

Extended Lissamphibia: a tale of character non-independence, analytical parameters and islands of trees

Ana Serra Silva

To cite this article: Ana Serra Silva (2024) Extended Lissamphibia: a tale of character non-independence, analytical parameters and islands of trees, Journal of Systematic Palaeontology, 22:1, 2321620, DOI: [10.1080/14772019.2024.2321620](https://doi.org/10.1080/14772019.2024.2321620)

To link to this article: <https://doi.org/10.1080/14772019.2024.2321620>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



View supplementary material [↗](#)



Published online: 02 Apr 2024.



Submit your article to this journal [↗](#)




View related articles [↗](#)



View Crossmark data [↗](#)



Extended Lissamphibia: a tale of character non-independence, analytical parameters and islands of trees

Ana Serra Silva^{a,b,*#} 

^aDepartment of Life Sciences, Natural History Museum, London, SW7 5BD, UK; ^bSchool of Earth Sciences, University of Bristol, Bristol, BS8 1RL, UK

(Received 15 June 2023; accepted 18 February 2024)

The age, content and inter-order relationships of crown Lissamphibia remain a debated topic in vertebrate systematics. Recent phylogenetic analyses of fossil amphibians were used to propose an extended Lissamphibia, with Anura and Caudata nested in Dissorophoidea and with Gymnophiona nested in Stereospondyli, but this hypothesis was not supported by subsequent studies on updated matrices. In a parsimony context, the extended Lissamphibia hypothesis was shown to result from the effects of large island bias on the majority-rule consensus, which masked the presence of topologies supporting the restricted Lissamphibia hypothesis, with all extant orders nested in Dissorophoidea or in Stereospondyli. Re-analysing this dataset, taking into account the presence of inapplicable and polymorphic character states and revising the scores for logically non-independent characters, shows that the phylogenies inferred from the morphological data matrix used to propose the extended Lissamphibia hypothesis are not robust to changes in analytical parameters and that great care should be taken when analysing fossil amphibian datasets. With the set of most parsimonious trees inferred from the unrevised matrix used to propose the extended Lissamphibia hypothesis, I also demonstrate that the phenomenon of large island bias extends to phylogenetic networks, but not to topology-based tests of taxonomic instability that do not rely on split-frequencies.

Keywords: Lissamphibia; taxonomic instability; large island bias; phylogenetic networks; character non-independence; inapplicable data

Introduction

The systematics of Amphibia Gray, 1825 remains a topic of much debate for both neontologists and palaeontologists (e.g. Anderson et al., 2008; Hime et al., 2021; Pardo et al., 2017; Simões et al., 2023; Siu-Ting et al., 2019). While there are still open questions regarding the intra-order relationships of salamander and frog lineages, subclass-wide molecular analyses predominantly support a monophyletic Lissamphibia Haeckel, 1866 (the clade comprising all descendants of the last common ancestor of the extant orders Gymnophiona Müller, 1832, Caudata Fisher von Waldheim, 1813 and Anura Fisher von Waldheim, 1813 and the extinct family Albanerpetidae Fox & Naylor, 1982) under the Batrachia Latreille, 1800 hypothesis of extant lissamphibian relationships (Fig. 1A), which places Gymnophiona as sister to Caudata + Anura (e.g. Frost et al., 2006; Pyron & Wiens, 2011). There is, however, evidence of multiple gene trees supporting the Procera Feller and Hedges, 1998 (frogs as sister to other lissamphibians) and Acauda *sensu* Hime et al., 2021 (salamanders sister

to caecilians and frogs) hypotheses (e.g. Hime et al., 2021; Siu-Ting et al., 2019), evidencing the presence of a complex genome-level evolutionary history. In a palaeontological context, while many phylogenetic studies support the Batrachia hypothesis they disagree on the placement of caecilians in relation to extinct fossil lineages, including albanerpetids (e.g. Gardner, 2001; Kligman et al., 2023), which has implications for the taxonomic content and age of Lissamphibia (e.g. Anderson et al., 2008; Pardo et al., 2017; Ruta & Coates, 2007; Schoch et al., 2020).

Pardo et al. (2017) described and attempted to phylogenetically place a new fossil temnospondyl from the Triassic of North America, *Chinlestegophis jenkinsi*. Their Bayesian analyses found strong support for a close relationship between *Chinlestegophis* and extant Gymnophiona, but a distant relationship between Gymnophiona and Batrachia (albanerpetids were not sampled), which combined with the congruence between the majority-rule consensus (MRC; Margush & McMorris, 1981) of their parsimony and Bayesian analyses, led them to propose an extended Lissamphibia

*Corresponding author. Email: a.da-silva@nhm.ac.uk, ana.silva.17@ucl.ac.uk

#Present address: Department of Genetics, Evolution and Environment, University College London, London, WC1E 6BT, UK.

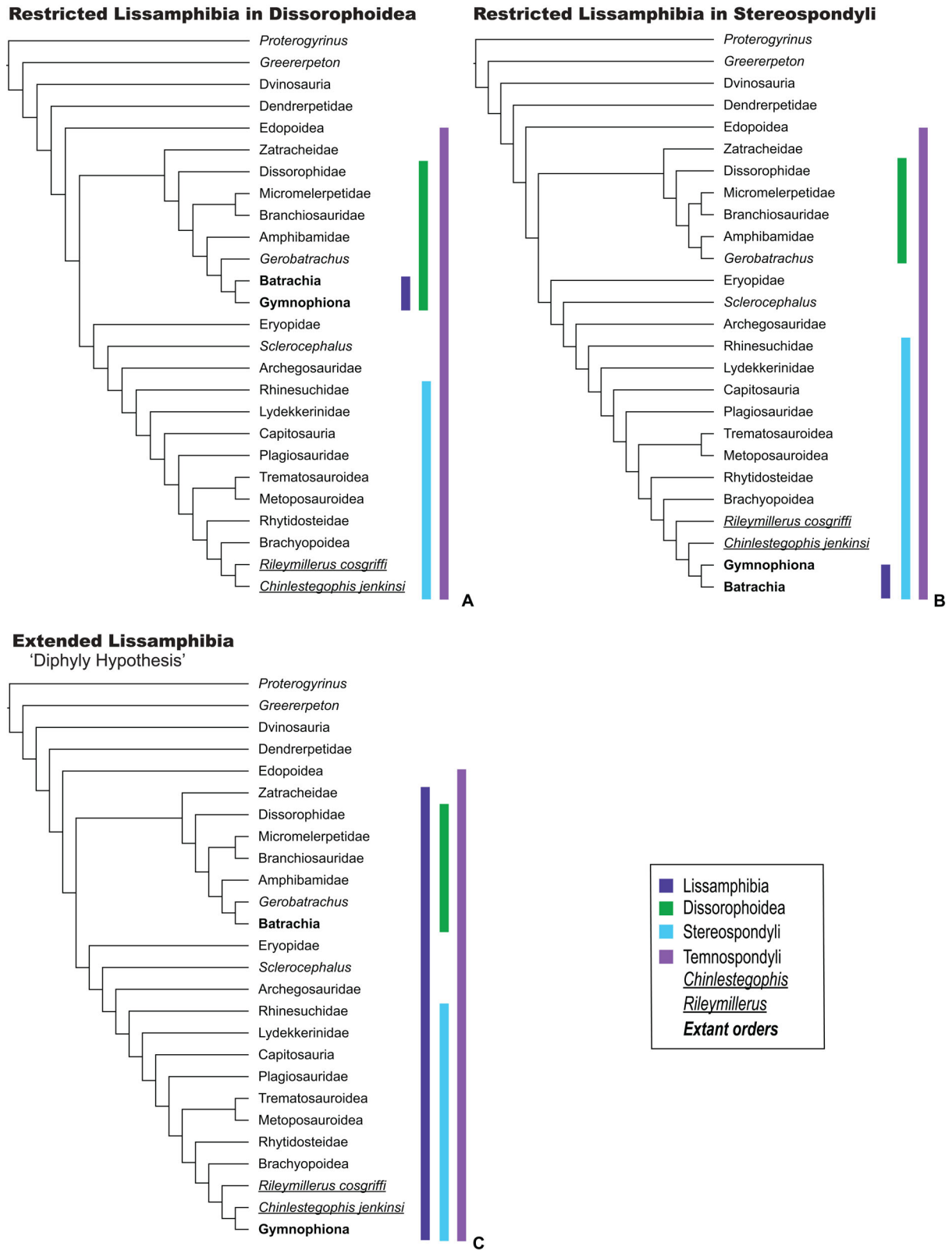


Figure 1. Illustrated hypotheses of lissamphibian relationships based on simplified partitioned-by-island consensus (Serra Silva & Wilkinson, 2021a) of the set of most parsimonious trees inferred from Pardo et al.'s (2017) matrix with labelled higher taxonomy for **A**, restricted Lissamphibia in Dissorophioidea; **B**, restricted Lissamphibia in Stereospondyli; and **C**, extended Lissamphibia.

