RHEEDEA
Journal of the Indian Association for Angiosperm Taxonomy
RESEARCH ARTICLE
Vol. 34(3): 165–182 (2024)
ISSN: 0971-2313 (Print edition)
ISSN: 2582-2438 (Online edition)
https://dx.doi.org/10.22244/rheedea.2024.34.03.01

How to best utilise molecular data to support new taxa – a practical guide in less than 5000 words

Möller M.*, Atkins H. & K. Nishii
Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh – EH3 5LR, Scotland, UK
*E-mail: mmoeller@rbge.org.uk

Abstract: Increasingly, molecular data are used in support of new taxa (species or genera). Understanding species` delimitation is key in this process, and also an essential prerequisite for a stable taxonomy and classification as well as for conservation. The strongest approach to delineate species is to utilize the congruence between the distribution of morphological characteristics and the molecular phylogenetic tree structure. One key aspect here is the inclusion of multiple samples per species. The approach is applicable irrespective of data source, Sanger sequencing or NGS approaches. A set of scenarios and recommendations for a best practice workflow are presented here for taxonomists considering to include molecular phylogenetic data in support of new taxa.

Keywords: Gesneriaceae, Molecular phylogeny, Species delineation, Species delimitation, Taxonomy.

Introduction

The last few years have seen a remarkable increase in the number of species new to science as exemplified by the plant family Gesneriaceae in China (Möller *et al.*, 2016; Möller, 2019; Wen *et al.*, 2019, 2021). Increasingly molecular analyses are included in publications, in support of taxa (species or genus), and this comes with opportunities and challenges (*e.g.*, Meier & Wheeler, 2008). Approaches to include molecular data, when executed properly, can successfully support and delineate new species and examples can be found in the literature (*e.g.*, Boluda *et al.*, 2022; Hassanpour *et al.*, 2023) and has been integrated in modern species concepts (De Queiroz, 2007). Here we describe an approach utilizing the congruence of morphological characteristics with a molecular phylogenetic tree structure (*e.g.*, Duminil & Di Michele, 2009; Ranasinghe *et al.*, 2016; Janeesha *et al.*, 2023).

Since the addition of molecular data is comparatively costly compared to traditional taxonomic work, authors might want to consider the appropriate use of molecular data. In some cases for new species that are morphologically distinctly different from all other existing congeners perhaps morphology itself can be enough. For example the resupinate flowers of *Boea resupinata* Zich & B.Gray (2021) are unique in the genus, as are the monochasial branching inflorescences of *Primulina anisocymosa* F.Wen, Xin Hong & Z.J.Qiu (Hong *et al.*, 2019). In these cases, the addition of molecular data to these studies may increase the support but are not necessarily needed.

In other cases, molecular data has been essential in systematically placing a new species in the correct genus. For instance, the morphological circumscription has recently widened in some Gesneriaceae genera, such as *Oreocharis* Benth. (Möller *et al.*, 2011b), or *Primulina* Hance (Wang *et al.*, 2011; Weber *et al.*, 2011a), after their redefinition based on molecular phylogenetic studies. The widening of morphological diversity in a genus can cause uncertainties in the placement of new species in the correct genus

Received: 16.04.2024; *Revised & Accepted*: 15.08.2024
Published Online: 30.09.2024

and the addition of molecular data are helpful in this respect *e.g.*, by demonstrating these species fall within the clade of a particular genus with strong support. This is neatly demonstrated in the genus *Petrocodon* Hance. It was, in its first inception, characterised by a white sub-urceolate corolla, and two stamens in anterior position (Hance, 1883). It was expanded to include species with purple, red or yellow flowers, cylindrical to infundibuliform corollas, and five stamens in actinomorphic corollas (Weber *et al.*, 2011b). This morphological widening of the generic concept in *Petrocodon,* and the availability of molecular data in public databases, has prompted the inclusion of molecular phylogenetic analyses in support of new taxa. The placement of *Petrocodon hunanensis* X.L.Yu & Ming Li when described was also supported by molecular phylogenetic analyses (Yu *et al.*, 2015), perhaps because the four stamens were outside of the generic diversity of *Petrocodon* at that time. Another example of generic misplacement is *Primulina guangxiensis* Yan Liu & W.B.Xu, which was moved to *Petrocodon* (as *Petrocodon guangxiensis* (Yan Liu & W.B.Xu) W.B.Xu & K.F.Chung) after molecular phylogenetic analyses were performed (Liu *et al.*, 2011; Xu *et al.*, 2014). Here, hidden homoplasies in corolla morphology caused the species to be systematically placed in the wrong genus due to its similarity in corolla morphology with the type species of *Primulina*, *P. tabacum* Hance.

At the generic rank, the addition of molecular data is essential for the establishment of new genera showing a paucity of morphological differences, *e.g.*, for *Glabrella* Mich.Möller & W.H.Chen, where the only diagnostic characters for the genus are the glabrous appearance of the plant and the presence of short stems, distinguishing it from *Oreocharis* (Möller *et al.*, 2011b, 2014). Another example is *Langbiangia* Luu, C.L.Hsieh & K.F.Chung, that differs only by spirally-arranged leaves from *Primulina* (with opposite phyllotaxis), or by its acaulescent habit and unilocular ovary from *Deinostigma* W.T.Wang & Z.Y.Li (caulescent and at least basally bilocular ovary) (Luu *et al.*, 2023). Indeed, without molecular data species of *Middletonia* C.Puglisi would have remained in *Paraboea* (C.B.Clarke) Ridl. (Puglisi *et al.*, 2016). The genus differs from *Paraboea* only by a farinose glandular indumentum on the ovary and free and erect anthers. At species rank, there are a number of examples where molecular data has been critical in identifying species which are genetically distinct but have very similar morphology to other species in the genus. These include *Oreocharis oriolus* J.Hu & F.Wen (Hu *et al.*, 2023), *Henckelia umbellata* Kanthraj & K.N.Nair (Kanthraj *et al.*, 2023) or *Petrocodon ainsliifolius* W.H.Chen & Y.M.Shui and *Petrocodon viridescens* W.H.Chen, Mich.Möller & Y.M.Shui (Chen *et al.*, 2014).

Morphology-driven molecular phylogeny-supported species delimitation

The present guide is based on an approach that is based on the congruence between clade structure and distribution of morphological characteristics at the tips of a phylogenetic tree (*cf.*, Ranasinghe *et al.*, 2016; Janeesha *et al.*, 2023). Congruence of the distribution of morphological characteristics with the clade structure of molecular phylogenetic trees are seen as a critical step in the delimitation of species, including new ones. Carried out properly, this combination of molecular phylogeny and morphological characters in delimiting species is a very powerful method resulting in a stable taxonomy. It could also be helpful in the selection of characters for diagnosing new taxa from existing ones (*cf.* Nishii *et al.*, 2015). A few fundamental criteria to infer the molecular phylogenetic tree against which the morphology is to be evaluated should be considered, such as ingroup and outgroup sampling, inclusion of relevant samples, number of samples per species, molecular marker choice, source of existing molecular data, phylogenetic tree-building algorithms, and presentation of results. In the following a practical guide to sampling and analysis of morphological-molecular phylogenetic species delimitation is provided.

A pragmatic and practical approach to phylogenetic species delimitation

Fundamental to any application of a species concept is the delimitation of a species, and molecular data have opened the window to new tools (*e.g.*, Meier & Wheeler, 2008). There is no fundamental difference between the use of Sanger or NGS data for species delimitation, except perhaps for deep mega-sampled studies, which are expensive and often impractical for the description of a single species. For pragmatic reasons, here, we focus on a molecular phylogenetic approach.

1. Sampling

Ingroup: The species under study are members of the ingroup and should include ideally all existing species in a genus, but at the least those covering the full geographic or morphological range, with particular attention paid to including peripheral ones (geographically or morphologically), as these often include key samples. Those species that are morphologically similar to the new species and those occurring nearby should also be prioritised.

It is also very important to include type species of genera in cases where there is the potential for generic boundaries to be redrawn. Without these, new generic delimitation cannot be reliably drawn as the placement of some species remain uncertain (e.g., Wen *et al.*, 2022).

It would be difficult to prescribe a minimum proportion of included species as this would depend on the species number in a genus and the question being asked, but as a good starting point we would propose a minimum of two-third of the species in a genus. Needless to state that the ingroup samples should be a natural, monophyletic group. Where this is not immediately clear from previous published work, the inclusion of several outgroup samples is required (see below).

Outgroup and root: Suitable outgroup samples should not be part of the ingroup, be closely related to the ingroup, and not too distant as to cause sequence alignment ambiguities as likely in some published studies (*e.g.*, Hong *et al.*, 2019). The outgroup should also include the samples on which to root the phylogenies. Information on suitable outgroup candidates and species to root the trees usually come from previously published studies and should be consulted and cited as such. The purpose of the outgroup samples and rooting of a phylogenetic tree is to provide polarity (the direction of evolutionary change) and a reference point for the genetic diversity among the study samples. The use of several outgroup samples is required to further test the monophyly of the ingroup samples.

Sampling size: One of the most critical aspects in species delimitation is the decision on the number of samples for a target species to be included and can be between a single sample or population genetic-level sampling of 20–30 individuals (*e.g.*, Luikart & Cornuet, 1998; Ward & Jasieniuk, 2009; Hale *et al.*, 2012). The latter might be too high for species delimitation, but a single sample per species is insufficient as it does not represent the genetic depth of a species and is one of the most common errors (*e.g.*, Guo *et al.*, 2015; Yu *et al.*, 2015). The use of single samples per species in phylogenetic analyses is appropriate for unravelling relationships between species but does not delimit the species (see Fig. 1). This is illustrated in the very comprehensive phylogenetic analysis, that despite the inclusion of a large number of species (> 80% of the genus) does not provide evidence for a new species as only one sample per species was included (Xu *et al.*, 2023).

Here, we follow recommendations from DNA barcoding, where the use of 5–10 individuals per species is suggested (Meyer & Paulay, 2005; Knowles & Carstens, 2007), and for globally distributed species this is increased to 11–15 to capture the full genetic depth of a species (Yao *et al.*, 2017). It should perhaps be mentioned at this point that DNA barcoding is widely applied for the identification and delimitation of taxonomically well-understood species but is not suitable for undescribed species (Meyer & Paulay, 2005; Meier & Wheeler, 2008).

