units (OTUs) in this study (table 2; supplementary files 3–6).

Method 1: Automated Barcode Gap Discovery (ABGD)

The individual gene data matrices were uploaded into the ABGD webserver (*http://www.abi.snv.jussieu.fr/public/abgd/abgdweb.html*; Puillandre *et al*. 2011). The analysis was conducted based on distance method (Jukes-Cantor {JC69}, Kimura {K80}, and Simple distance), partitioning of the sequences into potential species according to the barcode gap (X = 1.0 and 1.5), range of Prior Intraspecific divergence (P-min and P-max), Nb bins (20), while other parameters were in default settings. We performed strict priors to optimize our results (table 2a; supplementary file 3).

Table 2. Summary of Species delimitation: (a) ABGD, (b) bPTP, (c) TaxonDNA and (d) GMYC

(*a*)

				Jukes-Cantor (JC69)		Kimura		Simple Distane	
Region	RGW(X)	Partitions	Prior intraspecific divergence	IP	RP	IP	RP	IP	RP
matK	1.0	1–9	0.001–0.059	1 and 8	17–20	34	36	1 and 18	4
	1.5	1–9	0.001–0.059	1	–	1	–	1	–
rbcL	1.0	1–8	0.001–0.035	1 and 4	5–6	1–20	22	1	–
	1.5	1–8	0.001–0.035	–	–	–	–	1	–

(*b*)

Region	OTUs	A. Rate	Merge	Split	Ens	Mean	Support values
matK	25	0.72169	100222	99778	17 and 45	31.46	0.262–1.00
rbcL	46	0.86491	49845	50155	13 and 50	34.82	0.205–1.00

(*c*)

Threshold (%)	OTU
1	25/53
3	13/53
4	11/53

(*d*)

Methods	Region	Likelihood of null model	Maximum likelihood of GMYC model	Likelihood ratio:	LR test	No. of ML clusters (CI)	No. of ML entities (CI)	Threshold time
sGMYC	*matK*	331.4405	333.8913	4.901479	0.08622978	13 (3–15)	17 (3–33)	– 0.01300022
	rbcL	348.367	349.0871	1.440099	0.6765274	5 (1–18)	5 (1–52)	– 0.0253254
mGMYC	*matK*	331.4405	335.733	8.584973	0.01367089*	11 (8–14)	27 (8–32)	– 0.01240913 – 0.006682476 – 0.002590421
	rbcL	348.367	350.2918	3.849566	0.1459074	14 (1–16)	20 (1–30)	– 0.01480931 – 0.009475366 – 0.003676032

RGW relative gap width; *IP* initial partition; *RP* Recursive partition; *A. rate* Acceptance rate, *Ens* estimated number of species, *matK/rbcL* (for table 2c)

Method 2: Bayesian Posterior Tree Poisson (bPTP)

The Bayesian implementation of the PTP model for species delimitation utilizes the number of substitutions inferred from branch lengths of the ML tree (Zhang *et al*. 2013). The bPTP analysis was performed using the webserver (*http://species.h-its.org*; Saddhe *et al*. 2017). This web interface only allows a single phylogenetic tree as input at a time. We uploaded the rooted ML tree (Newick format) generated via CIPRES portal to conduct the bPTP delimitation analysis using the following parameters: No. of MCMC generations = 200,000 (> 50 taxa), thinning = 100, burn-in 25%, seed = default and specified the outgroups to improve the delimitation results. The result adds Bayesian support values (0 ≤ 1) to species delimitation (table 2b; supplementary file 3).

Method 3: TaxonDNA

The species cluster analysis for each marker was conducted at different threshold values (1%, 3%, and 4%) in TaxonDNA version 1.8 (Meier *et al*. 2006). The analysis provides cluster/OTUs for query sequences (53 species, excluding the outgroups) based similarities and generates specific detail on every cluster, which corresponds to species similarities (table 2c; supplementary file 3).

Method 4: General Mixed Yule Coalescent (GMYC)

We used the MrBayes nexus file to generate the ultrametric tree required for the GMYC analysis. First, we set the initial priors for each of the markers in BEAUti version 1.8 (Drummond and Rambaut 2007; Drummond *et al*. 2012). This included the models from the MrModeltest (table 1), tree prior (Yule process), 10, 000000 generations, sampling at every 1000 step, and other parameters were in default settings. The Yule tree is commonly used to explain the net speciation rate between different species (Nee 2006; Gernhard 2008). After that, we generated the input files (XML) for further analysis in BEAST version 1.8.4 (Drummond and Rambaut 2007). We evaluated the accuracy of the BEAST priors based on the Effective Sample Size (ESS > 200) in Tracer version 1.5 (Rambaut and Drummond 2009). We set the first 25 % of the samples as burn-in while the remaining trees were combined to generate the maximum clade credibility tree in TreeAnnotator v1.6.1 (Rambaut and Drummond 2010). For the GMYC analysis (table 2d; supplementary file 6), we conducted both the single (sGMYS) and multiple (mGMYS) thresholds in R platform (R Core Team, 2012) using the APE (Paradis *et al*. 2004) and SPLITS (Ezard *et al*. 2009) packages. We summarized the support values into three categories: weak (≤ 0.49), moderate (0.5–0.79), and strong (> 0.8).

3. Results

3.1 *Sequence information*

The aligned length of the *matK* and *rbcL* regions were 378 bp and 154 bp respectively after removing the variable sites in all the sequences. The sequence analysis revealed variations in the nucleotide frequencies for each locus; *matK* (A = 27.97%, T = 34.90%, C = 18.40%, and G = 18.73%) and *rbcL* (A = 26.33%, T = 29.10%, C = 22.09%, and G = 22.47%). Our analysis inferred that *matK* (PIC = 180, CI = 0.721; RI = 0.894, RC = 0.644 and tree length = 380) is highly informative than *rbcL* (PIC = 33; CI = 0.611, RI = 0.880, RC = 0.538 and tree length = 72). Evolutionary models for each gene are presented in table 1. The homogeneity test revealed different substitution patterns between the pair of sequences for the individual gene loci. The test discovered a total of 1379 possible comparisons among the 53 species. There were 117 significant ($P < 0.05$) species sequence comparisons and the number of significant comparisons for the *matK* and *rbcL* were 65 and 52 respectively (supplementary file 2).

3.2 *Phylogenetic relationships of the Lamiaceae and Verbenaceae*

The phylogenetic trees (ML and BI) constructed for the individual plastid DNA marker- *matK* and *rbcL* yielded conflicting topologies with varying support values (figure 1; supplementary figure 1). For the *matK*, the ML (BS = 25–100 %) and BI = (PP = 0.53–1.00) tree presents a fully resolved relationships to two large clades of the Verbenaceae and Lamiaceae. It supported the placement of traditional Verbenaceae genera (*Clerodendrum*, *Gmelina*, *Premna*, *Tectona*, and *Vitex*) within the Lamiaceae. The *rbcL* gene presents dissimilar topologies (BS = 0–100% / PP = 0.63–1.0), and failed to resolve the phylogenetic relationships (non-monophyletic) at both the generic (e.g. *Ocimum* L., *Scutellaria* L., and *Vitex*; supplementary figure 1) and family levels. However, the analyses ascertained the phylogenetic relationship between the traditional Verbenaceae and the core Lamiaceae. We could not clearly define the circumstance of the genera *Orthosiphon* Benth. and *Tectona* because only one species was included in this study.