hypothesis (Fig. 1C). Under this hypothesis, Lissamphibia no longer consists of a restricted clade nested within Dissorophoidea Bolt, 1969 (usually referred to as the Lissamphibia monophyly hypothesis; here it will be referred to as either the restricted Lissamphibia or restricted in Dissorophoidea hypothesis), but rather it spans Dissorophoidea and Stereospondyli von Zittel, 1887 (as sampled by Pardo et al. [2017]), which has paradigm-shifting implications for the clade's age and evolutionary history. This Dissorophoidea + Stereospondyli hypothesis, one of the polyphyly hypotheses that is sometimes referred to as the Lissamphibia diphyly hypothesis (e.g. Santos et al., 2020), is not accommodated under the definition of Lissamphibia used in this work, henceforth I will refer to it as the extended Lissamphibia hypothesis (see Fig. 1 for relevant Lissamphibia hypotheses). However, subsequent work has not supported extended Lissamphibia, with a slightly revised and extended morphological data matrix recovering the restricted Lissamphibia clade nested in Dissorophoidea and a distantly related *Chinlestegophis* in Stereospondyli (Schoch et al., 2020). Furthermore, Serra Silva and Wilkinson's (2021a) renewed work on islands of trees (Maddison, 1991) suggests that Pardo et al.'s (2017) conclusions stem from dismissal of tree set heterogeneity and the phenomenon of large island bias. The latter is particularly pronounced for this dataset, since the MRC of all most parsimonious trees (MPTs) was very resolved ($\rho = 0.96$, this measure is equivalent to Colless's [1980] consensus fork) and nearly identical to that of the largest 1-TBR (tree bisection-reconnection) island (PAUP*'s [Swofford, 2003] default branch-rearrangement for heuristic parsimony searches), yet the strict consensus (SC: Sokal & Rohlf, 1981) was comb-like (available from Serra Silva & Wilkinson [2021b]). Combined with Marjanović and Laurin's (2019) findings that analytical settings can drastically change the phylogenetic relationships inferred between Palaeozoic vertebrates, the topological heterogeneity of Pardo et al.'s (2017) set of inferred MPTs makes their data matrix a prime candidate for extensive re-analyses under revised character schemes and parametrizations to ascertain its robustness. Recently, Kligman et al. (2023) found support for the restricted Lissamphibia in Dissorophoidea, but the extensive character scoring and construction revisions make direct comparisons to the Pardo et al. (2017) matrix a less straightforward endeavour.

The taxonomic instability that characterizes the Pardo et al. (2017) dataset and leads to the multimodal island structure of the inferred MPTs (see Serra Silva & Wilkinson, 2021a, fig. 5a), also raises the interesting question of how the presence of multiple islands affects

methods for unstable taxa identification. While poorly resolved consensus trees are themselves diagnostic of taxonomic instability, multiple data- and topology-based methods have been developed to identify which, and sometimes why, taxa are unstable (e.g. Aberer et al., 2012; Thorley & Wilkinson, 1999; Wilkinson, 1995a). In addition to these, some non-parametric resampling techniques, primarily used to test tree robustness to data perturbations (e.g. first order taxon-jackknifing: Lanyon, 1985), can also be used to identify rogue taxa. However, these resampling approaches often require multiple rounds of tree inference, thus limiting their widespread use. While the effects of isolated or grouped unstable taxa have been tested for some topology-based methods (Wilkinson & Crotti, 2017), it seems the same has not been done for the presence of multiple tree subsets. Given that the presence of multiple subsets of equally optimal trees can substantially influence both tree search (e.g. Höhna & Drummond, 2011; Lakner et al., 2008; Olmstead et al., 1993; Sanderson et al., 2011) and the summary of inferred trees (e.g. Sharkey & Leathers, 2001; Serra Silva & Wilkinson, 2021a; Sumrall et al., 2001), it is not unreasonable to expect that topology-based taxonomic instability identification analyses, such as leaf stability (LS: Thorley & Wilkinson, 1999; Wilkinson, 2006), might also be affected by large island bias.

An alternative to topology-based instability tests, which displays split incompatibilities, is the use of (non-tree) phylogenetic networks. These are being increasingly used to summarize (multi)sets of phylogenetic trees where incomplete lineage sorting (ILS), hybridization or other reticulation events are suspected (reviewed in Elworth et al., 2019). While phylogenetic networks have been used primarily to explore incongruences in molecular datasets and/or trees inferred from molecular data, the latter often called consensus networks (Holland & Moulton, 2003), they have also started to be applied to morphological data. Using a set of craniodental characters, Caparros and Prat (2021) recently applied phylogenetic networks to the study of hominin evolution. However, despite the large size of the tree set they used for the network analyses (9639 trees), all trees belonged to the same 1-TBR island, and the network analyses would thus have been immune from large island bias. As such, it remains unclear whether the presence of multiple islands affects phylogenetic networks other than phylogenetic trees.

Here, using the MPTs inferred from Pardo et al.'s (2017) matrix, I explore how two topology-based instability tests, LS and relative bipartition information content (RBIC: Aberer et al., 2012), behave in the presence of (multiple) islands of trees (*sensu* Serra Silva &

Wilkinson, 2021a) and how island structure can be taken into account when using these methods. I also use two (non-tree) phylogenetic network methods to show that large island bias extends beyond the MRC. And, lastly, I re-analyse the Pardo et al. (2017) data matrix to explore how it behaves under different analytical parameters and how robust their conclusions on lissamphibian relationships are.

Materials and methods

Matrices and tree sets

The data matrices used for the phylogenetic analyses were obtained from Pardo et al.'s (2017) and Schoch et al.'s (2020) supplementary materials, and changes to these are detailed and discussed below. The original Pardo et al. (2017) data matrix consisted of 76 taxa and 345 morphological characters (six constant and 17 parsimony uninformative). Of the sampled taxa, seven are extant (Anura: *Leptodactylus* Fitzinger, 1826 and *Xenopus* Wagler, 1827; Caudata: *Ambystoma* Tschudi, 1838, *Cryptobranchus* Leuckart, 1821 and *Hynobius* Tschudi, 1838; Gymnophiona: *Epicrionops* Boulenger, 1883 and *Ichthyophis* Fitzinger, 1826), and all taxa except *Proterogyrinus* Romer, 1970 and *Greererpeton* Romer, 1969 are temnospondyls. Thus, this matrix assumes that Lissamphibia is nested within Temnospondyli von Zittel, 1887 and cannot be used to explore the Temnospondyli vs Lepospondyli von Zittel, 1887 hypothesis (see Marjanović & Laurin, 2019).

As for the characters, while most characters are defined as binary, the matrix also includes multistate characters with three and four coded states (see Pardo et al. [2017, supplementary material Part C] for character descriptions). The characters were treated as unordered and equally weighted for all analyses detailed below (see Marjanović et al. [2024] for an exploration of how ordering affects the trees inferred from this data matrix). Additionally, the matrix includes characters with polymorphic, missing and inapplicable states. There are 10 characters with taxa scored as polymorphic (characters 10, 17, 106, 160, 246, 248, 253, 267, 321 and 324). As for characters with missing (“?”) and inapplicable (“-”) states, 186 characters include missing data and 146 have both missing and inapplicable data (all characters with inapplicable data also have at least one taxon scored as missing). At the matrix level, $\approx 2\%$ and $\approx 25\%$ of cells consist of, respectively, inapplicable and missing data. For the differences between Pardo et al. (2017) and Schoch et al. (2020) see the subsection immediately preceding the Discussion section, below.

For the taxonomic instability, phylogenetic network and *a posteriori* taxon jackknifing analyses, unless otherwise

stated, the MPT and Bayesian distribution tree sets used were those inferred by Serra Silva and Wilkinson (2021a) from Pardo et al.'s (2017) original data matrix. Any tree searches that did not use default parameters, both under parsimony and Bayesian optimality criteria, are described below. All analyses were run on a 64-bit Windows 10 Enterprise system running on an Intel® Core™ i9-8950HK 2.90 GHz processor with 32Gb of RAM. All data matrices, scripts and output files are available in the Dryad repository: <https://doi.org/10.5068/D1RT1J>.

Taxonomic instability analyses

Methods to explore taxonomic instability can be broadly divided into data-based (e.g. safe taxonomic reduction [STR: Wilkinson, 1995a] and its heuristic extension Concatabominations [Siu-Ting et al., 2015], and *a priori* taxon jackknifing) and topology-based methods (e.g. LS, RBIC and *a posteriori* taxon jackknifing). While the presence of multiple subsets of trees is not expected to negatively influence data-based methods (even if we may expect some of, or all, the taxa contributing to island structure to be identified as unstable), topology-based methods may be prone to some of the problems highlighted for the MRC. In other words, if there is more topological variation between subsets of trees than within, could the set of identified unstable taxa change between analysing the full set of trees and analysing each subset separately? For ease of interpretation and discussion, I will focus solely on the set of 1-TBR/10-Robinson-Foulds (RF: Robinson & Foulds, 1981) tree islands Serra Silva and Wilkinson (2021a) extracted from the Pardo et al. (2017) set of MPTs. If we do expect topological variation to be greater between islands, we may sensibly expect four outcomes: a taxon is (1) stable within and between tree islands; (2) stable within islands, but unstable between; (3) unstable within islands, but stable between; or (4) unstable within and between islands (Table 1). These outcomes correspond to interactions between causes of instability and relationships between taxa in a (sub)set of trees. I hypothesize that taxa that are stable only within islands correspond to the taxa responsible for island structure (globally unstable taxa *sensu* Serra Silva & Wilkinson [2021a]) and their instability is caused, primarily, by homoplasy, i.e. independently evolved similarity. Taxa that are

Table 1. Instability patterns possible for any taxon present in a partitionable (multi)set of trees.

	Within islands	
Between islands	stable	unstable
Stable	++	+-
Unstable	-+	--

unstable only within islands, on the other hand, correspond to locally unstable taxa and are likely caused by missing data (potentially ILS in molecular datasets), although less severe levels of homoplasy (than seen in the globally unstable taxa) cannot be discarded as a contributing factor. Finally, taxa that are unstable both within and between islands likely have very large amounts of missing data and/or homoplasy, and inference analyses may benefit from their removal. Understanding if, and how, topology-based methods are influenced by the presence of multiple islands is especially important when dealing with datasets, like Pardo et al.'s (2017), where the SC is mostly unresolved, hinting at extensive taxon instability, but *a priori* data-based approaches, like STR (using PerlEQ v1.0: Jeffery and Wilkinson, unpub.; available at <https://uol.de/systematik-evolutionsbiologie/programme>), do not identify any taxa for safe removal, i.e. taxa whose removal does not affect the relationships between any other taxa in the matrix.