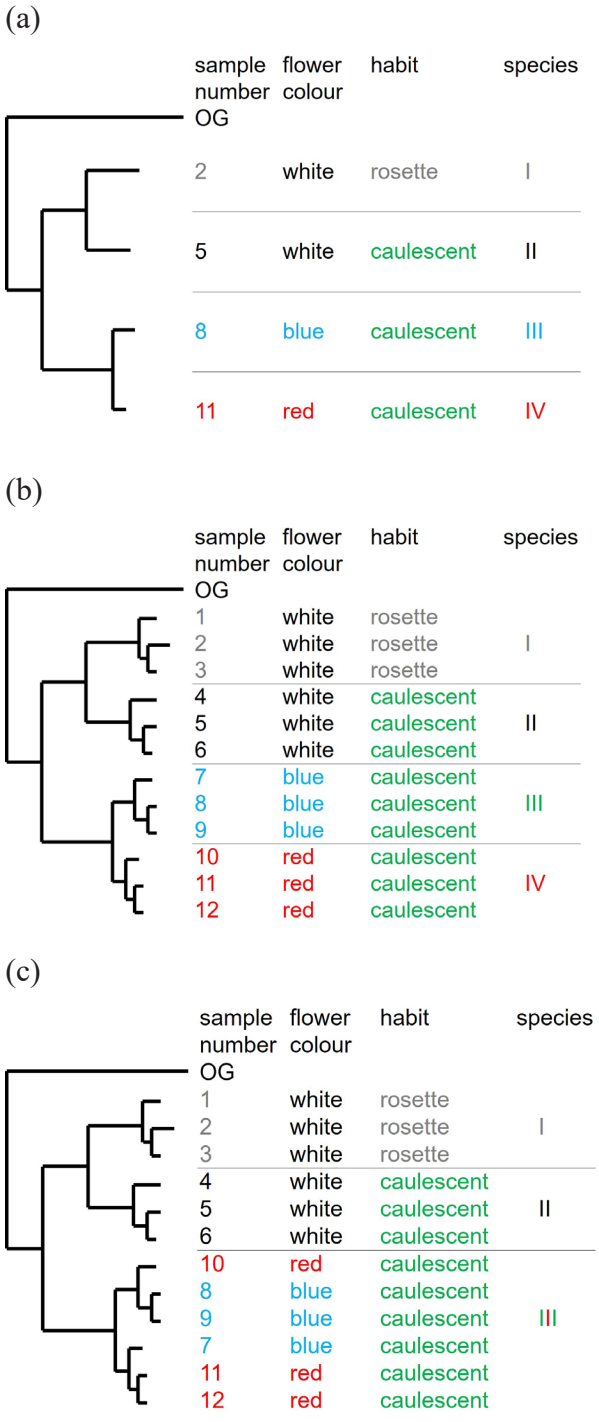


Fig. 1. Examples of the effect of the number of samples included per species in phylogenetic analyses on the interpretation of species delineation: **a.** Phylogram with one sample per species (I-III) and a new species IV (red flowered). The phylogram does not delimit the species as the genetic depth of the species are unknown; **b.** Phylogram with three samples included per species (I-IV), where the phylogram shows that the samples of the new species IV (red flowered) conform to species limits as they form a single clade without overlapping genetic ranges; **c.** Phylogram with three samples included per 'species', with the red-flowered samples scattered amongst those of the blue-flowered samples, indicating that the genetic gene pool of the two flower colour forms overlap and perhaps a species showing flower polymorphism exists.

For species delineation, we recommend the use of five samples per species, though a minimum would be at least three from different populations (*e.g.*, Tan *et al.*, 2023). These should come from areas across the distribution range of the species, and wider distributed ones should be represented by proportionally more samples. The reason for the inclusion of multiple samples of all species in the analysis is to detect the genetic boundaries and discontinuities between all species, to effectively delimit each species from the others, including any potential new ones (see Fig. 1). Where the new species is known from only the type locality, at least three samples should come from most distant parts of the population to capture any genetic variation that may exist across the locality.

2. Choice of molecular marker(s)

There is, in principle, no difference between the source of molecular data for species delimitation, whether from Sanger sequencing or NGS approaches, though the latter are sometimes impractical for the establishment of a single new species, and the analysis is then often compromised by the number of samples included due to cost implications.

Molecular data come principally from three different genomes, mitochondrial, chloroplast and nuclear, each with their own characteristics and caveats (*e.g.*, McKain *et al.*, 2018; Tyszka *et al.*, 2023). The mitochondrial genome is often too slowly evolving for species delimitation in plants, though they are sometimes used for population-level analyses within species where genetic variation have been found (*e.g.*, Grosser *et al.*, 2023). The chloroplast and nuclear genomes contain genes, introns and spacers, with increasing rates of evolution (*e.g.*, Small *et al.*, 1998; Stoebe *et al.*, 1999; Shaw *et al.*, 2005, 2007, 2014), with overall the nuclear genome (3 ×) and chloroplast (16 ×) evolving faster than mitochondrial genes (Wolfe *et al.*, 1987; Drouin *et al.*, 2008).

Selection of suitable candidate markers heavily depends on the taxonomic range studied. The aligned matrices should not contain ambiguously alignable regions and these should be removed before analysis (*e.g.*, Möller & Cronk, 1997). At species level for example, most often the nuclear ribosomal internal transcribed spacers (ITS) and chloroplast intron/spacers (*e.g.*, *atp*B-*rbc*L, *rpl*16, *trn*L-F, *trn*H-*psb*A) are used (*e.g.*, for Gesneriaceae see Möller & Clark, 2013).

The nuclear ribosomal internal transcribed spacer region (ITS) is the most-often used nuclear marker due to its abundant copies per cell and the availability of universal primers. The often-used original primers from White *et al.* (1990) were designed for fungi, although modified ones are available that are specific for plants (*e.g.*, Möller & Cronk, 1997; Douzery *et al.*, 1999). The ITS copies are greatly homogenised through concerted evolution (Wang *et al.*, 2023), though this breaks down where hybridisation is involved or more than one NOR locus (containing the ITS copies) per genome exist (Möller *et al.*, 2008; Puglisi *et al.*, 2011; Zhou *et al.*, 2017), where loci can diverge requiring cloning of PCR products (Denduangboripant & Cronk, 2000; Xiao *et al.*, 2010), as direct sequencing results in unusable polymorphic electropherograms. An elegant workaround would be an NGS approach using amplicon sequencing (amplicon-seq) (Nishii *et al.*, in prep) that can filter out majority copies (Paton, 2023).

The mitochondrial and chloroplast genomes are usually uniparentally inherited and in angiosperms usually maternally (*e.g.*, Mogensen, 1996; Greiner *et al.*, 2015), while nuclear genomes are biparentally inherited. Thus, the evolutionary history of a species is differently reflected by the three genomes, and incongruences/discordances between their phylogenies have been recorded repeatedly (*e.g.*, Soltis & Kuzoff, 1995; Rose *et al.*, 2020, and others). The underlying processes, hybridisation and incomplete lineage sorting, compound the analytical problem and are currently difficult to differentiate (*e.g.*, Pelser *et al.*, 2012; De Villiers *et al.*, 2013; Qin *et al.*, 2023). Phylogenies based on markers from one genome therefore represent gene trees rather than species trees. Thus, to fully unravel the status of a species, data from more than one genome, *i.e.*, chloroplast and nuclear origin, have to be included. The number of markers is equally important and accuracy in delimitation increases with increasing number of marker (*e.g.*, Knowles & Carstens, 2007). The more the better to obtain stable phylogenetic topologies, though minimum values have been discussed depending on the level of taxonomic sampling (Wortley *et al.*, 2005). The application of NGS-based methods is beneficial here, as these can cover hundreds of loci (*e.g.*, Ogutcen *et al.*, 2021; Yang *et al.*, 2023).

3. Data sources

Commonly, freshly silica-dried leaf material is used for DNA extraction (using mini preps *e.g.*, Doyle & Doyle, 1987; or commercial products), PCR amplification and sequencing. Though in some cases it is desirable and possible to use material from herbarium specimens although this may require booster PCRs (*e.g.*, Möller & Cronk, 2001; Särkinen *et al.*, 2012), or NGS approaches (*e.g.*, Ferrari *et al.*, 2023). Ideally, herbarium extractions and sequencing should be repeated to check for reproducibility. Obtaining data from herbarium specimens has the advantage that the specimens serve as, often verified, voucher specimens.

One of the most utilized sources for published sequence data is that of the NCBI, a data repository established in 1979 at the Los Alamos National Laboratory (US) and named GenBank in 1982 (https://en.wikipedia.org/wiki/GenBank). The database presently contains > 2.45 billion sequences, but these need to be treated with some caution when utilizing them for analyses, as these are uncurated submissions and often have no associated voucher specimen information and contain many erroneous entries, as recently indicated in Li *et al.* (2022). In addition, often only single samples per species are available making them of limited use for species delimitation (see above "1. Sampling").

4. Matrix building and alignment

DNA sequences evolve by base substitutions and indel (insertion and deletion) mutations, thus over evolutionary time diverge in length between species. A sequence alignment has the primary aim to adjust for these length differences by inserting alignment gaps in order to achieve this and to ensure that base pairs within a characters are homologous. Many programmes are available for an automated alignment (see https://www.bionity.com/en/encyclopedia/List_of_sequence_alignment_software.html), including the recently popular MAFFT (https://mafft.cbrc.jp/alignment/server/) (Katoh *et al.*, 2019). To ensure the primary homology of sequence positions the automated alignment need to be checked and optimised manually. Hypervariable regions where alternative alignment options exist need to be excluded as these would otherwise result in spurious tree topologies and/or lower branch support (*e.g.*, Möller & Cronk, 1997).