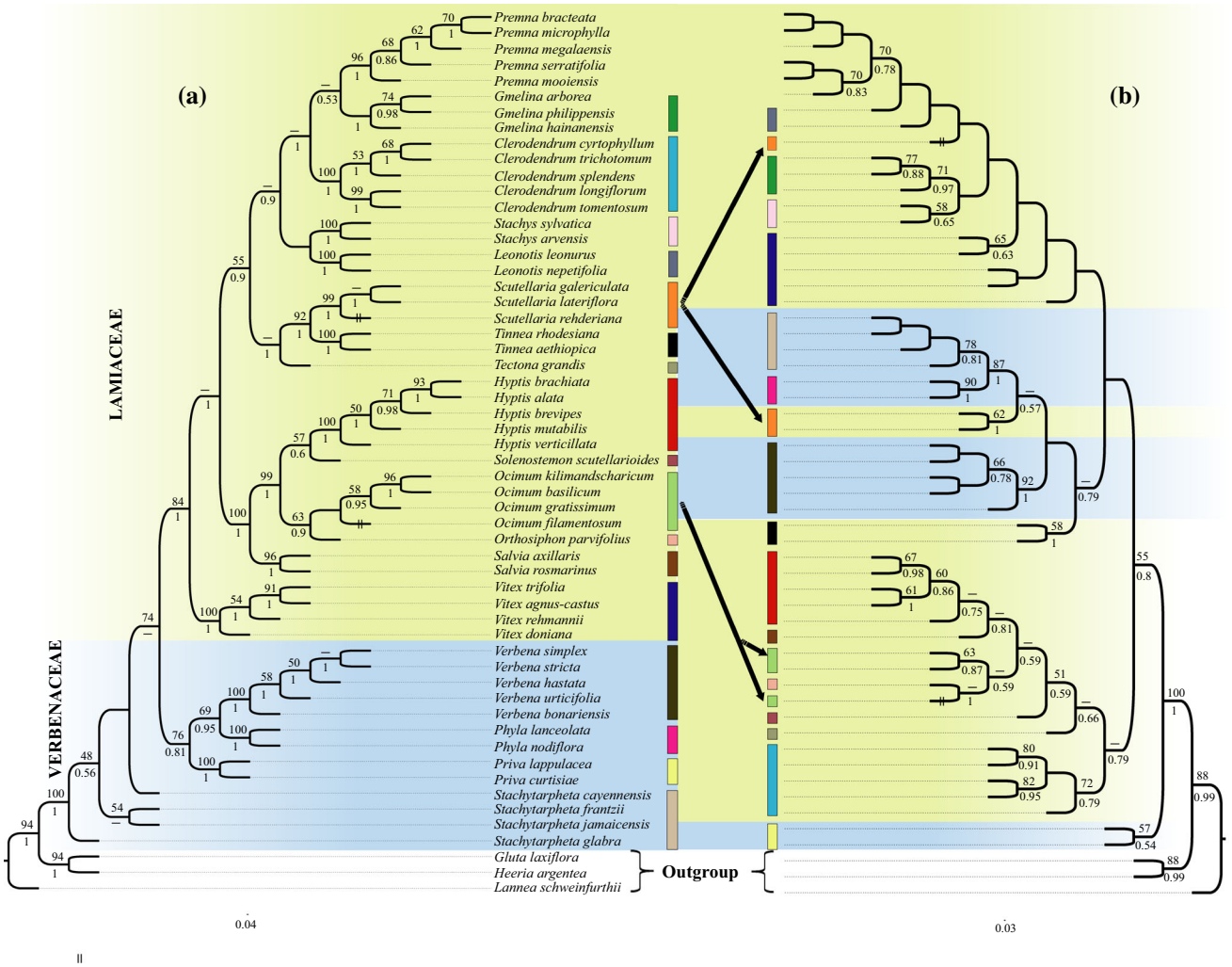

Figure 1. Maximum likelihood and Bayesian phylogenetic tree inferred from the single gene locus; (**a**) *matK* and (**b**) *rbcL*. The families Lamiaceae and Verbenaceae are indicated with different colors. The support values are presented on the branches (BS = above the branch and PP = below the branch). Only support values ≥ 50 % and ≥ 0.5 are displayed for the ML and BI trees respectively (supplementary figure 1). Colored boxes represent the genus while the dark arrows showed the topology differences. The double strokes on the tree branch indicate same species with different phylogenetic placement on the *matK* and *rbcL* tree.

3.3 *Species identification and delimitation*

The species discriminatory analysis generated varying OTUs according to the markers and the delimitation methods. Our analysis using the *rbcL* marker as implemented in the TaxonDNA (threshold 1%, 3%, and 4 %) yielded the highest number of OTUs. However, the resulting number OTUs differs among the methods. The detailed results are further presented (table 2a–d; figure 2; supplementary files 3–6).

3.3.1 *ABGD species delimitation:* The ABGD analysis revealed marginal differences in the partitions irrespective of the barcode gaps for each marker (*matK* = 1–9 and *rbcL* = 1–8). Highest OTUs were produced when X = 1.0, and the *matK* generated higher OTUs in contrast to *rbcL*. The K80 Kimura method had the highest OTU for both *matK* (initial = 34; recursive = 36) and *rbcL* (initial = 1–20; recursive = 22). The number of OTUs dropped when X = 1.5 (initial = 1) with no recursive partitions for most of the methods (table 2a; figure 2; supplementary file 3). The K80 Kimura method presents deeper discrimination power in the ABGD analysis.

3.3.2 *bPTP species delimitation:* The bPTPanalysis was returned with a high acceptance rate, merge, split, the estimated number of species, and mean values for