To explore whether the presence of large islands influences topology-based taxonomic instability analyses, I ran the full set of Pardo et al.'s (2017) MPTs, and each tree island, through LS analyses, as implemented in RogueNaRok (RNR: Aberer et al., 2012). All LS measures were used for the analyses: (1) LS_{max} , the normalized maximum (bootstrap) proportion for a quartet/triplet; (2) LS_{diff} , the normalized difference between the proportions of the two most supported quartet/triplet resolutions; and (3) LS_{ent} , the information theory-based entropic LS, which is the normalized negative sum of the product of the frequency of each quartet/triplet resolution and the log of that frequency (f) (Thorley, 2000; Thorley & Wilkinson, 1999; Wilkinson, 2006):

$$LS_{ent} = 1 - \frac{\sum (f \times \log(f))}{\log(f)} \quad (1)$$

The MPTs were also run through RNR's RBIC optimality criterion, set to optimize the SC. The RBIC is defined by Aberer et al. (2012) as the ratio between the sum of all support values (S) and the theoretical maximum support for a binary tree prior to taxon removal:

$$RBIC = \frac{\sum S}{100 \times (n - 2)}, \quad (2)$$

where $n - 2$ corresponds to the number of internal branches in a fully resolved unrooted tree with n tips ($n - 1$ for rooted trees), and the multiplication factor of 100 assumes the use of percentage-based support values (e.g. bootstrap proportions; Felsenstein, 1985). The SC was chosen due to the MRC biases detailed in Serra Silva and Wilkinson (2021a) and Sumrall et al. (2001). Under the SC the RBIC is equivalent to Pattengale

et al.'s (2011) relative information content optimization criterion, which consists of optimizing the number of bipartitions in the SC. Interestingly, this measure was also one of the parameters (the other being number of taxa) used by Wilkinson (1995b) to distinguish primary reduced consensus (RC) trees from other RC trees. Following Wilkinson and Crotti (2017), several dropset sizes (1, 5, 10 and 15) were used in the RBIC analyses, since unstable taxa may be missed if they are part of an unstable group, i.e. where the taxa in a clade are stable in relation to each other but the group is unstable in relation to the rest of the tree. Given the homogeneous nature of the Bayesian tree distribution explored by Serra Silva and Wilkinson (2021a), it would not be informative to the question of whether large island bias affects LS analyses and thus will only be run through the RBIC.

Phylogenetic networks

An alternative to consensus trees, clustering analyses and instability tests that can display at least some of the topological incongruences present in a (multi)set of trees in one graphical summary is phylogenetic networks. While there are multiple types of phylogenetic networks (summarized in Huson et al., 2010, fig. 4.1), including phylogenetic trees, I focus only on those used in Caparros and Prat (2021): consensus (Holland & Moulton, 2003) and reticulation networks (Huson et al., 2005). Consensus and reticulation networks differ in that, for a given (multi)set of trees, the former display all splits over a selected frequency threshold (setting this to 50% is equivalent to computing the MRC), and thus the split incompatibilities present in the tree (multi)set (Holland & Moulton, 2003). Reticulation networks, on the other hand, use those split incompatibilities to identify and display the presence, but not type, of reticulation events (Huson et al., 2005). In short, consensus networks summarize split incompatibilities, while reticulation networks display evolutionary histories. These approaches, particularly reticulation networks, have generally been applied to molecular datasets (including, but not restricted to, sequence data, inferred trees, matrix representation of splits, *etc.*), partly because they were developed with the aim of understanding the topological conflicts in (multi)sets of gene trees (e.g. Huson et al., 2005), and partly because reticulation events are often investigated at the molecular not morphological level (e.g. Cai et al., 2021; Suvorov et al., 2022).

Using SplitsTree v. 4.14.8 (Huson & Bryant, 2006) I computed the consensus and reticulation networks for the Pardo et al. (2017) MPTs. The consensus networks were computed for split frequency thresholds of 33%

(default), 20%, 10% and 0%, all displayed with ‘EqualAngle’. The reticulation networks were computed from the consensus networks, using the RECOMB2007 algorithm (Huson & Klopper, 2007), rooted on *Proterogyrinus*, and displayed with ‘EqualAngle120’.

Re-analysis of Pardo et al.’s (2017) data matrix

Congruence between the tree summaries of multiple phylogenetic analyses of the same dataset using distinct inference methods is often used as a sign of data robustness and used to bolster confidence on the inferred tree topologies (e.g. Pardo et al., 2017; San Mauro et al., 2014). However, when using the MRC to summarize inferred tree (multi)sets, this congruence can be positively misleading, particularly when large island bias is present. As noted by Serra Silva and Wilkinson (2021a), the MRCs of the parsimony and Bayesian inference analyses of Pardo et al.’s (2017) data matrix both support the extended Lissamphibia hypothesis, despite not being identical (RF = 19, see Pardo et al., 2017, fig. S7). Yet, the island structures of the tree sets yielded by the two analyses are highly disparate. The MPTs, under various metrics and thresholds (1-NNI, 1-SPR, 1-TBR, $2 \leq \text{RF} \leq 10$) can be partitioned into five well-defined islands that support distinct Lissamphibia make-ups and placements, whereas for the Bayesian tree distribution the island structure changes for each $2 \leq \text{RF} \leq 14$ (see Serra Silva & Wilkinson 2021a, fig. 5, table 2). Despite the difference in size and number of islands, the overall island structure for the Bayesian tree distribution consists of a large central island surrounded by large numbers of singleton or small islands that uniformly support the extended Lissamphibia hypothesis.

Moreover, topology-based instability analyses further emphasize the difference between the tree sets recovered by the parsimony and Bayesian analyses (see below). The results obtained by the inference and taxonomic instability analyses show that the Pardo et al. (2017) data matrix is not robust to changes in inference method, which is further supported by Schoch et al.’s (2020) analyses of a slightly modified version of the Pardo et al. (2017) data (see Results, below, on the differences between these matrices). Thus, to understand the conundrum of the phylogenetic placement of *Chinlestegophis* and *Gymnophiona*, we must first understand the data matrix itself and how the data respond to different analytical/parametric choices.

Resampling analyses. Pardo et al. (2017) used the topological congruence between the MRCs of their parsimony and Bayesian analyses to bolster their extended Lissamphibia hypothesis but did not refer to any resampling analyses under parsimony, despite including

bootstrap values in their parsimony MRC (in Pardo et al. 2017, fig. S7B, most branches have BP < 50%). Bootstrap proportions are commonly used as branch support measures in non-Bayesian tree inference. Serra Silva and Wilkinson (2021a) pointed out that, under parsimony, none of the splits that separate *Gymnophiona* and *Batrachia* had a bootstrap support (Felsenstein, 1985) greater than 50%; however, no analytical settings were reported. Thus, to compare across multiple resampling analyses, I ran a bootstrap analysis in PAUP* v. 4a165 (Swofford, 2003), with 1000 replicates under the same settings as the heuristic equal-weights parsimony search reported by Pardo et al. (2017) and the ‘MulTrees’ option selected. Delete-half (Lanyon, 1985) and Farris (delete 1/e \approx 36.8%: Farris et al., 1996) jackknife analyses were run with the same settings as the bootstrap analysis, except for the number of replicates, which were set to 100 due to memory constraints.

In addition to serving as measures of branch support, resampling techniques can also be used to identify unstable taxa. Thus, taxon jackknifing might help us identify the taxa whose data underpin the taxonomic instability that characterizes the Pardo et al. (2017) data matrix. Because a data-based first-order taxon jackknife requires an inference analysis each time a taxon is removed from the matrix, these resampling analyses can be quite time and resource intensive. A way of optimizing first-order taxon jackknifing is to use the output from (other) taxonomic instability analyses as a guide. Unfortunately, as mentioned above, the STR analysis (and its heuristic extension, Concatabominations: Siu-Ting et al., 2015) did not find any candidate taxa for safe removal. However, from the partitioned-by-island consensus reported by Serra Silva and Wilkinson (2021a), we know that *Chinlestegophis* always places with *Eocaecilia* Jenkins and Walsh, 1993 and/or *Rileymillerus* Bolt and Chatterjee, 2000. There is thus a trio of taxa for which we can explore whether removing any one of them from the matrix affects the number of trees and islands inferred from the resampled matrices, and whether, removed taxa aside, the topologies inferred are similar between taxon jackknifing runs. For matrix-based first-order taxon jackknifing analyses I ran three heuristic parsimony searches in PAUP*, with random addition sequence, default branch swapping and 1000 replicates. Each analysis differed only in the taxon removed: *Chinlestegophis*, *Eocaecilia* or *Rileymillerus*. For topology-based *a posteriori* taxon jackknifing, I explored the effects of removing different sets of taxa from the inferred MPTs. These analyses consisted of removing: (1) *Chinlestegophis*, *Eocaecilia* and/or *Rileymillerus* in every combination of one, two or three taxa; (2) taxa sorted by the levels of instability

computed by the all trees LS_{max} analysis (Supplemental material Table S1); and (3) each taxon on the tree iteratively. The *a posteriori* jackknife analyses were conducted in R v. 4.1.2 (R Core Team, 2021), with the resulting tree sets being checked for identical topologies and ran through the *xRFislands* function from the *islandNeighbours* v. 0.9.0 (Serra Silva & Wilkinson, 2021a) package to identify changes in 2-RF islands (this is equivalent to identifying 1-NNI islands: Chernomor et al., 2015).