5. Phylogenetic tree building

Three methods are predominantly used to reconstruct phylogenetic trees, Maximum Parsimony (MP), Maximum Likelihood (ML), and Bayesian Inference (BI). The first is character-based, the latter two substitution model-based. There are many sources describing the theoretical and practical application of these methods (*e.g.*, Page & Holmes, 1998; Felsenstein, 2004; Nei & Kumar, 2000; Lemey *et al.*, 2009; Baum & Smith, 2012). The methods have advantages and disadvantages: BI and ML have the advantages of providing a single phylogenetic tree, while MP analyses can result in one to many equally most-parsimonious trees, and these are usually combined into a consensus tree where the topology resolution is sometimes reduced. Though this is not a problem for species delineation when this concerns nodes that reside within a species clade. Best practice would be to perform at least two analyses, a character-based (*e.g.*, MP) and a substitution model-based (*e.g.*, BI or ML) method as the data are analysed differently (see above), and the resulting topologies can be believed with more confidence when identical (or congruent) topologies for the two methods result.

Branch (or clade) support for the tree topology comes in various forms depending on the phylogenetic method used. For ML and MP, bootstrap analyses are usually performed, for ML often with fewer replicates (1,000) because of the computational complexity of ML. For reproducibility of results in MP bootstrap analyses at least 10,000 replicates should be run and can be achieved even with larger matrices of several hundred samples by switching on TBR (tree-bisection-reconnection, a branch swapping option to optimise the trees) and switching off MulTrees (to obtain only one tree per replicate) in PAUP to speed-up the bootstrap analysis (Spangler & Olmstead, 1999; Möller *et al.*, 2009). Branch support values for BI analyses come in the form of posterior probabilities, generated from the 50% majority-rule consensus trees sampled trees in a BI run minus the burn-in, though these are not identical to ML/MP bootstrap values and can often be too optimistic (*e.g.*, Möller *et al.*, 2009).

There are other tree-building methods based on genetic distances between species, such as Neighbor Joining (NJ) or UPGMA (often used in DNA barcoding), though these are based on distance matrices and the distances calculated depend on the distance option used, ranging from uncorrected distances to model-based ML distances but the resulting topologies can differ (*e.g.*, Kim *et al.*, 1993; Tateno *et al.*, 1994). These are not explicit phylogenetic character-based methods though, and the danger of using distance methods is that natural groups (as defined as shared characters by descent from a single common ancestor) can be missed particularly where fast evolving, or poor sequences, matrix misalignments or large taxonomic distances are involved. Then, phylogenetically distant samples may appear genetically closer to phylogenetically unrelated samples and fall in the wrong clade (Fig. 2).

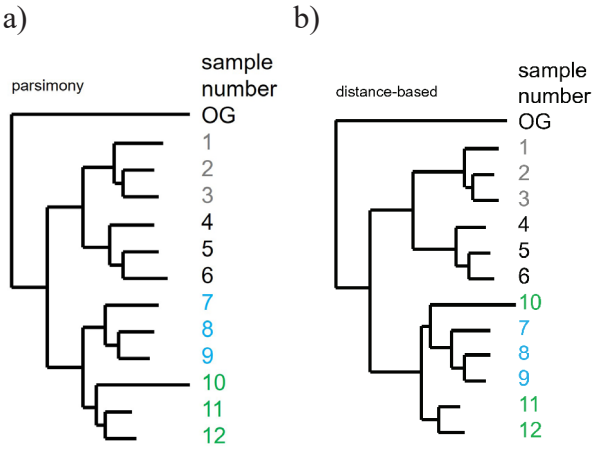


Fig. 2. Example of potential tree topology differences between parsimony and distance-based tree-building methods: **a.** Parsimony and **b.** Distance-based, where sample 10 falls among the blue samples; note the excessively long branch of sample 10.

When data from different genomes are used, *e.g.*, chloroplast and nuclear, tests for their combinability are available, such as the incongruence length difference (ILD) test (or partition homogeneity test) (Farris, 1995a, 1995b; Cunningham, 1997; Barker & Lutzoni, 2002). Where the test detects no incongruence, the data can be combined. Where there is significant incongruence, the phylogenetic data should be analysed separately, and the trees are shown separately (*e.g.*, Xu *et al.*, 2013; Huang *et al.*, 2017). Although there are some reported issues with the ILD test (discussed in Hipp *et al.*, 2004), one should adapt Swofford's (1991: 329) "admittedly non-Popperian position that an ambiguous solution that contains the truth is, in many situations, preferable to an unambiguous solution that is wrong." On the other hand, in cases where known phenomena, such as hybridisation or incomplete lineage sorting that results in topology incongruence and/or non-monophyletic taxa that cause the ILD test to fail (*e.g.*, De Villiers *et al.*, 2013), a pragmatic approach could regard only incongruences of key nodes supported by higher than 75% bootstrap values and/or 0.95 posterior probabilities as significant enough to prevent combination of datasets (Nishii *et al.*, 2015) (an example is provided in Fig. 3). As an alternative, samples of known hybrid origin could be excluded from the analysis and the ILD test repeated (*e.g.*, Möller *et al.*, 2020).The ILD test can also be used to interrogate the data quality if performed for sequences originating from different, or the same, genome with uniparental inheritance like those from the mitochondrion or chloroplast where there should be no incongruence (*e.g.*, Ranasinghe *et al.*, 2024). Failure of the ILD test here can point to a technical, rather than biological, issue concerning laboratory procedures, and the samples in question should be checked and, in the best case, sampling and sequencing repeated.

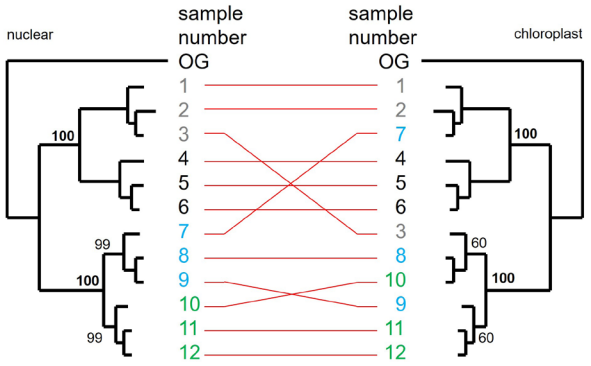


Fig. 3. Example of topology incongruences between nuclear and chloroplast phylogenies, where a combinability test, such as ILD, is likely to result in statistically significant incongruences; the reversed positions of samples 3 and 7 are highly supported true incongruences because of the maximum branch support values for the main clades (bold numbers), and in this case separate phylogenies should be shown; the reversed positions of samples 9 and 10 are not consistently highly supported, only for the nuclear tree (99%; with 60% in the chloroplast tree), and the data could be analysed combined; only relevant branch support values are shown.

6. Presentation of results

The reporting of genetic differences/distances between and within species is useful when multiple samples per species are included in the analysis and can be used to check for genetic discontinuities, *i.e.*, show genetic species limits. The reporting of this parameter just for the new species, however, is not useful as it does not put it in context with its closest phylogenetic relatives and cannot demonstrate a genetic distance between the species, or independent status of a new species (see Fig. 1b,c).

BI and ML analyses have the advantages of providing a single phylogenetic tree which should be displayed to show branch lengths as

average genetic substitutions along the branches (Fig. 4a,b). Similarly, a single MP phylogenetic tree depicted as a phylogram shows the number of genetic changes that occurred along the branches (Fig. 4c-e). MP analyses can result in many equally most-parsimonious trees, and when these hypotheses are combined into a consensus tree, the link to branch lengths are lost (Fig. 4d). In this case the display of one of the phylograms and the consensus tree is recommended. Trees, depicting branch lengths or substitution rates, illustrate clade depths, important for the visualisation of the distribution of genetic diversity across a tree, and allows a quick overview of the genetic distinctness and limits of species including the distinctness of a new species. This is not possible with trees depicted as cladograms as is usual for consensus trees which are often seen in published work (consensus trees can be 50% majority rule consensus trees, depicting only branches that occur in 50% or more of all most parsimonious trees, or a strict consensus tree where only those branches are depicted that are in all most parsimonious trees; Fig. 4f,g). Phylograms can also be informative of data or alignment quality, where poor sequence data or misaligned sequences will appear with excessively long branches (*e.g.*, Fig. 2).