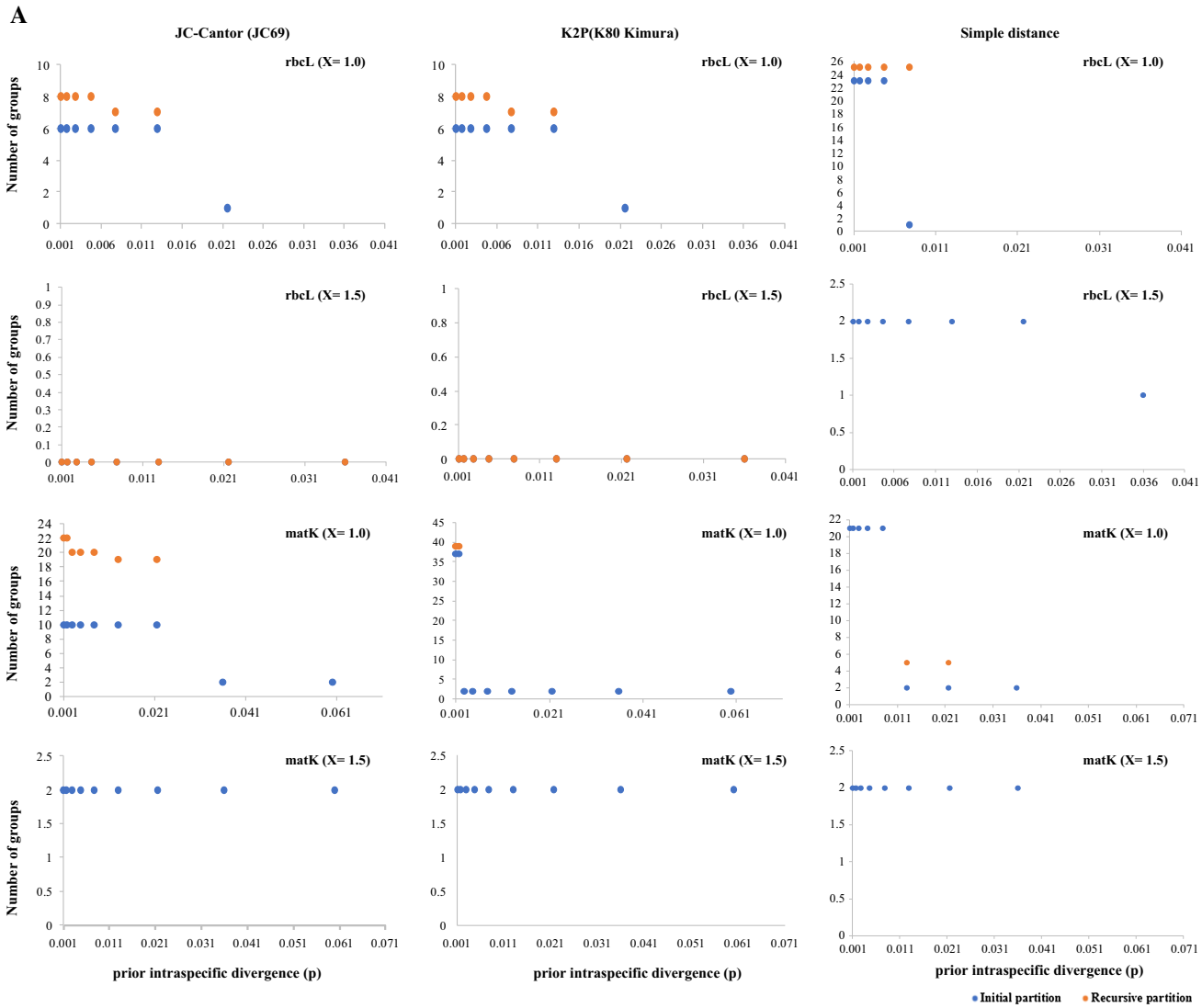

Figure 2. (**A**) Automatic partition produced by ABGD with three metrics (JC69, K80 and simple distance) and two X-values (X = 1, 1.5) for the *rbcL* and *matK*. Details of the partitions are shown in supplementary figure 3. (**B**) ABGD species delimitation analysis (b1) *matK* and (b2) *rbcL*. Species arrangement is similar to figure 1. Colors represent different OTUs: Green (genus > 1 which were Merged), Yellow (genus = 1, species > 1 which were Merged) and red (genus = 1, species = 1 which were Split). Barcode gap (1.0 and 1.5). The number in boxes corresponds to the same OTUs (see details in supplementary file 3).

rbcL compared to *matK* (table 2b; figure 2; supplementary file 4). The bPTP analysis presents the BS between 0.21 and 1.0, and 0.262 and 1.00 for *rbcL* and *matK,* respectively. Higher BS value on a node indicates that all descendants from this node are more likely to be from one species.

3.3.3 *TaxonDNA species delimitation:* The number of OTUs varies with the threshold values (1%, 3%, and 4 %), an indication of different evolutionary rates between the two barcode genes (table 2c; figure 2; supplementary file 4). The results remained constant for *rbcL* (53 OTUs), which resolved all the species identities (genus and species level) while the OTUs (11–25) decreases with increasing threshold values for the *matK* gene.

3.3.4 *GMYC species delimitation:* The sGMYC model generated a higher number of ML clusters and entities for the *matK* (13 and 17) than the *rbcL* (5 each for both). These were produced at different confident intervals and threshold time (table 2d; figure 2; supplementary file 6). The mGMYC produced three threshold times resulting in 11 ML clusters and 27 ML entities for the *matK,* while 14 ML clusters and 20 ML entities were generated for the *rbcL* (supplementary

B

(a)

Species	JC69 Jukes-Cantor				K80 Kimura				Simple Distance			
	Initial		Recursive		Initial		Recursive		Initial		Recursive	
	1.0	1.5	1.0	1.5	1.0	1.5	1.0	1.5	1.0	1.5	1.0	1.5
Premna bracteata	1	1	9	-	3	1	3	-	2	1	1	-
Premna microphylla	1	1	9	-	3	1	3	-	2	1	1	-
Premna menglaensis	1	1	9	-	3	1	3	-	2	1	1	-
Premna serratifolia	1	1	9	-	5	1	5	-	2	1	1	-
Premna mooiensis	1	1	9	-	4	1	4	-	2	1	1	-
Gmelina arborea	1	1	10	-	6	1	6	-	3	1	1	-
Gmelina philippensis	1	1	10	-	6	1	6	-	3	1	1	-
Gmelina hainanensis	1	1	10	-	6	1	6	-	3	1	1	-
Clerodendrum cyrtophyllum	2	1	2	-	12	1	12	-	9	1	1	-
Clerodendrum trichotomum	2	1	2	-	12	1	12	-	9	1	1	-
Clerodendrum splendens	2	1	2	-	12	1	12	-	9	1	1	-
Clerodendrum longiflorum	2	1	2	-	13	1	13	-	9	1	1	-
Clerodendrum tomentosum	2	1	2	-	34	1	34	-	9	1	1	-
Stachys sylvatica	4	1	4	-	17	1	17	-	13	1	3	-
Stachys arvensis	4	1	4	-	18	1	18	-	13	1	3	-
Leonotis leonurus	5	1	5	-	20	1	20	-	14	1	3	-
Leonotis nepetifolia	5	1	5	-	19	1	19	-	14	1	3	-
Scutellaria galericulata	1	1	12	-	8	1	8	-	5	1	1	-
Scutellaria lateriflora	1	1	12	-	8	1	8	-	5	1	1	-
Scutellaria rehderiana	1	1	12	-	8	1	8	-	5	1	1	-
Tinnea rhodesiana	1	1	13	-	9	1	9	-	6	1	1	-
Tinnea aethiopica	1	1	13	-	9	1	9	-	6	1	1	-
Tectona grandis	1	1	11	-	7	1	7	-	11	1	1	-
Hyptis brachiata	6	1	18	-	27	1	27	-	16	1	4	-
Hyptis alata	6	1	18	-	28	1	28	-	16	1	4	-
Hyptis brevipes	6	1	18	-	31	1	31	-	16	1	4	-
Hyptis mutabilis	6	1	18	-	29	1	29	-	16	1	4	-
Hyptis verticillata	6	1	18	-	30	1	30	-	16	1	4	-
Solenostemon scutellarioides	6	1	6	-	25	1	25	-	15	1	4	-
Ocimum kilimandscharicum	6	1	6	-	22	1	22	-	15	1	4	-
Ocimum basilicum	6	1	6	-	21	1	21	-	15	1	4	-
Ocimum gratissimum	6	1	6	-	23	1	23	-	15	1	4	-
Ocimum filamentosum	6	1	6	-	26	1	26	-	15	1	4	-
Orthosiphon parvifolius	6	1	6	-	24	1	24	-	15	1	4	-
Salvia axillaris	7	1	7	-	33	1	33	-	17	1	4	-
Salvia rosmarinus	8	1	8	-	32	1	32	-	18	1	4	-
Vitex trifolia	1	1	1	-	1	1	1	-	1	1	1	-
Vitex agnus-castus	1	1	1	-	1	1	1	-	1	1	1	-
Vitex rehmannii	1	1	1	-	2	1	2	-	1	1	1	-

(b)