Treatment of polymorphic taxa, missing and inapplicable data. Unlike aligned molecular data matrices, where characters consist of positions in a sequence and the states are dictated by the nucleotide or amino acid residue identified at each position, in morphological datasets characters and character states are delineated and identified by systematists (Wilkinson, 1995c). As such, morphological data matrices carry with them a certain degree of subjectivity, which can lead to different workers scoring the same character(s) differently for the same taxa (e.g. Pardo et al., 2017; Schoch et al., 2020; Marjanović et al., 2024). This subjectivity contributes factors such as taxonomic sampling, which can change/affect the decision of where/how to delineate characters and character states, uncertain homology between skeletal elements due to bone loss/fusion (e.g. Maddin et al., 2016; Schultze et al., 2008), and simple typographical errors and mistypes cannot be discounted either (for a discussion of how errors can accrue in a morphological matrix see Marjanović & Laurin [2019]). While the differing character scores between Pardo et al.'s (2017) and Schoch et al.'s (2020) matrices alone lead to the inference of trees displaying different evolutionary hypotheses, the Pardo et al. (2017) data matrix also includes two aspects of character scoring, polymorphic taxa and inapplicable data, whose presence and choice of how to analyse them can lead to the recovery of different (multi)sets of inferred trees (e.g. Brazeau et al., 2019; Nixon & Davis, 1991; Platnick et al., 1991).

Starting with polymorphic taxa, which can at best lead to changes in tree length (Nixon & Davis, 1991), the Pardo et al. (2017) matrix contains 10 characters for which at least one taxon was scored as polymorphic. In PAUP*, the user can choose to treat these character codings as uncertainty (the taxon might have any of the selected character states) or polymorphism (individuals in the taxon display one character state or the other[s]), which may lead to distinct character optimizations (see pp. 103–106 in the PAUP* manual for details: <https://phylosolutions.com/paup-documentation/paupmanual.pdf>). According to how the optimization for polymorphic taxa is implemented in PAUP*, choosing the

'msTaxa = polymorph' setting should only affect tree length, not topology. To confirm this, I ran the Pardo et al. (2017) data matrix through PAUP*, under the same parsimony settings described above, with the 'msTaxa = polymorph' option selected (additional analysis are summarized in Table 2). MrBayes v. 3.2.6 (Ronquist et al., 2012), on the other hand, always treats polymorphism as uncertainty (pp. 11–12 and 85–87 in the manual: http://mrbayes.sourceforge.net/mb3.2_manual.pdf), and thus no comparison to analysing polymorphism as variation is possible.

An important side note about analysing morphological/standard datasets in MrBayes is that, because the programme implicitly partitions datasets by each character's highest observed score – i.e. if a character's highest score in the matrix is '2', the character is assumed to have three possible states: 0, 1 or 2 – it occasionally makes assumptions about character states (see p. 148 in the manual). Also, when confronted with characters where the only numerical score present is 0, MrBayes assumes that those characters are binary (0,1). This was the case for characters 61 and 244 in Pardo et al.'s (2017) matrix, because all taxa were coded as 0, missing or inapplicable. These state-based assumptions make MrBayes' implicit partitioning scheme differ from an explicit partitioning scheme based on the character descriptions. A Bayesian analysis of a data matrix explicitly partitioned based on Pardo et al.'s (2017, supplementary information, Part C) character descriptions was run in MrBayes, under the Mk + G model, with two independent runs of four chains for 10,000,000 generations, sampled every 10,000, and with a relative burn-in of 25%.

Returning to character coding, most characters in the Pardo et al. (2017) matrix have at least one taxon scored as inapplicable and/or missing, with $\approx 2\%$ and $\approx 25\%$ of the matrix consisting, respectively, of inapplicable and missing data. While these character scores are often treated as equivalent (e.g. Schoch et al., 2020, see below), they represent two different types of 'unknowns'. Scoring characters as missing, commonly coded as '?', represents lack of knowledge/information. An example of this would be that, due to taphonomic processes, only the phalanges of an individual are fossilized. Anatomical knowledge dictates that the phalanges were at the end of a limb, but we cannot score any limb-specific characters because we are missing the necessary biological materials to do so. Scoring characters as inapplicable ('-'), on the other hand, denotes a logical impossibility. Using Maddison's (1993) classic example of tail presence and colour, if we apply conventional coding (Hawkins et al., 1997), sometimes referred to as contingent coding (Forey & Kitching,

Table 2. Summary of equal-weights maximum parsimony settings and results of Pardo et al.'s (2017) and Schoch et al.'s (2020) re-analyses. †Lissamphibia within Stereospondyli in one island. ‡Lissamphibia within Stereospondyli in two islands. *See Supplemental material Table S7 for character recoding.

Matrix	Polymorphism	Inapplicable data	Trees	Tree length	Islands	Lissamphibia
Pardo original	Uncertainty	Missing	882	1514	5	Extended and restricted†
Pardo, no <i>Chinlestegophis</i>	Uncertainty	Missing	18	1479	1	Restricted
Pardo, no <i>Eocaecilia</i>	Uncertainty	Missing	180	1471	1	Restricted
Pardo, no <i>Rileymillerus</i>	Uncertainty	Missing	162	1494	1	Restricted
Pardo original	Polymorphism	Missing	882	1532	5	Extended and restricted†
Pardo original	Uncertainty	Additional state	351	1699	2	Restricted‡
Pardo original	Polymorphism	Additional state	351	1717	2	Restricted‡
Pardo rescored*	Uncertainty	Missing	108	1499	2	Restricted
Pardo rescored*	Polymorphism	Missing	108	1517	2	Restricted
Pardo rescored*	Uncertainty	Additional state	270	1762	2	Restricted†
Pardo rescored*	Polymorphism	Additional state	270	1780	2	Restricted†
Schoch original	NA	NA	2	1392	1	Restricted
Schoch, 345 characters	NA	NA	3	1366	1	Extended
Schoch rescored*	NA	NA	2	1371	1	Restricted
Schoch rescored*, 345 characters	NA	NA	3	1351	1	Restricted

2000) or hierarchical characters (Hopkins & St. John, 2021), we would need two characters to explain the variation in tail colour: character A (tail present/absent) and character B (if present, tail blue/red). Thus, character B would be inapplicable for any taxa without a tail, as a tail cannot be both absent AND have colour. Such reconstructions can, however, occur during character optimization, affecting tree scores (e.g. Vignette 2 in Brazeau et al., 2019; Maddison, 1993). While there is still debate on how best to code variation that, just like the tail colour example, is logically and/or biologically dependent (e.g. Hawkins et al., 1997; Hopkins & St. John, 2021; Maddison, 1993; Wilkinson, 1995c), it is sometimes suggested that, once a presence/absence character has been defined, any new character that subdivides the original character's 'presence' state should be hierarchically/conditionally defined (e.g. Hawkins et al., 1997). In other words, if character A is tail presence/absence, we should not define a new three-state character (C) as tail absent, spotted tail, stripy tail, rather it should follow the same character definition as character B, i.e. character C – if present, tail spotted/stripy. In Pardo et al.'s (2017) matrix, this suggestion was not always followed, resulting in multiple instances of logically inconsistent/impossible character definitions and/or scores (Supplemental material Table S7, see below). These coding inconsistencies/impossibilities were rescored based exclusively on the independent character(s) definition/scoring (except for character 314, see Supplemental material Table S7). For this study, no

specimens were re-examined and, if a revision of character construction/definition was warranted, dependent characters were rescored conservatively as either inapplicable ('-') or missing ('?'). The decision of whether to rescore a character as inapplicable vs missing in the Pardo et al. (2017) matrix was based on the existing scores for a given taxon. Specifically, if most characters in a logically dependent set had been coded as inapplicable ('-') for a given taxon, then the characters necessitating recoding were changed to inapplicable. If, however, the majority of the set was coded as missing ('?') then the revised characters were recoded as such. This recoding strategy assumes that the use of missing/inapplicable states was consistent throughout the matrix. While there is sufficient evidence that missing and inapplicable scores have not been applied consistently (see Gee [2022] for a thorough discussion of the evolution of this family of data matrices), the primary goal of this recoding was to test whether, under default search parameters ('GapMode=Missing') and as few possible changes (including matrix size) to the Pardo et al. (2017) and Schoch et al. (2020) matrices, minimizing logical inconsistencies has an effect on the inferred topologies.

Lastly, the presence of inapplicable data also raises the question of how best to analyse these characters. Especially since, for as long as systematists have been debating how to code logically dependent characters, they have also been debating how to analyse these characters once conventional coding has been applied (e.g.

Hawkins et al., 1997; Maddison, 1993). PAUP* allows users to choose between treating inapplicable data as missing or as a new character state (see pp. 40–41 in the manual), with the former as default. To check whether treating inapplicable data as a distinct score from missing affected the sets of MPTs inferred from Pardo et al.'s (2017) matrix, I ran a tree search in PAUP* with the same settings as previous parsimony analyses, but with 'GapMode = NewState' selected.

As for methods developed specifically to deal with inapplicable data, Brazeau et al.'s (2019) algorithm is aware of inapplicable data during character optimization and does not require the relationship between hierarchical characters to be specified (unlike for example De Laet [2015] and Goloboff et al. [2021]). They found that treating inapplicable data as 'missing', 'new state' or 'inapplicable' often yields different sets of MPTs, for the same dataset, but not a clear pattern of variation in the results, i.e. no approach always found the shortest trees, the most resolved SC, the broadest sampling of tree space, etc. To compare Brazeau et al.'s (2019) approach to the settings offered in PAUP*, I analysed the original Pardo et al. (2017) matrix with Morphy v. 0.2 (Brazeau & Desjardins, 2020), running 1000 replicates with random addition sequence.

Recently, Simões et al. (2023) explored the effects of inapplicable data on tree space traversal and topological accuracy of the inferred (multi)sets, but this topic is beyond the scope of this study.