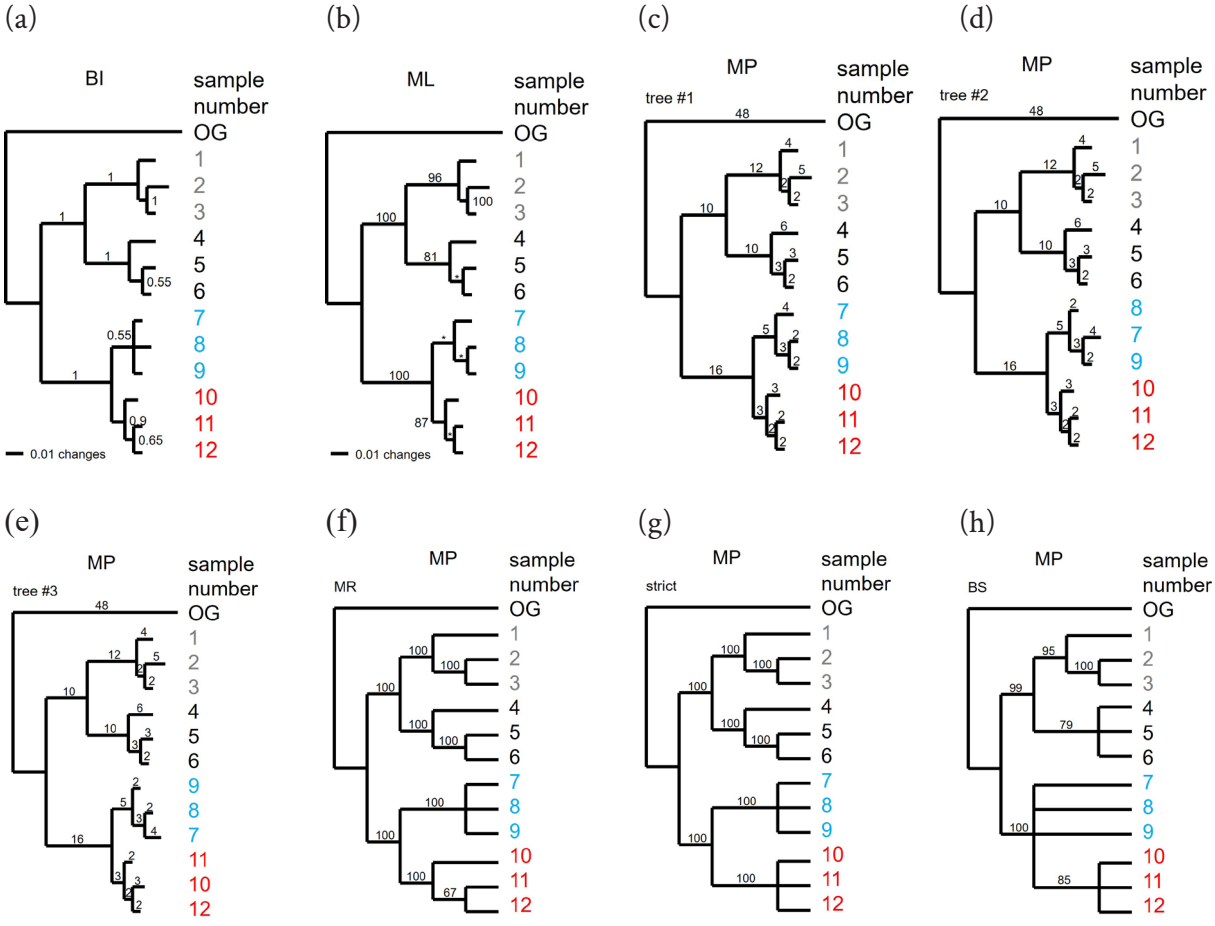


Fig. 4. Example of results for the same data from Bayesian Inference (BI), Maximum Likelihood (ML), and Maximum Parsimony (MP) analyses: **a.** 50% majority rule consensus BI tree with average branch lengths and posterior probabilities along the branches. Note the polytomy, absent in the ML analysis; **b.** Most likely ML tree depicted as phylogram with bootstrap values, note the fully resolved tree topology despite polytomies in the MP analysis (f, g), where the ML tree has branch support values below 50% (indicated with asterisks). These branches are likely artefacts and not real; **c**–**e.** Phylograms of three equally MP trees; **f.** their 50% majority rule consensus tree; **g.** their strict consensus tree. Note the absence of proportional branch lengths in (**f**) and (**g**); **h.** Parsimony bootstrap tree. Note the topology of the bootstrap tree is different from any other MP tree. Note also that a bootstrap tree is not a phylogenetic tree. Note further that trees (**f**–**h**) are depicted as cladograms because they represent summary trees of several to many individual trees, and the link to character changes along the branches are lost.

7. Examples of interpretation of results of morphological-molecular phylogenetic species delimitation

Example 1, complete species delimitation: Samples 1–9 represent existing species I (white flowers and rosette habit), II (white flowers and caulescent habit) and III (blue flowers and caulescent habit), and samples 10–12 represent a new species IV (red flowers and caulescent habit). In this example, full congruence between molecular phylogenetic tree topology and morphological characteristics is present and the four species are successfully delimited, with the new species being sister to species III (Fig. 5).

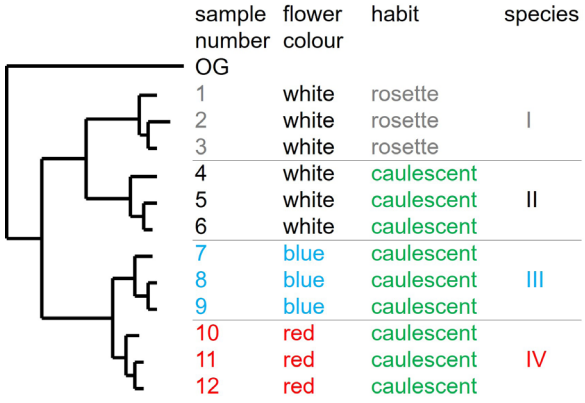


Fig. 5. Example 1 with complete delimitation of species. The samples fall in four clades, each clade comprises samples with unique combinations of characteristics, and four distinct species can be defined.

Example of species delimitation in Henckelia Spreng. (Gesneriaceae) based on ITS sequences: Six samples per species, representing two populations per species were included for analysis. The three species, *Henckelia moonii* (Gardner) D.J.Middleton & Mich.Möller, *H. wijesundarae* Ranasinghe & Mich.Möller, and *Henckelia walkerae* (Gardner) D.J.Middleton & Mich.Möller, each formed a clade in a larger phylogeny of all Sri Lankan *Henckelia* species (Ranasinghe, 2017). Based on this analysis, the six samples of *H. moonii* were selected as outgroup and root. The new species, *H. wijesundarae*, was originally recognized by Thwaites (1864) as *Chirita walkerae* var. *ß*. It was later described by Clarke (1883) as *Chirita walkerae* var. *parviflora* C.B.Clarke, then changed by Theobald & Grupe (1972) to *Chirita walkerae* subsp. *parviflora* (C.B.Clarke) Theob. & Grupe. A few years ago, it was synonymised by Weber *et al.* (2011a) under *Henckelia walkerae* and eventually recognised by Ranasinghe *et al.* (2016) as *Henckelia wijesundarae*. The last step was based on their morphological-molecular phylogenetic study of freshly collected material of 2–4 samples per population from six populations of three species. The analysis showed that the new species consistently differs from its sister species, *Henckelia walkerae*, in at least six qualitative morphological characters (Fig. 6).

Example 2, polymorphic species: Samples of a new species (red-flowered) are scattered among samples of an existing one, species III (blue-flowered) (Fig. 7). Incongruence between the molecular phylogenetic tree and morphological characters exist, seen as variable flower colour for species III including the newly collected samples. Two possible scenarios can be envisaged: either one species polymorphic for flower colour is accepted, or, since the two clades of species III are not that close genetically, additional morphological characters could be sought. In the second scenario, additional morphological characters, *e.g.*, leaf sclereids (absent/present) and pollen (single/tetrads), support the molecular topology and a redefinition of species III results in two newly defined species, III and IV, with polymorphic flower colours, but consistent leaf sclereids and pollen characteristics (Fig. 7b). In this scenario, our approach provides immediately two hard, qualitative, diagnostic characters, leaf sclereids and pollen, and such consistent characters could be used straight away in diagnoses (*cf.* Nishii *et al.*, 2015). Qualitative characters, such as size, shape, or degree of fusion of organs, may be unsuitable for taxon delimitation, because of their difficulty in scoring, high levels of variation and homoplasy, and the resultant scattered distribution of states across a phylogenetic tree (*e.g.*, Atkins *et al.*, 2021).

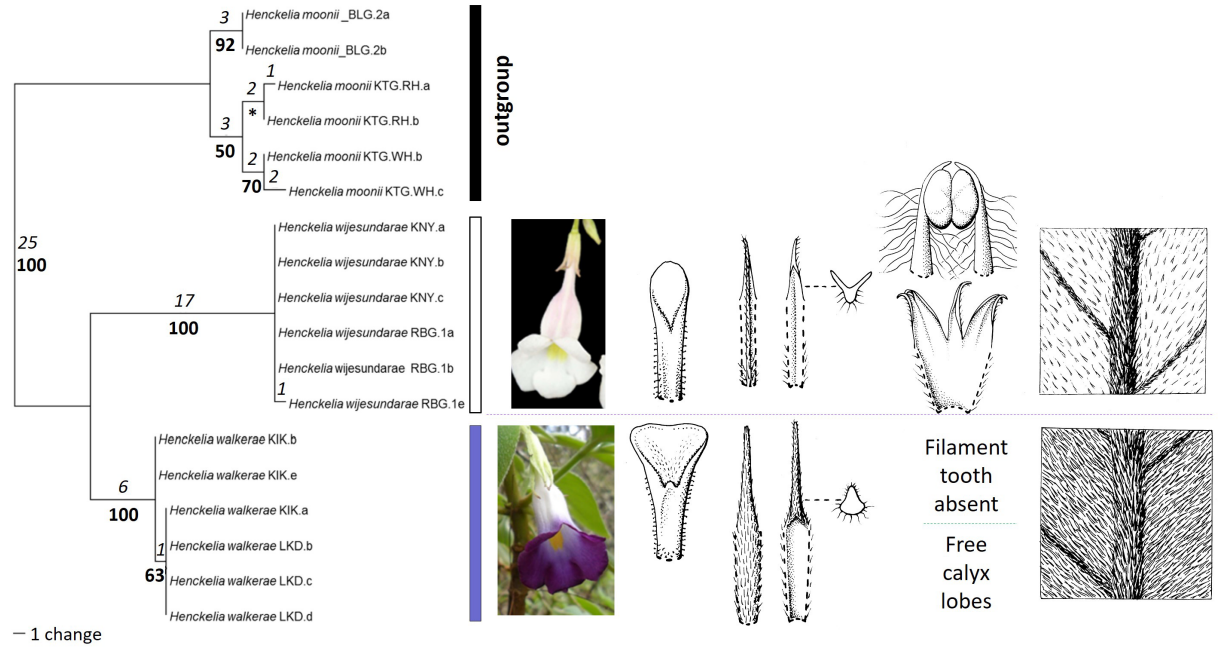


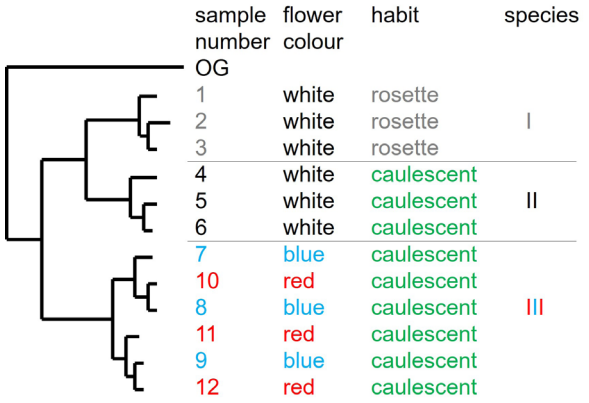
Fig. 6. Example of delimitation of three species in *Henckelia* Spreng. (Gesneriaceae) based on ITS sequences and sampling of six samples from two populations per species, and characters (stigma, sepals, calyx, filaments, abaxial leaf surface) separating the two species *Henckelia walkerae* (Gardner) D.J.Middleton & Mich.Möller and *H. wijesundarae* Ranasinghe & Mich.Möller. Numbers above and below branches are the number of nucleotide changes and bootstrap values respectively. Modified from Ranasinghe *et al.* (2016)..