Species	JC69 Jukes-Cantor				K80 Kimura				Simple Distance			
	Initial		Recursive		Initial		Recursive		Initial		Recursive	
	1.0	1.5	1.0	1.5	1.0	1.5	1.0	1.5	1.0	1.5	1.0	1.5
Premna mooiensis	2	-	2	-	7	-	7	-	1	1	-	-
Premna obtusifolia	2	-	2	-	7	-	7	-	1	1	-	-
Premna bracteata	2	-	2	-	7	-	7	-	1	1	-	-
Premna serratifolia	2	-	2	-	7	-	21	-	1	1	-	-
Premna microphylla	2	-	2	-	7	-	21	-	1	1	-	-
Premna menglaensis	2	-	2	-	7	-	21	-	1	1	-	-
Leonotis leonurus	2	-	2	-	3	-	3	-	1	1	-	-
Leonotis nepetifolia	2	-	2	-	3	-	3	-	1	1	-	-
Scutellaria rehderiana	2	-	2	-	3	-	3	-	1	1	-	-
Gmelina philippensis	2	-	2	-	6	-	6	-	1	1	-	-
Gmelina arborea	2	-	2	-	6	-	6	-	1	1	-	-
Gmelina hainanensis	2	-	2	-	6	-	6	-	1	1	-	-
Stachys sylvatica	2	-	2	-	2	-	2	-	1	1	-	-
Stachys arvensis	2	-	2	-	2	-	2	-	1	1	-	-
Vitex agnus-castus	2	-	2	-	5	-	5	-	1	1	-	-
Vitex trifolia	2	-	2	-	5	-	5	-	1	1	-	-
Vitex rehmannii	2	-	2	-	5	-	5	-	1	1	-	-
Vitex doniana	2	-	2	-	5	-	5	-	1	1	-	-
Vitex meliocarpa	2	-	2	-	5	-	5	-	1	1	-	-
Stachytarpheta frantzii	1	-	1	-	1	-	1	-	1	1	-	-
Stachytarpheta cayennensis	1	-	1	-	1	-	1	-	1	1	-	-
Stachytarpheta jamaicensis	1	-	1	-	1	-	1	-	1	1	-	-
Stachytarpheta glabra	1	-	1	-	1	-	1	-	1	1	-	-
Phyla nodiflora	1	-	5	-	18	-	18	-	1	1	-	-
Phyla lanceolata	1	-	5	-	18	-	18	-	1	1	-	-
Scutellaria galericulata	2	-	2	-	12	-	12	-	1	1	-	-
Scutellaria lateriflora	2	-	2	-	12	-	12	-	1	1	-	-
Verbena hastata	3	-	3	-	17	-	17	-	1	1	-	-
Verbena simplex	3	-	3	-	17	-	17	-	1	1	-	-
Verbena stricta	3	-	3	-	17	-	17	-	1	1	-	-
Verbena urticifolia	3	-	3	-	17	-	17	-	1	1	-	-
Verbena bonariensis	3	-	6	-	17	-	22	-	1	1	-	-
Tinnea rhodesiana	2	-	2	-	12	-	12	-	1	1	-	-
Tinnea aethiopica	2	-	2	-	12	-	12	-	1	1	-	-
Hyptis alata	2	-	2	-	10	-	10	-	1	1	-	-
Hyptis brachiata	2	-	2	-	9	-	9	-	1	1	-	-
Hyptis brevipes	2	-	2	-	10	-	10	-	1	1	-	-
Hyptis mutabilis	2	-	2	-	10	-	10	-	1	1	-	-
Hyptis verticillata	2	-	2	-	13	-	13	-	1	1	-	-
Salvia rosmarinus	2	-	2	-	11	-	11	-	1	1	-	-
Ocimum basilicum	2	-	2	-	8	-	8	-	1	1	-	-

Figure 2. continued

file 5). According to past study, we only considered the results of the sGMYC as the mGMYC may over/under parametise the number of ML clusters and objects (Fujisawa and Barraclough 2013). For both markers, the results are mostly consistent with the clades recovered from the ML and BI phylogeny except for the splits of *T. grandis* L., *Salvia* L. spp. and *Solenostemon scutellarioides* (L.) Codd to produce additional OTUs for the *matK*. In GMYC, only the genus *Stachytarpheta* Vahl. was strongly delineated (> 0.8), and seven genera (*Stachys* L., *Vitex*, *Gmelina*, *Verbena*, *Phyla* Lour, *Leonotis* (Pers.) R.Br., and *Clerodendrum*) were moderately delineated (0.5–0.75) for the *matK*, while the majority of the nodes received low support values for both the *matK* and *rbcL* markers.

3.3.5 *Comparison of the species delimitation methods based on the two gene loci:* Both the ML and BI inferred from the *matK* supported the monophyly of the Lamiaceae and Verbenaceae clade compared to the *rbcL* (table 2, supplementary files 3–6, Figures 1–3). Comparatively, the number of OTUs generated form all the methods are largely inconsistent. The bPTP and TaxonDNA analyses conducted for *rbcL* marker mainly delineate most species in *rbcL* gene, thus, relatively yielded a higher number of OTUs. For instance, all the species clearly delineated to species-level for TaxonDNA in *rbcL*, thus, displayed higher intraspecific and interspecific divergence among the groups. On the other hand, the GMYC method recovered the most consistent clades for the *matK*.

4. Discussion

4.1 *Sequence information and disparity index test*

Information retrieved from the public GenBank repositories has been successfully utilized in many scientific platforms, e.g., biogeography, conservation, ecology, and evolutionary studies (Benson *et al*. 2013). Accessibility to GenBank resources has contributed to the increasing rate of genetic studies because it saves cost and time. However, our observations agrees with previous report on the authenticity of sequences

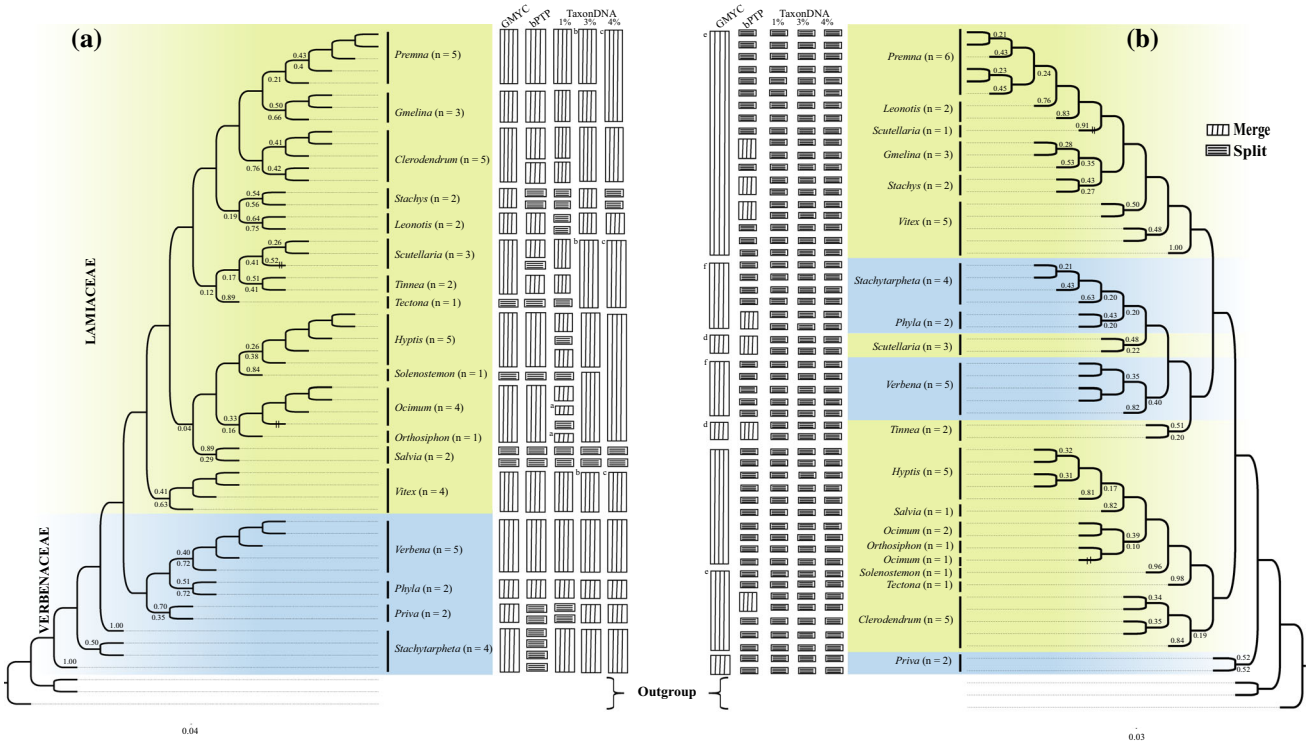

Figure 3. The tree topology and annotation is similar to figure 1. The support values of the bPTP and GMYC are presented above and below the branches respectively. Patterned vertical boxes indicate the OTUs and genetic groups. Genus/species that do not have sister relationships on the phylogenetic tree but formed the same OTUs are indicated with similar alphabets (supplementary files 3–6).

available in the GenBank (Griffiths *et al.* 2006). The *matK* and *rbcL* genes are some of the most available loci in the GenBank following their general acceptability as universal DNA barcode markers for plants (CBOL 2009; Parmentier *et al.* 2013; Gonzalez *et al.* 2009. Likewise, these loci are the most available marker for the families Lamiaceae and Verbenaceae.