Results

Taxonomic instability

Most parsimonious tree set(s). Starting with all 882 MPTs, LS analyses identified crown lissamphibians as the most unstable taxa (Supplemental material Table S1), with *Chinlestegophis* within the 15 most unstable taxa (taxon instability ranking varies slightly between LS measures). The RBIC results for the set of MPTs changed with dropset size, with RBIC-1, 5 and 10 recovering *Chinlestegophis* as the single most unstable taxon, and its removal increasing the number of bipartitions in the SC from 35 to 36. RBIC-15, however, recovered two sequential sets of unstable taxa, the first included all crown lissamphibians (both extant and extinct) and *Chinlestegophis*, removal of which drives the number of bipartitions in the SC to 48, and the second set made up of *Laidleria* Kitching, 1957, whose placement differs between the four smallest islands and the largest island (Serra Silva & Wilkinson, 2021a, figs 3a–d, 2b, respectively), removal of this taxon adds another bipartition to the SC. Thus, removal of the taxa

identified by RBIC-15 on the full set of trees increases the resolution of the SC from 35 to 49 bipartitions.

When the islands were individually tested for the presence of unstable taxa, however, all LS measures exceed 0.94 (Supplemental material Tables S2–S6), and the RBIC analyses identify *Trematosaurus* Burmeister, 1849 as unstable in island-72 and -216 (Serra Silva & Wilkinson 2021a, fig. 3b, d). Thus, just as in the MRC scenario, identification of unstable taxa in the presence of multiple tree subsets yields different results if we look at those sets individually, or if we look at the full (multi)set. In this example, for the within-island analyses, the taxa with the lowest stability scores can be found in the areas of local instability identified in Serra Silva and Wilkinson (2021a, figs 2b, 3a–d, table 1). In short, with Pardo et al.'s (2017) trees, testing all MPTs identified the group(s) of taxa whose position changed between, but not within, islands of MPTs as unstable (crown Lissamphibia and *Chinlestegophis*), but the within-islands analyses found only 'locally' unstable taxa (e.g. *Trematosaurus*). Additionally, RBIC analyses with dropsets >1 identifying groups of unstable taxa may also be indicative of an underlying island structure in the (multi)set of trees being analysed.

To summarize, the taxonomic instability analyses on the MPTs recovered three of the four predicted outcomes: (1) most taxa are stable within and between islands; (2) crown Lissamphibia and *Chinlestegophis* are stable within islands, but unstable between them; (3) *Trematosaurus* is unstable within some islands, but stable between them; but (4) no taxa were identified as unstable within and between islands. Following from the above hypothesized links between instability patterns and causes, Lissamphibia and *Chinlestegophis*' instability is likely caused by homoplasy, particularly between *Chinlestegophis*, *Eocaecilia* and *Rileymillerus* (see Re-analysis, below, for details), while the instability of taxa like *Trematosaurus* is likely due to missing data, although low levels of homoplasy cannot be discarded. Thus, understanding how unstable taxa behave in the presence of multiple islands of trees is not essential to interpreting taxonomic instability analyses, but understanding the relationship between island structure and causes of instability might help mitigate misinterpretations of topological heterogeneity.

Bayesian posterior distribution. While the RBIC analyses of the MPTs recovered different sets of unstable taxa for different dropset sizes (see above), the same analyses for the Bayesian tree distribution recovered *Apateon* von Meyer, 1844, *Iberospondylus* Laurin and Soler-Gijón, 2001, *Lapillopsis* Warren and Hutchinson, 1990 and *Sangaia* Dias-da-Silva and Marsicano, 2006 as the most unstable taxa for all dropsets. The only

variations between the results of the latter analyses are whether *Iberospondylus* and *Lapillopsis* should be removed separately or as a unit, and the order in which the four taxa should be removed.

Phylogenetic networks

For ease of visualization, Nexus formatted network files and EPS image files are available from <https://doi.org/10.5068/D1RT1J>. The areas of instability mentioned from here on correspond to those identified and discussed by Serra Silva and Wilkinson (2021a) in their work dealing with the generalization and exploration of islands of trees. The consensus network for the default split frequency threshold shows most of the split incompatibilities centred on the areas of local instability VIII and IX (Serra Silva & Wilkinson 2021a, fig. 2a, b), along with the possible solutions for the polytomies corresponding to areas I, II, V and VI (Fig. 2A; Serra Silva & Wilkinson, 2021a, figs 2b, 3). The 20% threshold displays split incompatibilities from areas VII to XII, which share taxa with areas VIII and IX, and the solutions to polytomies I to VI (Fig. 2B; Serra Silva & Wilkinson, 2021a, figs 2b, 3). As in the two largest islands, *Chinlestegophis* is placed with *Rileymillerus* and *Gymnophiona*, all nested within *Stereospondyli* in the 33% and 20% threshold consensus networks. At thresholds 10% and 0% the consensus networks are (near) uninterpretable (Fig. 2C, D), due to the very large number of split incompatibilities present in the input tree set. However, in the 10% threshold network *Chinlestegophis* is found with *Rileymillerus*, while *Gymnophiona* and *Batrachia* are each in well-defined clades, and in the 0% threshold network the relationship between *Chinlestegophis* and *Rileymillerus* breaks down, but *Gymnophiona* and *Batrachia* are still recovered as clades. These results are consistent with what we know of the island structure of Pardo et al.'s (2017) MPTs. First, areas of instability VII–XII are each present in at least one of the four largest islands, share taxa (e.g. *Sangaia* contributes to areas VIII and XII) and/or are in close proximity in at least two islands. Additionally, many of the taxa in these local instability areas were also recovered by most island-specific LS analyses as some of the least stable taxa (e.g. *Trematosaurus*, *Trematolestes* Schoch, 2006, *Benthosuchus* Efremov, 1937). That the 33% threshold consensus network displays split incompatibilities that match up with areas of local instability (VIII and IX) restricted to the largest island (c. 55% of all trees), and the 20% threshold network starts recovering split incompatibilities present in the next largest islands (c. 25% and c. 10% of all trees), suggests that, much like the MRC, consensus networks are also affected by the large island bias phenomenon.

This is further supported by the amount of split incompatibilities found between *Chinlestegophis*, *Rileymillerus*, *Gymnophiona* and *Batrachia*.

The reticulation networks yielded very similar results to the consensus networks, as expected from the latter being the input for the identification of reticulation events. For the default threshold, two reticulation events were identified between taxa contributing to areas VIII and IX, and single events were identified for the isolated polytomies corresponding to areas I, II, V and VI (Fig. 3A). The 20% threshold reticulation network finds single reticulation events for the localized polytomies (areas I to VI) and a minimum of eight reticulation events in areas VII–XII, including the two found under the 33% threshold (Fig. 3B). Unfortunately, the large number of taxa and high level of split incompatibilities in this dataset makes the networks hard to interpret and the reticulation nodes in area VII/X hard to parse. It is, however, still possible to recognize the placement of *Chinlestegophis* with *Rileymillerus* and *Gymnophiona* within *Stereospondyli*, i.e. the extended Lissamphibia hypothesis. The 10% network displays a very complex history of reticulations with all Dissorophoidea and *Stereospondyli* taxa (see Figs 1 and 4 for higher order amphibian taxonomy) being involved in, or descended from, at least one reticulation event (Fig. 3C). This analysis still finds *Chinlestegophis* nested in *Stereospondyli*, but neither the extended nor restricted Lissamphibia are displayed. This reflects the drastic topological shifts between islands-18 and -90 (restricted Lissamphibia in Dissorophoidea), island-72 (restricted Lissamphibia in *Stereospondyli*), and islands-216 and -486 (extended Lissamphibia). The computation of the reticulation network for the 0% threshold consensus network was cancelled after running for 36 hours, as the next slowest analysis (10% threshold) took less than 5 minutes. However, given the consensus network for this threshold (Fig. 2D), we might expect a reticulation network at least as complex as the one inferred for the 10% threshold. Thus, consensus and reticulation networks, much like the MRC, are prone to large island bias, with reticulation networks not always computable in the presence of high levels of split incompatibility.

Resampling analyses

Starting with the bootstrap analysis, it weakly supported (53%) the restricted Lissamphibia and was uninformative in regard to *Chinlestegophis*'s placement. This result might lead us to think that even though the islands that display the restricted Lissamphibia make up only c. 20% of the inferred MPTs, they may in fact be better supported by the underlying data. The jackknife analyses, however, weakly support