If incongruence between molecular tree topology and morphological data persists after adding additional morphological characters (Fig. 7c), perhaps it represents a case of incomplete lineage sorting or involves hybridisation, and irrespective of methods/data used to address the problem (for molecular data often seen as discordance/incongruence in tree topologies between different genome markers, or multiple nuclear markers), the issue is not solvable but results in a polymorphic species (Knowles & Carstens, 2007; Nopporncharoenkul *et al.*, 2016). These plants may have often a complex evolutionary history which cannot be easily solved, except with much larger sampling and multi-genomic data (*e.g.*, Joly *et al.*, 2009; Kong *et al.*, 2023, and references therein). Studies of tropical rainforest taxa, such as *Andira* Lam. have recovered examples of non-monophyly where widespread dominant tropical rainforest taxa have morphologically distinct, often range-restricted taxa nested within them (Pennington & Lavin, 2016) which do not necessarily require taxonomic redefinition as they are seemingly examples of incipient speciation, with their progenitor species showing incomplete lineage sorting. The authors advocate "that using a coalescent framework for species delimitation may result in better agreement with traditional morphological delimitations than methods based on genetic exclusivity criteria."

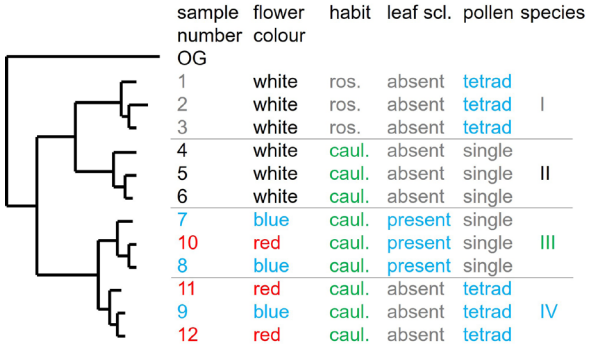
In Gesneriaceae, a highly complex case of non-monophyly seems to be represented by the genus *Raphiocarpus* Chun, as the species fall in several clades of subtribe Didymocarpinae (Möller *et al.*, 2011a). This does not seem to represent a case of incipient speciation as the clades have a significant genetic depth (Möller *et al.*, 2011a), and is not based on too few markers (ITS and *trn*LF), as an extensive NGS study of several hundred markers shows a similar pattern of non-monophyly (Yang *et al.*, 2023). This is more likely representing a case similar to *Chirita*, where the species are grouped by a plesiomorphic characteristic (Weber *et al.*, 2011a), which is uninformative for a phylogeny-based classification. In this case, only an in-depth sampling across the distribution range of the genera involved, *i.e.*, southern China and particularly the

presently relatively poorly sampled Vietnam would be necessary to resolve this case.

(a)



(b)



(c)

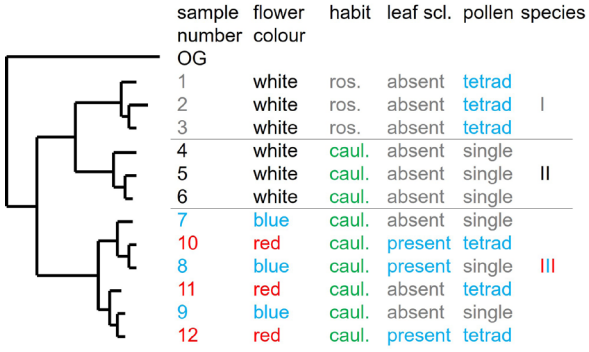
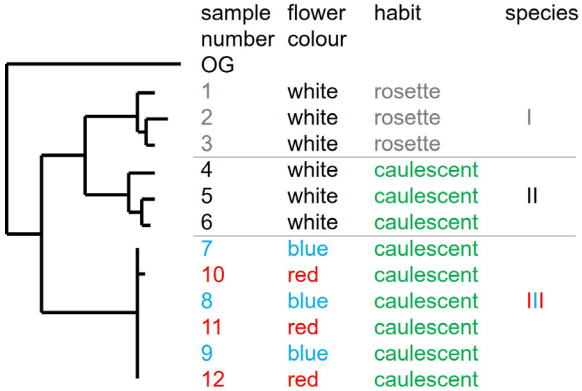


Fig. 7. Example 2, polymorphic species: **a.** Samples of a new species (samples 10-12) scattered among samples of an existing species (samples 7-9); **b.** Additional characters (leaf sclereids: absent/present; pollen: single/tetrads) indicate the existence of two species with polymorphic flower colours; **c.** Or confirm the existence of one species with complex character combination (see text for further explanations).

Example 3, species in a polytomy: In this example, a new species has identical or very similar sequence data as the samples of the blue-flowered species, and thus the phylogeny involves a polytomy (Fig. 8a). If the polytomy is only present in a parsimony consensus tree, *i.e.*, represents a soft polytomy, perhaps applying another phylogenetic method, *e.g.*, BI or ML may resolve the polytomy. If a hard polytomy exists, *i.e.*, in the absence of parsimony informative sequence differences between the samples 7–12, perhaps the acquisition of additional, faster evolving molecular markers can resolve the polytomy. In the example here, the new species IV is successfully delineated from species III after adding more molecular markers (Fig. 8b).

(a)



(b)

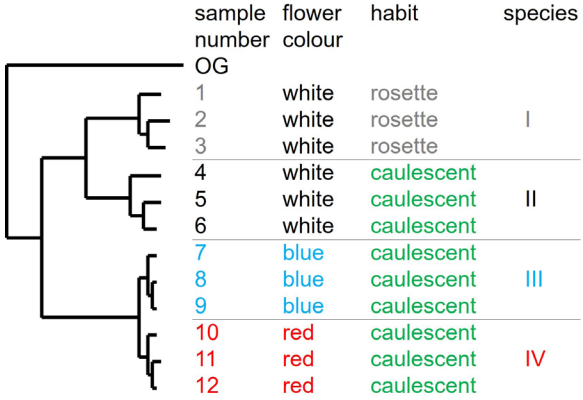
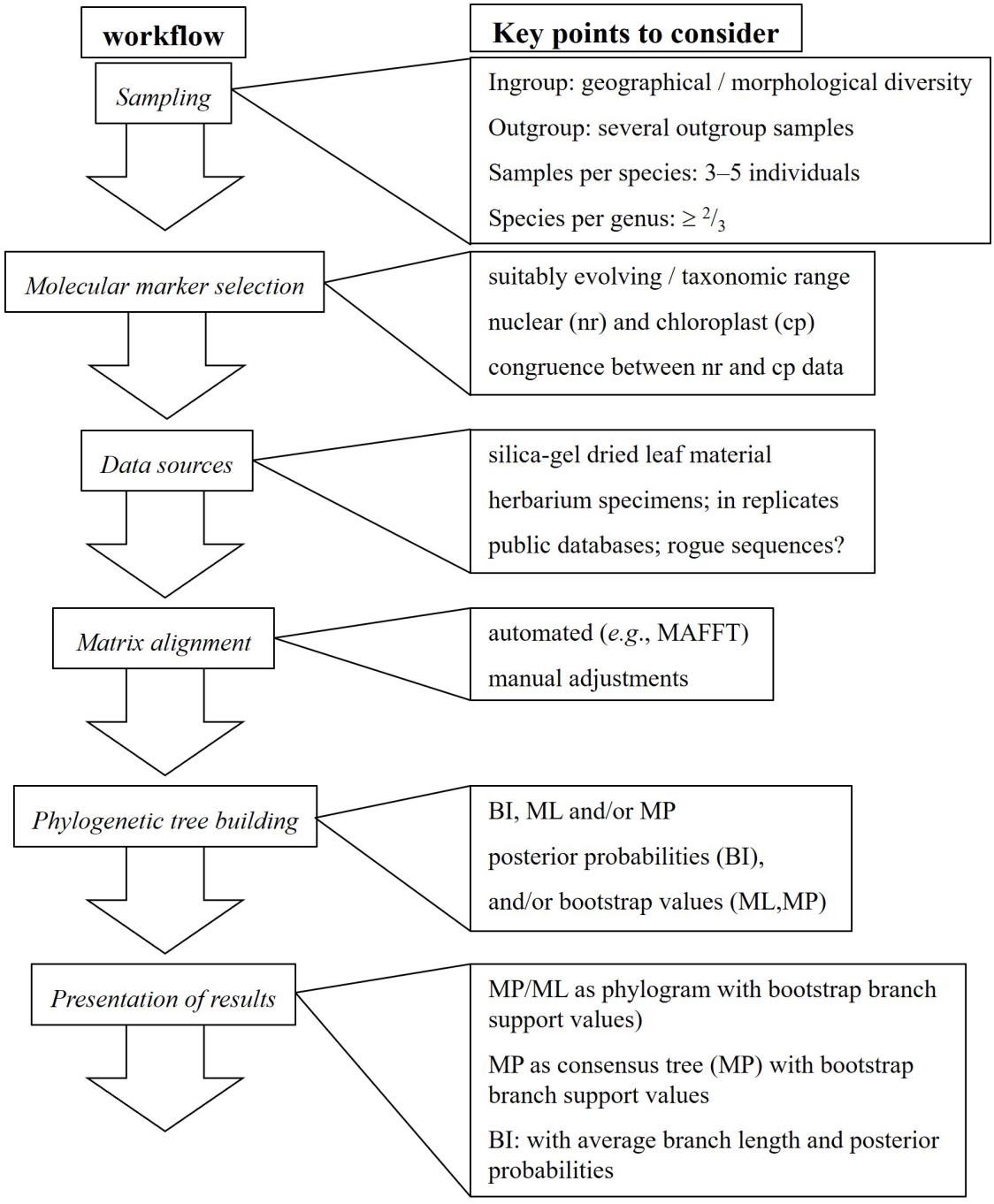


Fig. 8. Example 3, species in a polytomy: **a.** Samples of a new species and an existing species have very similar or identical sequences and fall in one polytomy; **b.** After the addition of data from an additional marker/s, the two species are resolved as two separate species.