The *matK* had higher nucleotide frequency and most parsimonious compared to *rbcL*. The DIT test may be a reputable means to recognize lineages that have undergone substantial dissimilar evolutionary processes, as demonstrated in our study. Interestingly, our finding is in line with the previous record on the high number of significant sequence pair for the *matK* compared to *rbcL* (Ashfaq *et al.* 2013) and revealed the species discriminatory power of these universal DNA barcodes in plants (Mattia *et al.* 2011; Stoeckle *et al.* 2011). Previous studies highlighted hybridization and different evolutionary rates in the plastid genes to be accountable for the genetic variation in plants (Fazekas *et al.* 2008). Importantly, the unique plastid DNA inheritance of pattern in plant species could be as a result of hybridization and/ecological adaptation overtime. These may be the underlying of factors responsible for the degree of substitution pattern between the group of species studied.

4.2 *Phylogenetic relationships of the families Lamiaceae and Verbenaceae*

Our phylogenetic analyses (BI and ML) based on the *matK* and *rbcL* yielded conflicting topologies between the two families (figure 1; supplementary figure 1). The *matK* successfully resolved the relationship into two large monophyletic clades (Verbenaceae and Lamiaceae) with strong supports while *rbcL* failed to address the tie. The phylogenetic relationship of two genera (*Tectona* and *Orthosiphon*) couldn't be ascertained due to only one sample represented in the study. Convincingly, the *matK* supported the expansion of the Lamiaceae to include *Clerodendrum*, *Gmelina*, *Premna*, *Tectona,* and *Vitex–proforma* Verbenaceae. Our results agree with the previous phylogenetic analysis based on chloroplast gene markers (Fazekas *et al.* 2009; Schäferhoff *et al.* 2010; Yuan *et al.* 2010;

Refulio-Rodriguez and Olmstead 2014). The topology retrieved from the *matK* gene presents a more explicit sister relationship among the genera- *Vitex*, as sister to *Salvia*; *Orthosiphon parvifolius* + monophyletic genus *Ocimum*; *Solenostemom scutellarioides* + *Hyptis* L. spp. and *Clerodendrum* spp. + *Gmelina* spp. + *Premna* spp. are sisters (figure 1) (Chen *et al*. 2014; Li *et al*. 2016; Fazekas *et al*. 2009; Elansary *et al*. 2017). The topology presented in this study is linked with sampling frequency. The *matK* provides better insight into the phylogenetic and systematic analysis of these families. Plastid genes have been successfully used to resolve complicated relationships at various taxonomic ranks (Ma *et al*. 2014). Generally, plastid genes (e.g., *matK*) can determine taxa relationship because it is more conservative than the nuclear DNA region (e.g., *ITS*). Species phylogenetic relationships inference based on evolutionary changes and patterns may result in ambiguous phylogenetic inferences (Bos and Posada 2015). The unresolved relationship could be attributed to standard genetic information amongst different genera, imperfect taxonomy due to species complex, incomplete lineage sorting, and inter/intra divergence variations (Fazekas *et al*. 2009). Therefore, we recommend further studies to include more representative samples for robust phylogenetic inference.

4.3 *Species delimitation using matK and rbcL*

Beyond any reasonable doubt, independent floral evolution and lack of floral parts especially, among closely related species, restrict morphological-based classification. The delimitation power > 70% (Hollingsworth *et al.* 2011) and strength of plant DNA barcodes amongst distantly related species have been demonstrated in the past (Kress *et al*. 2009; Burgess *et al*. 2011). This study affirms the inter/intraspecies delineation power of the *matK* and *rbcL* genes in plants (Mattia *et al*. 2011; Stoeckle *et al*. 2011). Most studies applied the single locus *matK* and *rbcL* regions to resolve the relationships amongst these families (Wagstaff *et al.* 1998; Chen *et al.* 2014; Li *et al.* 2016) rather than the species delimitation approaches for this group. Noteworthy, the application of these gene loci has been widely accepted as universal plant coding (CBOL 2009). The effectiveness of the DNA barcodes in the current study was based on their ability to differentiate among species. We evaluated the success rate of species identification using the four species delimitation methods.

Typically, species discrimination power varies with gene markers. Past studies have reported on the species delimitation power of several plastid DNA barcodes at different taxonomic levels (Kress *et al*. 2010). In this study, our results showed that r*bcL* could be a better choice of barcoding for the Lamiaceae and Verbenaceae compared to *matK* (figure 3). Similarly, high species discriminatory power was reported among the tropical coral islands of Xisha (Li *et al*. 2018). This is incongruence with the low species identification rate of *rbcL* reported in previous studies (Liu *et al*. 2011; Zhang *et al*. 2015). The *rbcL* resolved all the species clustered in the *matK* (inter-generic relationship, e.g., *Ocimum* + *Orthosiphon*; *Premna* + *Scutellaria* + *Tinnea* Kotschy ex Hook.f. + *Tectona* and *Vitex*, except *Salvia*). Low identification rate of plastid genes (e.g., *matK* and *rbcL*) in many taxonomic groups have been reported in previous studies, e.g., Schisandraceae (Zhang *et al*. 2015; Zamiaceae (Nicolalde-Morejón *et al*. 2010), Combretaceae (Gere *et al*. 2013), ornamental plants in Egypt (Elansary *et al*. 2017), and among the conifers in Mexico (Hernández-León *et al*. 2018). Also, the utilization of non-coding plastid and nuclear genes as an improvement on species discrimination amongst the closely related species has been demonstrated (Hollingsworth *et al.* 2011). Typically, extreme genetic disparities between lineages may clarify the reasons for different species identification power of DNA barcodes (Hollingsworth *et al.* 2011).

However, findings have proved the success of the barcode gap and distance methods to delimit plant groups (Zhang *et al*. 2012). Noteworthy, the maternal inheritance and different evolutionary rates in the chloroplast genomes perhaps play a vital role resulting in the inequitable delineation power of *matK* and *rbcL* (Dong *et al.* 2014; Vinitha *et al.* 2014; Guo *et al.* 2016). The barcode gap offers reliable means for species identification. However, it hasn't been a condition to delimit species (Saarela *et al.* 2013). Based on this study, the *rbcL* tends to provide more species identification resolution into the species than what was observed in *matK*. This may support the recommendation of the CBOL group on the universal of *rbcL* for land plants. The failure to separate closely related/complex species using the core barcodes is well-known (Liu *et al*. 2011; Kress *et al*. 2010; Kress *et al*. 2009; Jiang *et al*. 2011; Zhang *et al*. 2012). Previous studies associated this failure to wrong morphological based identification by taxonomist (due to loss of certain taxonomic dichotomous characters, e.g. fruits or flowers), high similarity to a wrongly identified reference sequence in the DNA sequences, database, and absence of the barcode gap (Jiang *et al*. 2011; Fu *et al*. 2011; Cowan and Fay 2012; Tang *et al.* 2014).