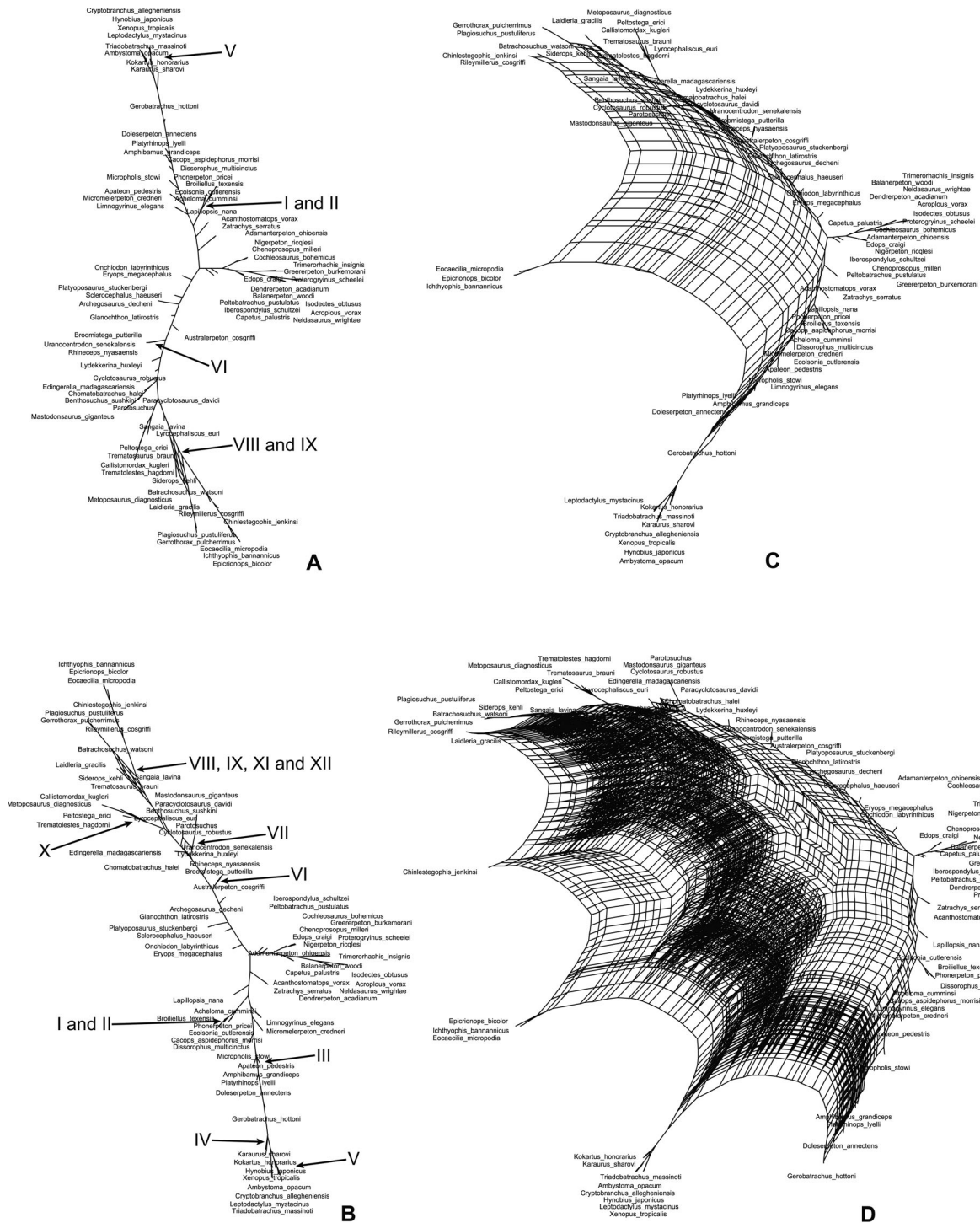


Figure 2. Consensus networks of Pardo et al.'s (2017) set of most parsimonious trees for **A**, 33%, **B**, 20%; **C**, 10% and **D**, 0% split frequency thresholds. Roman numerals correspond to the areas of instability identified in Sera Silva and Wilkinson's (2021a) figures 2b and 3a–d. High resolution image files for each network are available from <https://doi.org/10.5068/D1RT1J>.

Gerobatrachus+*Batrachia* (delete-half = 54%; Farris = 57%) and *Rileymillerus*+*Chinlestegophis*+*Gymnophiona* (delete-half = 55%; Farris = 54%) but are uninformative regarding the relationship between these two clades,

and thus restricted vs extended Lissamphibia. With only 11 out of 345 characters not having taxa with missing ('?') and/or inapplicable ('-') character states, it is not surprising that the trees summarizing the resampling

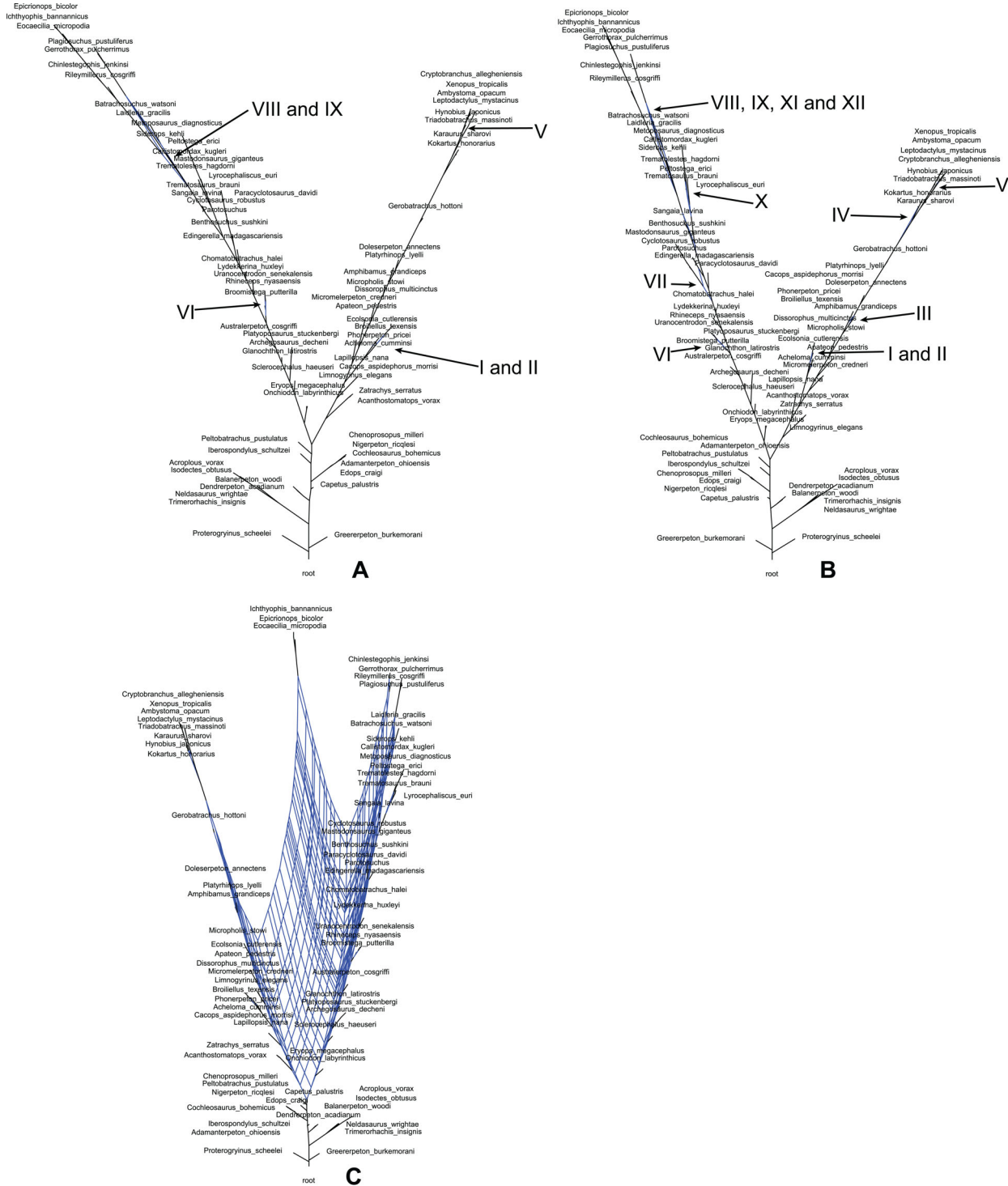


Figure 3. Reticulation networks of Pardo et al.'s (2017) set of most parsimonious trees for **A**, 33%, **B**, 20% and **C**, 10% split frequency thresholds. Reticulate nodes are in dark blue, as per SplitsTree default drawing settings. High resolution image files for each network are available from <https://doi.org/10.5068/D1RT1J>.

analyses are poorly resolved (e.g. Dos Santos & Falaschi, 2007; Wilkinson, 2003). Unfortunately, memory constraints prevented saving the trees inferred for

each resampling replicate, and thus also the check of whether the large island bias present in the heuristic parsimony search also affected the bootstrap and/or

jackknife analyses (commonly summarized with the MRC). That different support measures find generally low branch support across the inferred relationships and that they yield conflicting results regarding the relationship between Batrachia and Gymnophiona, not only confirm that the MRC is a poor summary of the MPTs but also that the Pardo et al. (2017) data matrix is not robust to perturbations.

The first-order taxon jackknifing analyses, on the other hand, strongly support a restricted Lissamphibia, with all three analyses recovering a single 1-TBR island, with restricted Lissamphibia nested within Dissorophoidea (Fig. 4A–C, Table 2). The run without *Chinlestegophis* recovered 18 MPTs with three unstable subtrees: (*Dissorophus* Cope, 1895, *Broiliellus* Williston, 1914, *Cacops* Williston, 1910), (*Acheloma* Cope, 1882, *Phonerpeton* Dilkes, 1990, *Ecolsonia* Vaughn, 1969) and (*Rhineceps* Watson, 1962, *Uranocentron* van Hoepen, 1915, *Broomistega* Shishkin & Rubidge, 2000). These subtrees correspond to the three areas of instability present in all 1-TBR islands identified from the MPTs inferred from the original Pardo et al. (2017), areas I, II and VI (Serra Silva & Wilkinson, 2021a, fig. 2b, table 1). However, for the two runs where *Chinlestegophis* is retained the number of MPTs recovered increases comparatively to the no-*Chinlestegophis* analysis, and *Chinlestegophis* is placed as the sister taxon to the remaining taxon (*Rileymillerus* or *Eocaecilia*) (Table 2, Fig. 4B, C). The increase in number of MPTs is reflected by an increase in taxonomic instability. In addition to the three subtrees above, the relationship between and placement of *Edingerella* and *Benthosuchus* also becomes uncertain, and, with the removal of *Rileymillerus*, so does the subtree (*Paracyclotosaurus* Watson, 1958, *Mastodonsaurus* Jäger, 1828, *Cyclotosaurus* Fraas, 1889). Interestingly, these taxa all correspond to those identified as least stable in the partitioned-by-islands LS analyses (Supplemental material Tables S2–S6). Taken together, the matrix-based resampling analyses make it clear that

the uncertainty surrounding the lissamphibian relationships is anchored by the *Chinlestegophis-Eocaecilia-Rileymillerus* triad.