8. Workflow

For convenience a workflow for morphological-molecular phylogenetic species delimitation is assembled showing the main steps and key points to consider:



Concluding thoughts

The above recommendations are designed to encourage taxonomists to use molecular data most effectively by applying a few common-sense provisions to effectively delimit new taxa (species or genera). Once a sufficiently large baseline data set is assembled based on multiple samples per species, the efforts required for the addition of new ones is relatively small and would require the gathering and inclusion of multiple samples of the new species. The congruence of morphological characteristics obtained from these samples with the resulting tree topologies will further support the protologue descriptions by focussing on key characters and strengthen the resulting taxonomic decisions. It is hoped that the above pragmatic approach will be helpful to guide and encourage future molecular-morphological taxonomic work.

Acknowledgements

The Royal Botanic Garden Edinburgh (RBGE) is supported by the Scottish Government's Rural and Environmental Science and Analytical Services Division. We are grateful to Professor Pete Hollingsworth (Director of Science) for access to RBGE facilities. We like to thank Dr Subhani Ranasinghe, then Peradeniya Botanical Garden, Sri Lanka, for organising a species delimitation workshop that inspired us to prepare guidelines that presented the basis for the recommendations given here. We also acknowledge constructive comments of two anonymous reviewers on an earlier version of the manuscript.

Literature Cited

ATKINS H.J., BRAMLEY G.L.C., NISHII K., MÖLLER M., OLIVAR J.E.C., KARTONEGORO A. & M. HUGHES 2021. Sectional polyphyly and morphological homoplasy in Southeast Asian *Cyrtandra* (Gesneriaceae): consequences for the taxonomy of a mega-diverse genus. *Plant Systematics and Evolution* 307: 60. https://doi.org/10.1007/s00606-021-01784-x

BARKER F.K. & F.M. LUTZONI 2002. The utility of the incongruence length difference test. *Systematic Biology* 51(4): 625–637. https://doi.org/10.1080/10635150290102302

BAUM D.A. & S.D. SMITH 2012. Tree thinking: an introduction to phylogenetic biology. Roberts and Co., Greenwood Village, CO.

BOLUDA C.G., CHRISTE C., NACIRI Y., & L. GAUTIER 2022. A 638-gene phylogeny supports the recognition of twice as many species in the Malagasy endemic genus *Capurodendron* (Sapotaceae). *Taxon* 71(2): 360–395. https://doi.org/10.1002/tax.12676

CHEN W.H., MÖLLER M., SHUI Y.M., WANG H., YANG J.B. & G.Y. LI 2014. Three new species of *Petrocodon* (Gesneriaceae), endemic to the limestone areas of Southwest China, and preliminary insights into the diversification patterns of the genus. *Systematic Botany* 39(1): 316–330. https://doi.org/10.1600/036364414X678189

CLARKE C.B. 1883. Cyrtandreae. *In*: DE CANDOLLE A. & C. DE CANDOLLE (eds.), *Monographiae Phanerogamarum*. Volume 5/1. Masson G., Paris, pp. 1–303.

CUNNINGHAM C.W. 1997. Can three incongruence tests predict when data should be combined? *Molecular Biology and Evolution* 14: 733–740. https://doi.org/10.1093/oxfordjournals.molbev.a025813

DE QUEIROZ 2007. Species concepts and species delimitation. *Systematic Biology* 56(6): 879–886. https://doi.org/10.1080/10635150701701083

DE VILLIERS M.J., PIRIE M.D., HUGHES M., MÖLLER M., EDWARDS T.J. & D.U.B. BELLSTEDT 2013. An approach to identify putative hybrids in the 'coalescent stochasticity zone', as exemplified in the African plant genus *Streptocarpus* (Gesneriaceae). *New Phytologist* 198(1): 284–300. https://doi.org/10.1111/nph.12133

DENDUANGBORIPANT J. & Q.C.B. CRONK 2000. High intraindividual variation in internal transcribed spacer sequences in *Aeschynanthus* (Gesneriaceae): implications for phylogenetics. *Proceedings of the Royal Society B, Biological Sciences* 267(1451): 1407–1415. https://doi.org/10.1098/rspb.2000.1157

DOUZERY E.J.P., PRIDGEON A.M., KORES P., LINDER H.P., KURZWEIL H. & M.W. CHASE 1999. Molecular phylogenetics of *Diseae* (Orchidaceae): a contribution from nuclear ribosomal ITS sequences. *American Journal of Botany* 86(6): 887–899. https://doi.org/10.2307/2656709

DOYLE J.J. & J.L. DOYLE 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

DROUIN G., DAOUD H. & J. XIA 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Molecular Phylogenetics and Evolution* 49(3): 827–831. https://doi.org/10.1016/j.ympev.2008.09.009

DUMINIL J. & M. DI MICHELE 2009. Plant species delimitation: a comparison of morphological and molecular markers. *Plant Biosystems* 143(3): 528–542. https://doi.org/10.1080/11263500902722964

FARRIS J., KÄLLERSJÖ S.M., KLUGE A.G. & C. BULT 1995a. Constructing a significance test for incongruence. *Systematic Biology* 44(4): 570–572. https://doi.org/10.2307/2413663

FARRIS J., KÄLLERSJÖ S.M., KLUGE A.G. & C. BULT 1995b. Testing significance of incongruence. *Cladistics* 10: 315–319. https://doi.org/10.1111/j.1096-0031.1994.tb00181.x

FELSENSTEIN J. 2004. Inferring phylogenies. Sinauer Associates, Sunderland, Massachusetts.

FERRARI G., ESSELENS L., HART M.L., JANSSENS S., KIDNER C., MASCARELLO M., PEÑALBA J.V., PEZZINI F., VON RINTELEN T., SONET

G., VANGESTEL C., VIRGILIO M. & P.M. HOLLINGSWORTH 2023. Developing the protocol infrastructure for DNA sequencing natural history collections. *Biodiversity Data Journal* 11: e102317. https://doi.org/10.3897/BDJ.11.e102317

GREINER S., SOBANSKI J. & R. BOCK 2015. Why are most organelle genomes transmitted maternally? *BioEssays* 37: 80–94. https://doi.org/10.1002/bies.201400110

GROSSER M.R., SITES S.K., MURATA M.M., LOPEZ Y., CHAMUSCO K.C., LOVE HARRIAGE K., GROSSER J.W., GRAHAM J.H., GMITTER F.G. JR. & C.D. CHASE 2023. Plant mitochondrial introns as genetic markers - conservation and variation. *Frontiers in Plant Sciences* 14: 1116851. https://doi.org/10.3389/fpls.2023.1116851

GUO J., PAN B., LIU J., XU W.B. & K.F. CHUNG 2015. Three new species of *Primulina* (Gesneriaceae) from limestone karsts of China based on morphological and molecular evidence. *Botanical Studies* 56(34): 1–13. https://doi.org/10.1186/s40529-015-0115-5

HALE M.L., BURG T.M. & T.E. STEEVES 2012. Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. *PLoS ONE* 7: e45170. https://doi.org/10.1371/journal.pone.0045170

HANCE H.F. 1883. New Chinese Cyrtandreae. *Journal of Botany* 21: 165–170.

HASSANPOUR S., MOAZZENI H., SONBOLI A., HOSEINI S., PIRANI A., JOHARCHI M.R. & C. OBERPRIELER 2023. Molecular and morphological data reveal a new species of *Sclerorhachis* (Compositae, Anthemideae) and the reassessment of another species of the genus. *Plant Systematics and Evolution* 309: 10. https://doi.org/10.1007/s00606-022-01840-0

HIPP A.L., HALL J.C. & K.J. SYTSMA 2004. Congruence versus phylogenetic accuracy: revisiting the incongruence length difference test. *Systematic Biology* 53(1): 81–89. https://doi.org/10.1080/10635150490264752

HONG X., KEENE J., QIU Z.J. & F. WEN 2019. *Primulina anisocymosa* (Gesneriaceae), a new species with a unique inflorescence structure from Guangdong, China. *PeerJ* 7: e6157. https://doi.org/10.7717/peerj.6157

HU J., ZHANG J.Y., HE H., YU D.X., JIANG H., LIU Q. & F. WEN 2023. *Oreocharis oriolus*, a new species of Gesneriaceae in a sclerophyllous oak community from Yunnan, Southwest China. *Ecology and Evolution* 13(6): e10174. https://doi.org/10.1002/ece3.10174

HUANG J., XIANG X.G., LU Y.B., PAN B. & Q. ZHANG 2017. *Hemiboea pterocaulis* comb. et stat. nov. (Gesneriaceae), a new critically endangered species segregated from *H. subcapitata*. *Nordic Journal of Botany* 2018: e01468. https://doi.org/10.1111/njb.01468

JANEESHA A.P., RANASINGHE S., MÖLLER M., DILNA P.K. & S. NAMPY 2023. Revision of *Rhynchoglossum* (Gesneriaceae) in South India and Sri Lanka, reinstating *R. scabrum* based on a multi-sample molecular-morphological approach. *Plant Systematics and Evolution* 309: 26. https://doi.org/10.1007/s00606-023-01858-y

JOLY S., MCLENACHAN P.A. & P.J. LOCKHART 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. *American Naturalist* 174(2): E54–E70. https://doi.org/10.1086/600082

KANTHRAJ A.S., RANA T.S., TIWARI A. & K.N. NAIR 2023. Taxonomic reappraisal of *Henckelia hookeri* and *H. collegii-sancti-thomasii* (Gesneriaceae): two little-known endemic species of the Eastern Himalaya in India. *Kew Bulletin* 78: 625–640. https://doi.org/10.1007/s12225-023-10090-3

KATOH K., ROZEWICKI J. & K.D. YAMADA 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* 20(4): 1160–1166. https://doi.org/10.1093/bib/bbx108

KIM J.Y., ROHLF F.J. & R.R. SOKAL 1993. Accuracy of phylogenetic estimation using the Neighbor-Joining method. *Evolution* 47(2): 471–486. https://doi.org/10.2307/2410065

KNOWLES L.L. & B.C. CARSTENS 2007. Delimiting species without monophyletic gene trees. *Systematic Biology* 56(6): 887–895. https://doi.org/10.1080/10635150701701091

KONG H.H., CONDAMINE F.L., YANG L.H., HARRIS A.J., FENG C., WEN F. & M. KANG 2022. Phylogenomic and macroevolutionary evidence for an explosive radiation of a plant genus in the Miocene. *Systematic Biology* 71(3): 589–609. https://doi.org/10.1093/sysbio/syab068

LEMEY P., SALEMI M. & A.M. VANDAMME 2009. The phylogenetic handbook second edition: a practical approach to phylogenetic analysis and hypothesis testing. Cambridge University Press, Cambridge.