Correct and comprehensive molecular database, particularly for plant species, is still unrealistic in the current state.

The current study represents one of the efforts to evaluate the similarity of barcoding outputs from multiple delimitations. The tree-based identification is used as an initial step to delineate among closely/distantly related species, despite their limitations among the various species (Knowles and Carstens 2007; Hamilton *et al*. 2014). In this study, it is evident that the *matK* and *rbcL* genes produced different topologies and displayed varying discrimination power concerning the delimitation methods. Again, our study validates integrative species delimitation methods for barcoding analysis (Hamilton *et al.* 2014; Weiss *et al.* 2014). Our study revealed the *rbcL* marker analyzed with bPTP and TaxonDNA as the most reliable species delimitation methods, while the *matK* generated genetic groups/OTUs in these methods (figure 3). The genetic clusters generated from the bPTP and GMYC models differ for the markers. Although both models are alternatives, but operate on different algorithms. Comparatively, the bPTP requires less computation time than the GMYC because it utilizes a non-ultrametric tree.

Notable is the split of two genera (*Stachytarpheta* and *Phyla*) of the family Verbenaceae in the bPTP, which was merged in the GMYC. Also, the ABGD method generated markedly different genetic clusters for both markers (figure 2). Our study revealed that the K80 Kimura method outperformed other methods irrespective of the relative gap width (e.g., genus *Hyptis*). Hence, it may be proposed as a species-level delimitation tool in ABGD for plants. Interestingly, the TaxonDNA analysis based on the *rbcL* with a threshold value (1%, 3%, and 4 %) demarcated all the species of the families Lamiaceae and Verbenaceae studied. Thus, this finding disagrees with the poor performance previously reported (Vinitha *et al.* 2014; Saddhe *et al.* 2017).

Based on our results, when two or more species cluster, it may indicate that there are many species present within the populations and not necessarily weak interspecific or intraspecific divergence of the markers. Our study provides supporting evidence because the sampling was limited to the genus-level in the study. Thus, the GMYC presents low support values for most nodes between the species of the two families studied. However, the GMYC generated genetic clusters that are of taxonomic relevance. It is worthy to note that the accuracy of the species delimitation methods is influenced by the number of gene markers (Hamilton *et al.* 2014; Saddhe *et al.* 2017), Species sampling frequency (Pollock *et al*. 2002; Saddhe *et al*. 2017), and accurate selection of species delimitation algorithms. The current study presents the comparative discrimination power of the two core DNA barcodes to provide insights into the phylogeny and species-level delimitation of the families Lamiaceae and Verbenaceae. However, we suggest future DNA barcodes studies to include more representative species from these families with combined gene markers for robust taxonomic assignment. Our findings showed that the TaxonDNA analysis based on *rbcL* was the most effective barcoding approach for identifying the species of the Lamiaceae and Verbenaceae.

5. Conclusions

Our study confirms the two plastid markers *matK* and *rbcL* as a reliable plant barcodes based on ABGD, bPTP, TaxonDNA, and GMYC species delimitation methods. The *matK* gene fully resolved phylogenetic relationships of the families Lamiaceae and Verbenaceae. Our study supported the revision of the family Lamiaceae to include the *proforma* core Verbenaceae species. Also, we established the efficiency of the *rbcL* gene for species-level identifications using the TaxonDNA method at 1%, 3%, and 4% threshold values. The success of species discrimination analysis relies on accurate morphological-based identification, availability of correct sequences in the public molecular database and geographic diversity of species. On this note, we suggest further studies to include more species and multiple DNA barcode markers for robust evaluation and inferences for these families. It is evident that the systematic revision of both families will benefit from accurate species identification. In sum, the two plastid loci may be significant in solving problems concerning ecology and evolutionary relationships among plants. Also, the utilization of these markers sheds more light on the links of co-existing species.

Acknowledgements

The authors gratefully acknowledge Dr. Nneji Lotanna Micah of the Kunming Institute of Zoology, Chinese Academy of Sciences, China, and all those who have contributed towards the success of this research. Also, we appreciate the intellectual guidance of the anonymous reviewers that improved the quality of this manuscript.

References

Angiosperm Phylogeny Group 2009 An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* **161** 105–121

Abu-Asab MS and Cantino PD 1992 Pollen morphology in subfamily Lamioideae (Labiatae) and its phylogenetic implications; in *Advances in Labiate Science* (Eds.) RM Harley and H Reynolds (Royal Botanic Gardens, Kew, Richmond) pp. 97–112

Ashfaq M, Asif M, Anjum ZI and Zafar Y 2013 Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (*Gossypium*: Malvaceae). *Mol. Ecol. Resour.* **13** 574–582

Atkins S 2004 Verbenaceae; in *The families and genera of flowering plants* (Ed) JW Kadereit (Springer-Verlag, Berlin, Germany) pp. 449–468

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS and Donoghue MJ 1995 The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Ann. Mo. Bot. Gard.* **82** 247–277

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J and Sayers EW 2013 GenBank. *Nucl. Acids Res.* **41** D36–D42

Bos DH and Posada D 2015 Using models of nucleotide evolution to build phylogenetic trees. *Dev. Comp. Immunol.* **29** 211–227

Bremer B, Bremer K, Heidari N, Erixon P, Olmstead RG, Anderberg AA, Kallersjo M and Barkhordarian E 2002 Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. *Mol. Phyl. Evol.* **24** 274–301

Burgess K, Fazekas A J, Kesanakurti PR, Graham SW, Husband BC and Barrett SCH 2011 Discriminating plant species in a local temperate flora using the *rbcL* plus *matK* DNA barcode. *Methods Ecol. Evol.* **2** 333–340

Cantino PD 1992 Evidence for a polyphyletic origin of the Labiatae. *Ann. Mo. Bot. Gard.* **79** 361–379

Castresana J 2000 Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17** 540–552

CBOL 2009 Plant Working Group, A DNA barcode for land plants. *Proc. Nat. Acad. Sci.* USA **106** 12794–12797

Chen J, Zhao J, Erickson DL, Xia N and Kress WJ 2015 Testing DNA barcodes in closely related species of *Curcuma* (Zingiberaceae) from Myanmar and China. *Mol. Ecol. Resour.* **15** 337–348

Chen, Y P, Li B, Olmstead RG, Cantino PD, Liu ED and Xiang CL 2014 Phylogenetic placement of the enigmatic genus *Holocheila* (Lamiaceae) inferred from plastid DNA sequences. *Taxon* **63** 355–366

Cowan RS and Fay MF 2012 Challenges in the DNA barcodingof plant material. *Mol. Biol.* **862** 23–33

Cronquist A 1981 *An integrated system of classification of flowering plants* (Columbia University Press, New York)

Dahlgren R 1983 General aspects of angiosperm evolution and macrosystematics. *Nordic J. Bot.* **3** 119–149

De Mattia F, Bruni I, Galimberti A, Cattaneo F, Casiraghi M and Labra M 2011 A comparative study of different DNA barcoding markers for the identification of some members of Lamiacaea. *Food Res Int.* **44** 693–702

Dong W, Cheng T, Li C, Xu C, Long P and Zhou CCS 2014 Discriminating plants using the DNA barcode *rbcL*b:an appraisal based on a large data set. *Mol. Ecol. Resour.* **14** 336–343

Doyle JJ, Doyle JL, Ballenger JA, Dickson EE, Kajita T and Ohashi H 1997 A phylogeny of the chloroplast gene *rbcL* in the Leguminosae: taxonomic correlations and insights into the evolution of nodulation. *Am. J. Bot.* **84** 541–554