As for the *a posteriori* topology-based jackknife analyses, despite its clear effect on the matrix-based analyses, removal of *Chinlestegophis* from the inferred trees does not change the number of MPTs. In fact, *a posteriori* removal of *Chinlestegophis*, *Eocaecilia* and *Rileymillerus* – whether we remove one, two or all three of these taxa – affects only the number of tips on the MPTs, not the number of trees or island structure. Using LS_{max} (Supplemental material Table S1) as the guide for taxon removal order, I had to remove *Eocaecilia*, *Epicrionops*, *Ichthyopsis* and *Karaurus* Ivachnenko, 1978 before the number of trees dropped to 648 and that of 2-RF islands to four. When any external information was ignored and all taxa tested, the taxa whose removal affected tree number the most were *Broomistega*, *Rhineceps* and *Uranocentron*, which correspond to polytomy VI (one of the three areas of instability present in all islands; Serra Silva & Wilkinson, 2021a, table 1). The removal of any one of these taxa decreased the size of the MPT set to 294 trees, removing exactly two-thirds of the original trees, by virtue of breaking the polytomy (Table 3). Thus, with post-inference taxon jackknife we can identify locally unstable taxa but not the taxa causing island structure.

Treatment of polymorphic taxa, missing and inapplicable data

Polymorphic taxa under parsimony. As expected from how the optimization for polymorphic taxa is implemented in PAUP*, choosing the ‘msTaxa = polymorph’ setting affected only tree length, not topology. The MPTs resulting from the polymorphism-aware re-analyses were 18 steps longer than those under default settings, i.e. uncertainty (Table 2). Thus, while some software packages offer multiple settings for the analysis of polymorphic taxa, these settings do not

Table 3. Post-inference first order taxon jackknife. List of taxa whose removal from the set of inferred most parsimonious trees (MPTs) resulted in decrease in MPT number, the change in MPT set size (Δ Trees) and the size of the ensuing set of MPTs. Island number was not affected by taxon removal.

Taxon	Δ Trees	MPT set
<i>Broomistega</i> , <i>Rhineceps</i> , <i>Uranocentron</i>	588	294
<i>Trematosaurus</i>	540	342
<i>Ambystoma</i> , <i>Cryptobranchus</i> , <i>Hynobius</i>	387	495
<i>Acheloma</i> , <i>Broiliellus</i> , <i>Cacops</i> , <i>Dissorophus</i> , <i>Ecolsonia</i> , <i>Phonerpeton</i>	294	588
<i>Amphibamus</i> , <i>Platyrhinops</i>	288	594
<i>Karaurus</i> , <i>Kokartus</i>	234	648
<i>Batrachosuchus</i> , <i>Laidleria</i> , <i>Sangaia</i> , <i>Siderops</i>	162	720
<i>Trematolestes</i>	72	810
<i>Benthosuchus</i> , <i>Edingerella</i>	36	846

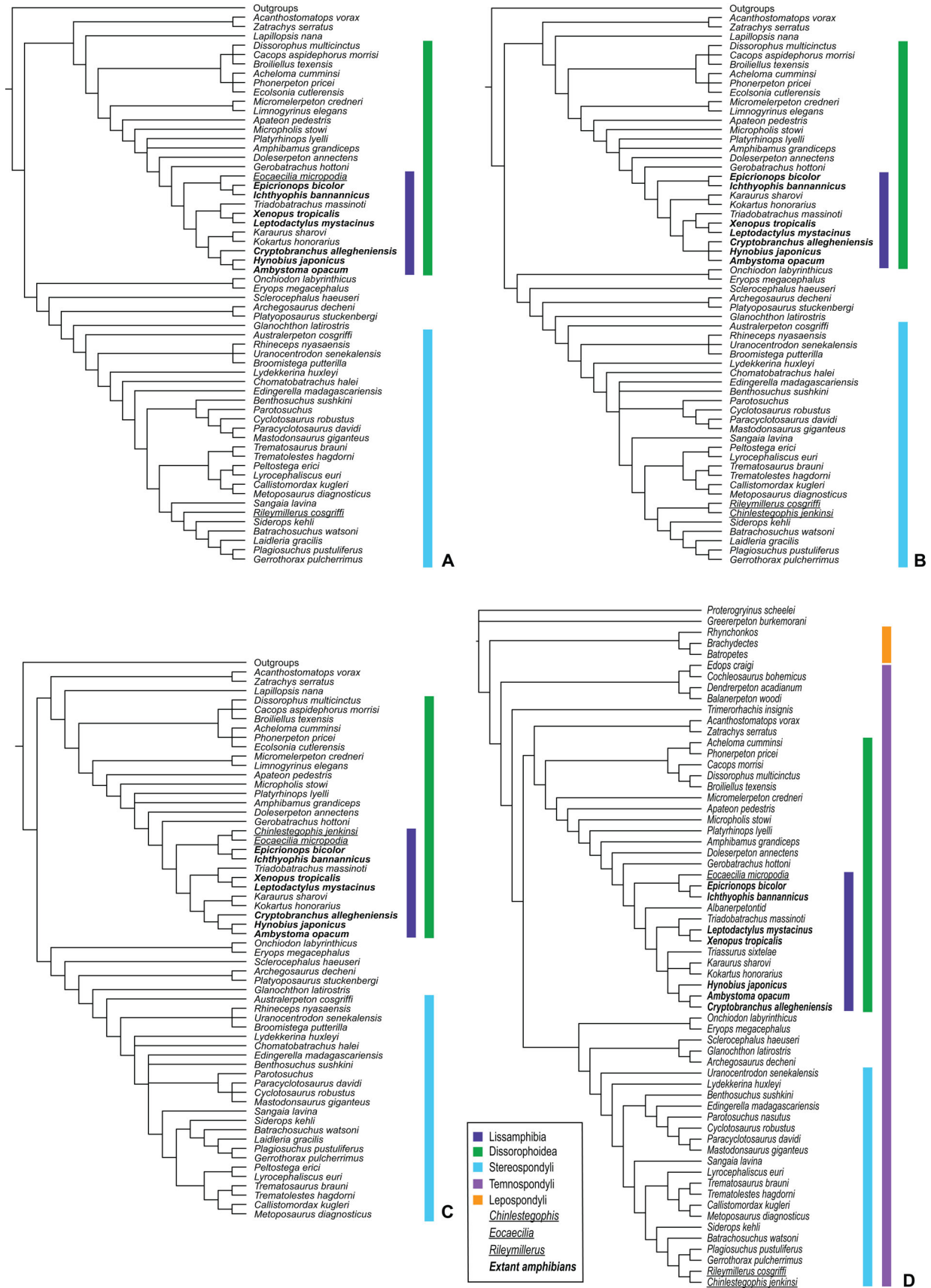


Figure 4. Strict consensus (SC) trees of the first-order taxon jackknife analyses omitting **A**, *Chinlestegophis*, **B**, *Eocaecilia* and **C**, *Rileymillerus*; and **D**, SC of the two most parsimonious trees yielded by analysis of the Schoch et al. (2020) matrix, with higher order taxonomy highlighted.

affect topology and can be ignored when comparing analytical settings, the only exception being if polymorphic taxa are coded as missing data (Nixon & Davis, 1991).

Explicit partitioning scheme in MrBayes. In contrast to the implicitly partitioned analyses (1502 trees), the analyses under the explicit character partitioning scheme yielded a set of 14,871 trees, with an RF = 3 between the MRCs of both analyses. However, unlike the analysis with MrBayes's implicit partition scheme, the set of trees inferred with the explicitly partitioned matrix displays a clear underlying island structure (Fig. 5). Unfortunately, *x*-RF island extraction analyses with the *islandNeighbours* package had to be aborted due to memory restrictions resulting from the size of the tree file. The less memory intensive clump analyses (Serra Silva, 2022), under a break-only approach, identified six clumps, with two recovering a restricted Lissamphibia in Dissorophoidea topology and four the extended Lissamphibia hypothesis. Given these results, it may be worthwhile to run morphological datasets through MrBayes under implicit and explicit partitioning schemes.

Inapplicable data. When Pardo et al.'s (2017) matrix was analysed with 'GapMode = NewState', PAUP* yielded 351 MPTs in two 1-TBR islands (Table 2). Interestingly, the new set of MPTs displays the restricted Lissamphibia hypothesis, but with the clade nested in Stereospondyli not Dissorophoidea, which can

also be seen in island-72 of the default search (Serra Silva & Wilkinson, 2021a, fig. 3b). This placement was found for both the original and rescored matrices (Table 2). Thus, as has been shown for other datasets (e.g. Brazeau et al., 2019; Maddison, 1993), the choice of how to treat inapplicable data clearly affects how many MPTs are found, their length, and, even, the topologies found.

Analysing the original Pardo et al. (2017) matrix with Morphy yielded 585 MPTs (351 of which were also found in the original analysis), all displaying the extended Lissamphibia hypothesis. Using the *xRFislands* function from the *islandNeighbours* R package (Serra Silva & Wilkinson, 2021a), three island structure patterns were identified for the Morphy MPTs. With an RF = 2 threshold the function identifies four islands (72, 108, 162 and 243 trees), at $4 \leq RF \leq 10$ two islands (180 and 405 trees), and at RF = 12 all MPTs are part of a single island. The primary topological differences between the 10-RF islands are the placement of (*Rileymillerus*+*Chinlestegophis*+*Gymnophiona*) within Stereospondyli and the internal relationships of Batrachia, patterns that can also be seen in the two largest 1-TBR islands extracted from the 882 MPT set.

Beyond the treatment of inapplicable data, after the inconsistent/impossible codings were revised and corrected, and the matrix re-analysed under default settings, PAUP* yielded 108 MPTs in two 1-TBR islands, all displaying the restricted Lissamphibia in Dissorophoidea hypothesis (Table 2). These islands correspond to the 1-TBR islands-18 and -90 of the original parsimony analysis (Serra Silva & Wilkinson, 2021a, fig. 3a–c). This means that the extended Lissamphibia hypothesis rests, at least partially, on character construction, not just on analytical choices/parameters.

Thus, the phylogenetic trees inferred from Pardo et al.'s (2017) matrix vary greatly, not only between analyses under different optimality criteria, but also between analyses where the treatment of inapplicable data and polymorphic taxa varies (Table 2).