LI P.W., LIU F.P., HAN M.Q., SMITH J.F. & Y.Z. WANG 2022. Molecular and morphological evidence supports the inclusion of *Deinostigma* into *Metapetrocosmea* (Gesneriaceae). *Annals of the Missouri Botanical Garden* 107: 447–466. https://doi.org/10.3417/2022767

LIU Y., XU W.B. & Y.S. HUANG 2011. *Primulina guangxiensis* sp. nov. (Gesneriaceae) from a karst cave in Guangxi, China. *Nordic Journal of Botany* 29: 682–686. https://doi.org/10.1111/j.1756-1051.2011.01089.x

LUIKART G. & J. CORNUET 1998. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology* 12(1): 228–237. https://doi.org/10.1111/j.1523-1739.1998.96388.x

LUU H.T., HSIEH C.L., CHUANG C.R., CHEN C.W., TRAN N.T., VU N.L. & K.F. CHUNG 2023. *Langbiangia*, a new genus of Gesneriaceae endemic to Langbiang Plateau, southern Vietnam and a taxonomic endeavor to achieve key targets of the post-2020 global biodiversity framework. *PLoS ONE* 18(5): e0284650. https://doi.org/10.1371/journal.pone.0284650

MCKAIN M.R., JOHNSON M.G., URIBE-CONVERS S. EATON D. & Y. YANG 2018. Practical considerations for plant phylogenomics. *Applications in Plant Sciences* 6(3): e1038. https://doi.org/10.1002/aps3.1038

MEIER R. & B. WHEELER 2008. DNA sequences in taxonomy, opportunities and challenges. *The New Taxonomy* 65–127.

MEYER C.P. & G. PAULAY 2005. DNA barcoding: error rates based on comprehensive sampling. *PLOS Biology* 3(12): e422 https://doi.org/10.1371/journal.pbio.0030422

MOGENSEN H.L. 1996. The hows and whys of cytoplasmic inheritance in seed plants. *American Journal of Botany* 83: 383–404. https://doi.org/10.1002/j.1537-2197.1996.tb12718.x

MÖLLER M. 2019. Species discovery in time: an example from Gesneriaceae in China. *Guangxi Sciences* 26(1): 1–16. https://doi.org/10.13656/j.cnki.gxkx.20190307.002

MÖLLER M. & Q.C.B. CRONK 1997. Origin and relationships of *Saintpaulia* (Gesneriaceae) based on ribosomal DNA internal transcribed spacer (ITS) sequences. *American Journal of Botany* 84(8): 956–965. https://doi.org/10.2307/2446286

MÖLLER M. & Q.C.B. CRONK 2001. Evolution of morphological novelty: a phylogenetic analysis of growth patterns in *Streptocarpus* (Gesneriaceae). *Evolution* 55(5): 918–929. https://doi.org/10.1111/j.0014-3820.2001.tb00609.x

MÖLLER M. & J.L. CLARK 2013. The state of molecular phylogenetic work in the family Gesneriaceae: a review. *Selbyana* 31(2): 95–125. https://www.jstor.org/stable/24894284

MÖLLER M., KOKUBUGATA G. & K. JONG 2008. Aspects of genome evolution in Gesneriaceae: patterns of 45S-nrDNA site evolution based on evidence from fluorescent *in-situ* hybridization (FISH). *International Journal of Plant Sciences* 169(5): 667–676. https://doi.org/10.1086/533609

MÖLLER M., PFOSSER M., JANG C.G., MAYER V., CLARK A., HOLLINGSWORTH M.L., BARFUSS M.H.J., WANG Y.Z., KIEHN M. & A. WEBER 2009. A preliminary phylogeny of the "didymocarpoid Gesneriaceae" based on three molecular data sets: incongruence with available tribal classifications. *American Journal of Botany* 96(5): 989–1010. https://doi.org/10.3732/ajb.0800291

MÖLLER M., FORREST A., WEI Y.G. & A. WEBER 2011a. A molecular phylogenetic assessment of the advanced Asiatic and Malesian didymocarpoid Gesneriaceae with focus on non-monophyletic and monotypic genera. *Plant Systematics and Evolution* 292: 223–248. https://doi.org/10.1007/s00606-010-0413-z

MÖLLER M., MIDDLETON D.J., NISHII K., WEI Y.G., SONTAG S. & A. WEBER 2011b. A new delineation for *Oreocharis* incorporating an additional ten genera of Chinese Gesneriaceae. *Phytotaxa* 23: 1–36. https://doi.org/10.11646/phytotaxa.23.1.1

MÖLLER M., CHEN W.H., SHUI Y.M., ATKINS H. & D.J. MIDDLETON 2014. A new genus of Gesneriaceae in China and the transfer of *Briggsia* species to other genera. *Gardens' Bulletin Singapore* 66(2): 195–205. https://biostor.org/reference/217698

MÖLLER M., WEI Y.G., WEN F., CLARK J.L. & A. WEBER 2016. You win some you lose some: updated generic delineations and classification of Gesneriaceae – implications for the family in China. *Guihaia* 36(1): 44–60. https://doi.org/10.11931/guihaia.gxzw201512015

MÖLLER M, LIU J., LI Y., LIE J.H., YEA L.J., MILL R., THOMAS P., LI D.Z. & L.M. GAO 2020. Repeated intercontinental migrations and recurring hybridizations characterise the evolutionary history of yew (*Taxus* L.). *Molecular Phylogenetics and Evolution* 153: e106952. https://doi.org/10.1016/j.ympev.2020.106952

NEI M. & S. KUMAR 2000. *Molecular evolution and phylogenetics.* Oxford University Press, Oxford.

NISHII K., HUGHES M., BRIGGS M., HASTON E., CHRISTIE F., DEVILLIERS M.J., HANEKOM T., ROOS W.G., BELLSTEDT D.U. & M. MÖLLER 2015. *Streptocarpus* redefined to include all Afro-Malagasy Gesneriaceae: molecular phylogenies prove congruent with geography and cytology and uncovers remarkable morphological homoplasies. *Taxon* 64(6): 1243–1274. https://doi.org/10.12705/646.8

NOPPORNCHAROENKUL N., SOONTORNCHAINAKSAENG P., JENJITTIKUL T., CHUENBOONNGARM N. & U. VIBOONJUN 2016. *Kaempferia simaoensis* (Zingiberaceae), a new record for Thailand: evidence from nuclear ITS2 sequence analyses. *Thai Journal of Botany* 8: 81–91.

OGUTCEN E., CHRISTE C., NISHII K., SALAMIIN N., MÖLLER M. & M. PERRET 2021. Phylogenomics of Gesneriaceae using targeted capture of nuclear genes. *Molecular Phylogenetics and Evolution* 157: 107068. https://doi.org/10.1016/j.ympev.2021.107068

PAGE R.D.M. & E.C. HOLMES 1998. *Molecular evolution: a phylogenetic approach*. Wiley-Blackwell, Cambridge.

PATON J. 2023. What's in an NOR? Nuclear ribosomal DNA variation in *Aeschynanthus* (Gesneriaceae) – what can be revealed with genome skimming? MSc thesis (unpublished), The University of Edinburgh.

PELSER P.B., ABBOTT R.J., COMES H.P., MILTON J.J., MÖLLER M., LOOSELEY M.E., CRON G.V., BARCELONA J.F., KENNEDY A.H., WATSON L.E., BARONE R., HERNÁNDEZ F. & J.W. KADEREIT 2012. The genetic ghost of an invasion past: colonization and extinction revealed by historical hybridization in *Senecio*. *Molecular Ecology* 21(2): 369–387. https://doi.org/10.1111/j.1365-294X.2011.05399.x

PENNINGTON R.T. & M. LAVIN 2016. The contrasting nature of woody plant species in different neotropical forest biomes reflects differences in ecological stability. *New Phytologist* 210(1): 25–37. https://doi.org/10.1111/nph.13724

PUGLISI C., WEI Y.G., NISHII K. & M. MÖLLER 2011. *Oreocharis* × *heterandra* (Gesneriaceae): a natural hybrid from the Shengtangshan Mountains, Guangxi, China. *Phytotaxa* 38: 1–18. https://doi.org/10.11646/phytotaxa.38.1.1

PUGLISI C., YAO, T.L., MILNE R., MÖLLER M. & D.J. MIDDLETON 2016. Generic recircumscription in the Loxocarpinae (Gesneriaceae), as inferred by phylogenetic and morphological data. *Taxon* 65(2): 277–292. https://doi.org/10.12705/652.5

QIN H.T., MÖLLER M., MILNE R., LUO Y.H., ZHU G.F., LI D.Z., LIU J. & L.M. GAO 2023. Multiple paternally inherited chloroplast capture events associated with *Taxus* speciation in the Hengduan Mountains. *Molecular Phylogenetics and Evolution* 189: 1–12. https://doi.org/10.1016/j.ympev.2023.107915

RANASINGHE S.W. 2017. Molecular species delimitation, taxonomy and biogeography of Sri Lankan Gesneriaceae. PhD thesis (Unpublished), The University of Edinburgh.