Dumas P, Barbut J, Le Ru B, Silvain JF, Clamens AL, *et al* 2015 Phylogenetic molecular species delimitations unravel potential new species in the pest genus *Spodoptera* GueneÂe, 1852 (Lepidoptera, Noctuidae). *PLoS ONE* **10** e0122407

Drummond AJ and Rambaut A 2007 BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7** 214

Drummond AJ, Suchard MA, Xie D and Rambaut A 2012 Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29** 1969–1973

Elansary HO, Ashfaq M, Ali HM and Yessoufou K 2017 The first initiative of DNA barcoding of ornamental plants from Egypt and potential applications in horticulture industry. *PLoS ONE* **12** e0172170

Ezard T, Fujisawa T and Barraclough TG 2009 splits: SPecies' LImits by Threshold Statistics. R package version. 1.0–14/r31 (*http://R-Forge.R-project.org/projects/splits/*)

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M and Barrett SC 2008 Multiple multilocus DNA barcodes from the plastid genome discriminate plantspecies equally well. *PLoS ONE* **3** e2802

Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, Barrett SCH, Newmaster SG, Ibabaei MH and Husband BC 2009 Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Mol. Ecol. Resour.* **9** 130–139

Fu YM, Jiang W M and Fu CX 2011 Identification of species within *Tetrastigma* (Miq.) Planch. (Vitaceae) based on DNA barcoding techniques. *J. Syst. Evol.* **49** 237–245

Fujisawa T and Barraclough TG 2013 Delimiting species using single-locus data and the generalized mixed yule coalescent approach: a revised method and evaluation on simulated data sets. *Syst. Biol.* **62** 707–24

Gere J, Yessoufou K, Daru BH, Mankga LT, Maurin O and van der Bank M 2013 Incorporating trnH-psbA to the

core DNA barcodes improves significantly species discrimination within southern African Combretaceae, in Z.T. Nagy, T. Backeljau, M. De Meyer, K. Jordaens (Eds) DNA barcoding: a practical tool for fundamental and applied biodiversity research. *ZooKeys* **365** 127147

Gernhard T 2008 The conditioned reconstructed process. *J. Theor. Biol.* **253** 769–778

Gonzalez MA, Baraloto C, Engel J, Mori SA, Pétronelli P, Riéra B, Roger A, Thébaud C and Chave J 2009 Identification of amazonian trees with DNA barcodes. *PLoS ONE* **4** e7483

Griffiths AJF, Miller H, Suzuki D T, Lewontin RC and Gelbart WM 2006 *An Introduction to Genetic Analysis* (New York: W. H. Freeman)

Guo YY, Huang LQ, Liu ZJ and Wang XQ 2016 Promise and Challenge of DNA Barcoding in *Venus slipper* (Paphiopedilum). *PLoS ONE* **11** e0146880

GRIN 2004 GRIN genera sometimes placed in Verbenaceae (Germplasm Resources Information Network: United States Department of Agriculture)

Hamilton CA, Hendrixson BE, Brewer MS and Bond JE 2014 An evaluation of sampling effects on multiple DNA barcoding methods leads to an integrative approach for delimiting species: a case study of the North American tarantula genus Aphonopelma (Araneae, Mygalomorphae, Theraphosidae). *Mol. Phyl. Evol*. **71** 79–93

Hernández-León S, Little DP, Acevedo-Sandoval O, Gernandt DS, Rodríguez-Laguna R, Saucedo-García M, Arce-Cervantes O, Razo-Zárate R and Espitia-López J 2018 Plant core DNA barcode performance at a local scale: identification of the conifers of the state of Hidalgo, Mexico. *Syst. Biod*. **16** 791–806

Hollingsworth PM, Graham SW and Little DP 2011 Choosing and using a Plant DNA barcode. *PLoS ONE* **6** e19254

Hutchinson J, Dalziel JM, Keay RWJ and Hepper FN 1963 *Flora of West tropical Africa* (Crown Agents for Oversea Governments and Administrations, London)

Jiang Y, Ding C, Zhang L, Yang R, Zhou Y and Tang L 2011 Identification of the genus *Epimedium* with DNA barcodes. *J. Med. Plants Res.* **5** 6413–6417

Knowles LL and Carstens BC 2007 Delimiting species without monophyletic gene trees. *Syst. Biol.* **56** 887–895

Kress WJ and Erickson DL 2007 A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS ONE* **2** e508

Kress WJ, Erickson DL, Andrew-Jones F, Swenson NG, Perez R, Sanjur O and Bermingham E, 2009 Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proc. Nat. Acad. Sci.* **106** 18621–18626

Kress WJ, Erickson DL, Swenson NG, Thompson J, Uriarte M and Zimmerman JK 2010 Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican Forest Dynamics Plot. *PLoS ONE* **5** e15409

Kumar S and Gadagkar S R 2001 Disparity Index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. *Genetics* **158** 1321–1327

Kumar S, Stecher G and Tamura K 2016 MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33** 1870–1874

Lahaye R, Savolainen V, Duthoit S, Maurin O and van der Bank M 2008 A test of *psbK-psbI* and *atpF-atpH* as potential plant DNA barcodes using the flora of the Kruger National Park (South Africa) as a model system. Nature Precedings. Available from *http://hdl.handle.net/10101/npre.2008.1896.1*. Accessed 2 Feb 2020

Lavin M, Herendeen PS and Wojciechowski MF 2005 Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Syst. Biol.* **54** 530–549

Li B, Cantino PD, Olmstead RG, Bramley GLC, Xiang CL, Ma ZH, Tan YH and Zhang DX 2016 A large-scale chloroplast phylogeny of the Lamiaceae sheds new light on its subfamilial classification. *Sci. Rep.* **6** 34343

Li HQ, Chen JY, Wang S and Xiong SZ 2012 Evaluation of six candidate DNA barcoding loci in *Ficus* (Moraceae) of China. *Mol. Ecol. Resour.* **12** 783–790

Li S, Qian X, Zheng Z, Shi M, Change X, Li X, Liu J, Tu T and Zhang D 2018 DNA barcoding the flowering plants from the tropical coral islands of Xisha (China). *Ecol. Evol.* **8** 10587–10593

Liu J, Moeller M, Gao LM, Zhang DQ and Li DZ 2011 DNA barcoding for the discrimination of Eurasian yews (*Taxus* L., Taxaceae) and the discovery of cryptic species. *Mol. Ecol. Res.* **11** 89–100

Ma PF, Zhang YX, Zeng CX, Guo ZH and Li DZ 2014 Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae). *Syst. Biol.* **63** 933–950

Marx HE, O'Leary N, Yuan Y, Lu-Irving P, Tank DC, Mulgara ME and Olmstead RG 2010 A molecular phylogeny and classification of Verbenaceae. *Am. J. Bot*. **97** 1647–1663

Meier R, Kwong S, Vaidya G and Ng P K L 2006 DNA Barcoding and taxonomy in *Diptera*: a tale of high intraspecific variability and low identification success. *Syst. Biol*. **55** 715–728

Miller MA, Pfeiffer W and Schwartz T 2010 Creating the CIPRES science gateway for inference of large phylogenetic trees, in *Proceedings of the Gateway Computing Environments Workshop*, New Orleans, LA. pp. 1–8

Nee S 2006 Birth-death models in macroevolution. *Annu. Rev. Ecol. Evol. Syst.* **37** 1–17

Nicolalde-Morejón F, Vergara-Silva F, González-Astorga J and Stevenson D W 2010 Character-based, population-

level DNA barcoding in Mexican species of *Zamia* L. (Zamiaceae: Cycadales). *Mitochondrial DNA* **21** 51–59