RANASINGHE S.W., MILNE R., JAYASEKARA R., RUBASINGHE S. & M. MÖLLER 2016. *Henckelia wijesundarae* (Gesneriaceae), a new endemic species from Sri Lanka and lectotypification of *Henckelia walkerae* and *Chirita walkerae* var. *parviflora*. *Willdenowia* 46: 213–224. https://doi.org/10.3372/wi.46.46202

RANASINGHE S.W., NISHII K., MÖLLER M., ATKINS H.J., CLARK J.L., PERRET M. KARTONEGORO A., GAO L.M., MIDDLETON D.J. & R.I. MILNE 2024. Biogeographic history of the pantropical family Gesneriaceae with a focus on the Indian plate and diversification through the Old World. *Frontiers of Biogeography* 16(2): e61637. https://doi.org/10.21425/F5FBG61637

ROSE J.P., TOLEDO C.A.P., MORIARTY LEMMON E., LEMMON A.R. & K.J. SYSTMA 2020. Out of sight, out of mind: widespread nuclear and plastid-nuclear discordance in the flowering plant genus *Polemonium* (Polemoniaceae) suggests widespread historical gene flow despite limited nuclear signal. *Systematic Biology* 70(1): 162–180. https://doi.org/10.1093/sysbio/syaa049

SÄRKINEN T., STAATS M., RICHARDSON J.E., COWAN R.S. & F.T. BAKKER 2012. How to open the treasure chest? Optimising DNA extraction from herbarium specimens. *PLoS One* 7(8): e43808. https://doi.org/10.1371/journal.pone.0043808

SHAW J., LICKEY E.B., BECK J.T., FARMER S.B., LIU W.S., MILLER J., SIRIPUN K.C., WINDER C.T., SCHILLING E.E. & R.L. SMALL 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92(1): 142–166. https://doi.org/10.3732/ajb.92.1.142

SHAW J., LICKEY E.B., SCHILLING E.E. & R.L. SMALL 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* 94(3): 275–288. https://doi.org/10.3732/ajb.94.3.275

SHAW J., SHAFER H.L., LEONARD O.R., KOVACH M.J., SCHORR M. & A.B. MORRIS 2014. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. *American Journal of Botany* 101(11): 1987–2004. https://doi.org/10.3732/ajb.1400398

SMALL R.L., RYBURN J.A., CRONN R.C., SEELANAN T. & J.F. WENDEL 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany* 85(9): 1301–1315. https://doi.org/10.2307/2446640

SOLTIS D.E. & R.K. KUZOFF 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* 49(4): 727–742. https://doi.org/10.1111/j.1558-5646.1995.tb02309.x

SPANGLER R.E. & R.G. OLMSTEAD 1999. Phylogenetic analysis of Bignoniaceae based on the cpDNA gene sequences *rbc*L and *ndh*F. *Annals of the Missouri Botanical Garden* 86: 33–46. https://doi.org/10.2307/2666216

STOEBE B., HANSMAN S., GOREMYKIN V., KOWALLIK K.V. & W. MARTIN 1999. Proteins encoded in sequenced chloroplast genomes: an overview of gene content, phylogenetic information and endosymbiotic gene transfer to the nucleus. *In*: HOLLINGSWORTH

P.M., BATEMAN R.M. & R.J. GORNALL (eds), *Molecular systematics and plant evolution*. Taylor & Francis, London. pp. 327–352.

SWOFFORD D.L. 1991. When are phylogeny estimates from molecular and morphological data incongrugent? *In*: MIYAMOTO M.M. & J. CRACRAFT (eds.), *Phylogenetic analysis of DNA sequences*. Oxford University Press, Oxford. pp. 295–333.

TAN K., CHEN D.Y., SONG X.Q. & M.X. REN 2023. A new lithophilous species of Gesneriaceae, *Petrocodon rubrostriatus*, from the karst area of South Yunnan, China. *PhytoKeys* 230: 289–299. https://doi.org/10.3897/phytokeys.230.106358

TATENO Y., TAKEZAKI N. & M. NEI 1994. Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. *Molecular Biology and Evolution* 11(2): 261–77. https://doi.org/10.1093/oxfordjournals.molbev.a040108

THEOBALD W.L. & D.A. GRUPE 1972. Critical notes on Ceylon plants: a new name and combinations in the Gesneriaceae. *Ceylon Journal of Science, Biological Sciences* 10: 70–71.

THWAITES G.H.K. 1864. Cyrtandraceae. *In*: *Enumeratio plantarum Zeylaniae*. Dulau & Co., London. pp. 206–208.

TYSZKA A.S., BRETZ E.C., ROBERTSON H.M., WOODCOCK-GIRARD M.D., RAMANAUSKAS K., LARSON D.A., STULL G.W. & J.F. WALKER 2023. Characterizing conflict and congruence of molecular evolution across organellar genome sequences for phylogenetics in land plants. *Frontiers in Plant Sciences, Section Plant Bioinformatics* 14: 2023. https://doi.org/10.3389/fpls.2023.1125107

WANG W.C., ZHANG X.Z., GARCIA S., LEITCH A.R. & A. KOVAŘÍK 2023. Intragenomic rDNA variation - the product of concerted evolution, mutation, or something in between? *Heredity* 131: 179–188. https://doi.org/10.1038/s41437-023-00634-5

WANG Y.Z., MAO R.B., LIU Y. LI J.M., DONG Y., LI Z.Y. & J.F. SMITH 2011. Phylogenetic reconstruction of *Chirita* and allies (Gesneriaceae) with taxonomic treatments. *Journal of Systematics and Evolution* 49(1): 50–64. https://doi.org/10.1111/j.1759-6831.2010.00113.x

WARD S.M. & M. JASIENIUK 2009. Review: sampling weedy and invasive plant populations for genetic diversity analysis. *Weed Science* 57(6): 593–602. https://doi.org/10.1614/WS-09-082.1

WEBER A., MIDDLETON D.J., FORREST A., KIEW R., LIM C.L., RAFIDAH A.R., SONTAG S., TRIBOUN P., WEI Y.G., YAO T.L. & M. MÖLLER 2011a. Molecular systematics and remodelling of *Chirita* and associated genera (Gesneriaceae). *Taxon* 60(3): 767–790. https://doi.org/10.1002/tax.603012

WEBER A., WEI Y.G., PUGLISI C., WEN F., MAYER V. & M. MÖLLER 2011b. A new definition of the genus *Petrocodon* (Gesneriaceae). *Phytotaxa* 23: 49–67. https://doi.org/10.11646/phytotaxa.23.1.3

WEN F., LI S., XIN Z.B., FU L.F., HONG X., CAI L., QIN J.Q., PAN B., PAN F.Z. & Y.G. WEI 2019. The updated plant list of Gesneriaceae in China under the new Chinese naming rules. *Guangxi Science* 26(1): 37–63.

WEN F., WEI Y.G., FU L.F., XIN Z.B., LI S., HUANG Z.J., GE Y.Z. & D.C. MENG 2021. The checklist of Gesneriaceae in China. http://gccc.gxib.cn/cn/about-68.aspx

WEN F., XIN Z.B., HONG X., CAI L., CHEN X.Y., LIANG J.J., WANG H.F., MACIEJEWSKI S., WEI Y.G. & L.F. FU 2022. *Actinostephanus* (Gesneriaceae), a new genus and species from Guangdong, South China. *PhytoKeys* 193: 89–106. https://doi.org/10.3897/phytokeys.193.80715

WHITE T.J., BURNS T.D., LEE S.B. & J.W. TAYLOR 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In*: INNIS M.A., GELFAND D.H., SNINSKY J.J. & T.J. WHITE (eds.), *PCR protocols*. Academic Press, London. pp. 315–322.

WOLFE K.H., LI W.H. & P.M. SHARP 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences USA* 84(24): 9054–9058. https://doi.org/10.1073/pnas.84.24.9054

WORTLEY A.H., RUDALL P.J., HARRIS D.J. & R.W. SCOTLAND 2005. How much data are needed to resolve a difficult phylogeny? Case study in Lamiales. *Systematic Biology* 54(5): 697–709. https://doi.org/10.1080/10635150500221028

XIAO L.Q., MÖLLER M. & H. ZHU 2010. High nrDNA ITS polymorphism in the ancient extant seed plant *Cycas*: incomplete concerted evolution and the origin of pseudogenes. *Molecular Phylogenetics and Evolution* 55(1): 168–177. https://doi.org/10.1016/j.ympev.2009.11.020

XU G.L., LIANG L.F., CHEN D.Y. JING Z.F., ZUO X.H., ZUO Z.Y. & F. WEN 2023. *Primulina jiulianshanensis*, a new species of Gesneriaceae from Jiangxi province, China. *PhytoKeys* 226: 1–16. https://doi.org/10.3897/phytokeys.226.96351

XU W.B., LIU Y., KONO Y., CHANG H., PENG C.I. & K.F. CHUNG 2013. *Primulina cardaminifolia* (Gesneriaceae), a rare new species from limestone areas in Guangxi, China. *Botanical Studies* 54: 19. https://doi.org/10.1186/1999-3110-54-19

XU W.B., MENG T., ZHANG Q., WU W.H., LIU Y. & K.F. CHUNG 2014. *Petrocodon* (Gesneriaceae) in the limestone karsts of Guangxi, China: three new species and a new combination based on morphological and molecular evidence. *Systematic Botany* 39(3): 965–974. https://doi.org/10.1600/036364414X681437

YANG L.H., HARRIS A.J., WEN F., LI Z., FENG C., KONG H.H. & M. KANG 2023. Phylogenomic analyses reveal an allopolyploid origin of core Didymocarpinae (Gesneriaceae) followed by rapid radiation. *Systematic Biology* 72(5): 1064–1083. https://doi.org/10.1093/sysbio/syad029

YAO P.C., GAO H.Y., WEI Y.N., ZHANG J.H., CHEN X.Y. & H.Q. LI 2017. Evaluating sampling strategy for DNA barcoding study of coastal and inland halo-tolerant Poaceae and Chenopodiaceae: a case study for increased sample size. *PLoS ONE* 12(9): e0185311. https://doi.org/10.1371/journal.pone.0185311

YU X.L., LI M., ZHOU J.J. & P.W. LI 2015. *Petrocodon hunanensis* (Gesneriaceae), a new species identified by both morphological and molecular evidence from limestone area in Hunan, China. *Phytotaxa* 195(1): 65–72. https://dx.doi.org/10.11646/phytotaxa.195.1.4

ZHOU P., LI J. & M. MÖLLER 2017. Secondary contact, hybridization and polyploidization add to the biodiversity in the Hengduan Mountains, exemplified by the widespread *Corallodiscus lanuginosus* (Gesneriaceae). *Plant Systematics and Evolution* 303(5): 587–602. https://doi.org/10.1007/s00606-017-1392-0

ZICH F.A. & B. GRAY 2021. *Boea resupinata* Zich & B.Gray (Gesneriaceae), a new species from Cape York Peninsula, Queensland, Australia. *Austrobaileya* 11: 56–66. Available from: https://www.qld.gov.au/__data/assets/pdf_file/0017/209600/zich-and-gray-boea-resupinata-austrobaileya-v11-56-66.pdf (Accessed on 16.09.2024).