Olmstead RG, dePamphilis CW, Wolfe AD, Young ND, Elisens WJ and Reeves PA 2001 Disintegration of the Scrophulariaceae. *Am. J. Bot.* **88** 348–361

Olmstead RG, Jansen RK, Kim KJ and Wagstaff SJ 2000 The phylogeny of the Asteridae *s.l.* based on chloroplast *ndh*F sequences. *Mol. Phyl. Evol.* **16** 96–112

Oxelman B, Kornhall P, Olmstead RG and Bremer B 2005 Further disintegration of the Scrophulariaceae. *Taxon* **54** 411–425

Paradis E, Claude J and Strimmer K 2004 APE: analyses of phylogenetics and evolution in R language. *Bioinfomatics* **20** 289–290

Parmentier I, Duminil J, Kuzmina M, Philippe M, Thomas D W, Kenfack D, Chuyong GB, Cruaud C and Hardy OJ 2013 How effective are DNA barcodes in the identification of African rainforest trees? *PLoS ONE* **8** e54921

Pennisi E 2007 Taxonomy. Wanted: A barcode for plants. *Science* **318** 190–191

Pollock D, Zwick D J, McGuire JA and Hillis D M 2002 Increased taxon sampling is advantageous for phylogenetic inference. *Syst. Biol.* **51** 664–671

Posada D and Buckley TR 2004 Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* **53** 793–808

Puillandre N, Lambert A, Brouillet S and Achaz G 2011 ABGD, Automatic barcode discovery for primary species delimitation. *Mol. Ecol.* **21** 1864–1877

R Core Team 2012 R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rambaut A and Drummond A J 2009 Tracer version 1.5 [computer program] (*http://beast.bio.ed.ac.uk*)

Rambaut A and Drummond A J 2010. TreeAnnotator version 1.6.1 [computer program] (*http://beast.bio.ed.ac.uk*)

Rambaut A 2014 FigTree v1.4.2, a graphical viewer of phylogenetic trees (*http://tree.bio.ed.ac.uk/software/figtree/*)

Refulio-Rodriguez N F and Olmstead R G 2014 Phylogeny of Lamiidae. *Am. J. Bot.* **101** 287–299

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA and Huelsenbeck JP 2012 MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61** 539–542

Rozas J 2009 Polymorphism Analysis using DnaSP, in *Bioinformatics for DNA Sequence Analysis: Methods in Molecular Biology Series* (Ed) Posada D (Humana Press NJ USA) pp. 337–350

Ryding O 1995 Pericarp structure and phylogeny of the Lamiaceae-Verbenaceae-complex. *Plant Syst. Evol.* **198** 101–141

Saarela JM, Sokoloff PC, Gillespie LJ, Consaul LL and Bull RD 2013 DNA barcoding the Canadian Arctic flora: core plastid barcodes (*rbcL* + *matK*) for 490 vascular plant species. *PLoS ONE* **8** e77982

Saddhe AA, Jamdade RA, and Kumar K 2017 Evaluation of multilocus marker efficacy for delineating mangrove species of West Coast India. *PLoS ONE* **12** e0183245

Sahu S K, Singh R, and Kathiresan K, 2016 Multi-gene phylogenetic analysis reveals the multiple origin and evolution of mangrove physiological traits through exaptation, Estuarine, Coastal and Shelf. *Science* **183** 41–51

Sanders R W 2001 The genera of Verbenaceae in the southeastern United States. *Harvard Papers Bot.* **5** 303–358

Schäferhoff B, Fleischmann A, Fischer E, Albach DC, Borsch T, Heubl G and Müller KF 2010 Towards resolving Lamiales relationships: insights from rapidly evolving chloroplast sequences. *BMC Evol. Biol.* **10** 352

Scheen A, Bendiksby M, Ryding O, Mathiesen C, Albert VA and Lindqvist C 2010 Molecular phylogenetics, character evolution, and suprageneric classification of Lamioideae (Lamiaceae). *Ann. Mo. Bot. Gard.* **97** 191–217

Selvaraj D, Sarma RK, and Sathishkumar R 2008 Phylogenetic analysis of chloroplast *matK* gene from Zingiberaceae for plant DNA barcoding. *Bioinformation* **3** 24–27

Stamatakis A 2006 RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22** 2688–2690

Stoeckle MY, Gamble CC, Kirpekar R, Young G, Ahmed S and Little D P 2011 Commercial teas highlight plant DNA barcode identification successes and obstacles. *Sci. Rep.* **1** 42

Surya SD, Swati S, and Partadeb G 2014 Phylogenetic relationships among three species of the mangrove genus *Avicennia* found in Indian Sundarban, as revealed by RAPD analysis. *Asian J. Plant Sci.* **4** 25–30

Swofford DL 2002 PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4b10. Sinauer Associates, Sunderland, Massachusetts, USA

Talavera G and Castresana J 2007 Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56** 564–577

Tang CG, Humphreys AM, Fontaneto D and Barraclough TG 2014 Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. *Methods Ecol. Evol.* **5** 1086–1094

Thorne RF 1983 Proposed new realignments in the angiosperms. *Nordic J. Bot.* **3** 85–117

Vinitha RM, Kumar SU, Aishwarya K, Sabu M and Thomas G 2014 Prospects for discriminating Zingiberaceae species in India using DNA barcodes. *J. Integr Plant Biol.* **56** 760–773

Wagstaff SJ and Olmstead RG 1997 Phylogeny of the Labiatae and Verbenaceae inferred from *rbc*L sequences. *Syst. Bot.* **22** 165–179

Wagstaff SJ, Hickerson L, Spangler R, Reeves PA and Olmstead R G 1998 Phylogeny in Labiatae *s.l.*, inferred from cpDNA sequences. *Plant Syst. Evol.* **209** 265–274

Weiss M, Macher J N, Seefeldt MA, and Leese F 2014 Molecular evidence for further overlooked species within the *Gammarus fossarum* complex (Crustacea: Amphipoda). *Hydrobiologia* **721** 165–184

Xiang XG, Hu H, Wang W and Jin XH 2011 DNA barcoding of the recently evolved genus *Holcoglossum* (Orchidaceae: Aeridinae): a test of DNA barcode candidates. *Mol. Ecol. Resour.* **11** 1012–1021

Yan LJ, Liu J, Möller M, Zhang L, Zhang XM, Li DZ and Gao LM 2015 DNA barcoding of *Rhododendron* (Ericaceae), the largest Chinese plant genus in biodiversity hotspots of the Himalaya-Hengduan Mountains. *Mol. Ecol. Resour.* **15** 932–944

Yuan YW, Mabberley DJ, Steane DA and Olmstead RG 2010 Further disintegration and redefinition of *Clerodendrum* (Lamiaceae): implications for the understanding of the evolution of an intriguing breeding strategy. *Taxon* **59** 125–133

Zhang CY, Wang FY, Yan HF, Hao G and Hu CM 2012 Testing DNA barcoding in closely related groups of *Lysimachia* L. (Myrsinaceae). *Mol. Ecol. Resour.* **12** 98–108

Zhang J, Kapli P and Pavlidis P 2013 A. Stamatakis, A General species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29** 2869–2876

Zhang J, Chen, M, Dong X, Lin, R, Fan J and Chen Z 2015 Evaluation of four commonly used DNA barcoding loci for chinese medicinal plants of the family Schisandraceae. *PLoS ONE* **10** e0125574

Zou S, Fei C, Song J, Bao Y, He M, and Wang C 2016 Combining and comparing coalescent, distance and character-based approaches for barcoding microalgaes: a test with Chlorella-like species (Chlorophyta). *PLoS ONE* **11** e0153833

Corresponding editor: NG PRASAD