Differences between the Pardo et al. (2017) and Schoch et al. (2020) data matrices

When investigating the phylogenetic placement of the Triassic stem-salamander *Triassurus sixtelae* Ivakhnenko, 1978, Schoch et al. (2020) used a matrix slightly modified from the Pardo et al. (2017) matrix. They sought to bypass the issues of how to analyse polymorphic and inapplicable data by rescored all such character states as missing ('?'). However, as shown by Nixon & Davis (1991, p. 238) such scoring of polymorphisms as missing data can lead to "failure to find some or all of the [MPTs] that would be found if the

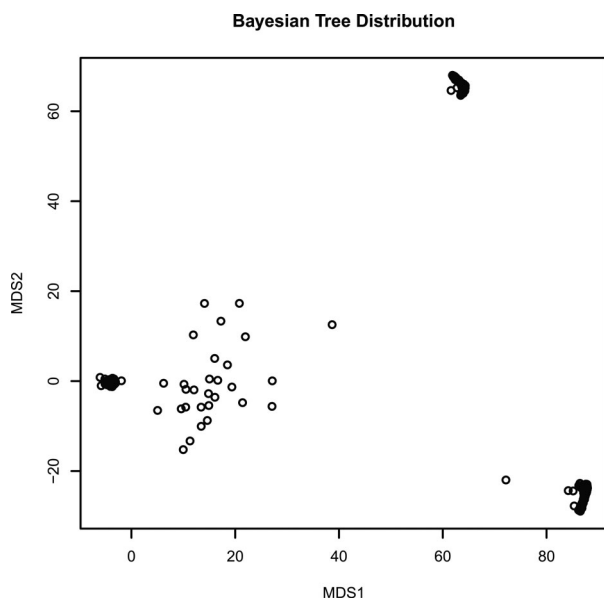


Figure 5. Multidimensional scaling plot of the Bayesian posterior distribution of trees obtained from an explicitly partitioned Pardo et al.'s (2017) matrix.

polymorphisms were scored". Schoch et al. (2020) further modified Pardo et al.'s (2017) matrix by adding 15 new characters, revising the scores for 10 others, and adjusting taxonomic sampling to include a chimaeric albanerpetid, three lepospondyls (*Brachydectes* Cope, 1868, *Batropetes* Carroll & Gaskill, 1971 and *Rhynchonkos* Schultze & Foreman, 1981) and exclude 19 taxa across Temnospondyli (see Schoch et al., 2020, supplementary information). This 62-taxon and 360-character matrix yielded two MPTs that display a restricted Lissamphibia in Dissorophoidea, with albanerpetids sister to Batrachia, and *Chinlestegophis* as sister to *Rileymillerus*, within Stereospondyli (Fig. 4D). Under equal-weights parsimony, resampling analyses found low to moderate support for the restricted Lissamphibia hypothesis (bootstrap = 77%; delete-half jackknife = 59%; Farris jackknife = 67%). Bayesian inference analyses of this matrix were attempted with MrBayes, using varying numbers of independent runs, heated chain temperatures and with or without starting trees. However, based on the recommended values for both the average standard deviation of split frequencies (ASDSF \leq 0.01) and the potential scale reduction factor (PSRF \approx 1.00), none of the analyses converged (longest attempt ran for 200,000,000 generations).

Further parsimony analyses reveal that it is the 15 additional characters in Schoch et al.'s (2020) matrix that are responsible for the recovery of the restricted Lissamphibia, since when these characters were removed the extended Lissamphibia hypothesis was inferred (Table 2). This occurred when analysing either all 62 taxa in Schoch et al.'s (2020) original matrix, or just the 57 taxa shared between Schoch et al.'s (2020) and Pardo et al.'s (2017) matrices. Additionally, while inapplicable data is no longer coded as such (all '-' recoded as '?'), character descriptions and scorings were not revised to accommodate the new characters (see Gee [2022] and Kligman et al. [2023] for a comprehensive revision of the added characters). Thus, hierarchical relationships between characters are not only still present in the matrix but their number increases (Supplemental material Table S7). This is confirmed by the analysis of the rescored matrix yielding shorter MPTs than the original matrix (Table 2), the same pattern found for Pardo et al.'s (2017) dataset. And, just as with the latter dataset, the rescored 345-character matrix no longer recovers an extended Lissamphibia, rather the inferred trees display a restricted Lissamphibia within Dissorophoidea (Table 2). In summary, despite stronger support for the restricted Lissamphibia hypothesis, minor changes to Schoch et al.'s (2020) data matrix can still recover the extended Lissamphibia hypothesis introduced by Pardo et al. (2017).

Discussion

Amphibian systematics have long been controversial, whether it be the relationships between the three extant orders (Anura, Caudata and Gymnophiona), debates over intra-order relationships and rooting (e.g. is Cryptobranchioidea Fitzinger, 1826 or Sirenidae Gray, 1825 the least nested salamander clade: Larson & Dimmick, 1993), or even which group of fossil amphibians did extant amphibians diverge from (lepospondyl vs temnospondyl hypothesis: for an in-depth review see Marjanović & Laurin [2019]). While there is an element of data-availability/taxonomic sampling to many of these conflicting hypotheses, gene trees that support all three possible permutations of the relationships between the extant amphibian orders have been inferred from phylogenomic datasets (e.g. Hime et al., 2021), and, more to the purpose of this study, it has been shown that choice of optimality criterion and data parametrization affect tree inference from both molecular and morphological datasets (e.g. Marjanović & Laurin, 2019; Siu-Ting et al., 2019). Thus, knowing that tree sets inferred from both molecular and morphological data (and from slightly altered datasets) can support multiple hypotheses of amphibian relationships, are we justified in ignoring tree set heterogeneity, particularly in analyses that support yet another novel phylogenetic hypothesis of amphibian relationships? Serra Silva and Wilkinson's (2021a) exploration of how tree islands affect the interpretation of Pardo et al.'s (2017) extended Lissamphibia, and Marjanović & Laurin's (2019) extensive revisions of an early tetrapod data matrix and the phylogenetic hypotheses supported by its analyses under multiple analytical settings, would suggest not.

Starting with tree (multi)set heterogeneity, the presence and effects of tree islands have been explored primarily in the contexts of tree search (e.g. Höhna & Drummond, 2011; Lakner et al., 2008; Olmstead et al., 1993) and consensus tree(s) building (e.g. Maddison, 1991; Serra Silva & Wilkinson, 2021a; Sharkey & Leathers, 2001; Sumrall et al., 2001). In the latter, the phenomenon of large island bias has led to the recognition that the MRC is often a poor summary of (multi)-sets of trees, as the trees in the largest island(s) drown the signal from trees in other islands. This insight urged workers to advocate either for the use of the SC (Sumrall et al., 2001) or weighted consensus (e.g. Serra Silva & Wilkinson, 2021a; Sumrall et al., 2001) if a single summary is desired, or for island-partitioned consensus (e.g. Maddison, 1991; Serra Silva & Wilkinson, 2021a). This idea of summarizing islands, or any such subset, of trees individually rests on the

assumption that there is more variation between islands than within, and that “a single consensus may ignore information in the data” (Hendy et al., 1988, p. 358). And, as shown in Serra Silva and Wilkinson (2021a), the ‘ignore[d] information’ can greatly affect the conclusions made about a group’s relationships and evolutionary history, with most MPTs inferred (under PAUP*’s default settings) from the Pardo et al. (2017) dataset supporting a greatly extended Lissamphibia while most islands supported the traditional restricted Lissamphibia. It is, thus, not foolish to question whether the presence of islands, more specifically large island bias, might also affect post-inference topology-based taxonomic instability tests.

The split frequency-based LS analyses above suggest that island bias is not an issue for the identification of unstable taxa, but that analyses of the complete (multi)-set and each of its islands can be complementary. The results of the analyses on the Pardo et al. (2017) MPTs found that the unpartitioned analyses identify the set of taxa whose placement changes between islands as the least stable taxa (Lissamphibia+*Chinlestegophis*: Supplemental material Table S1), whereas the partitioned-by-island analyses pick up on locally unstable taxa (Supplemental material Tables S2–S6). Thus, the unpartitioned analyses find the taxa responsible for island structure, while partitioned analyses find the taxa (partly) responsible for island size(s), i.e. locally unstable. While any taxon’s identification as unstable in either the partitioned or unpartitioned analyses warrants further investigation, whether the aim is to understand the causes of or decrease topological variation, the identification of a taxon as unstable at the global *and* local levels might be a sensible indicator for its removal.

The consensus optimization-based RBIC analyses, on the other hand, did not find any complementarity between partitioned and unpartitioned analyses, but did yield a potential indicator for the presence of islands, if these are not known *a priori*. RBIC’s default dropset is of one taxon; however, it was shown by Wilkinson and Crotti (2017) that this highly conservative parametrization can lead to the failure to identify all unstable taxa, particularly if they belong to an (internally stable) unstable clade. For the Pardo et al.’s (2017) MPTs, when the dropset is allowed to go up to 15 taxa, the RBIC identifies Lissamphibia+*Chinlestegophis* as the (group of) taxa whose removal shows the greatest improvement to the SC resolution. This result is unsurprising given they are the taxa contributing to island structure but is interesting in that it suggests that the recovery of groups of unstable taxa, rather than single unstable taxa, might be indicative of island presence. As such, while large island bias does not appear to affect

topology-based instability analyses, the latter can be used either to explore taxonomic instability within and between islands (if known), or as an indicator for the possible presence of island structure. Using an RBIC-like method as a test for the potential presence of islands also has the benefit of not requiring the extensive tree-to-tree distance calculations needed to visualize tree space occupation, and thus island structure, with multidimensional scaling (MDS: Fig. 5).

Given that phylogenetic trees are special cases of phylogenetic networks (Huson et al., 2010), and that with a 50% threshold a consensus network is also the MRC (Holland & Moulton, 2003), it is unsurprising that non-tree networks computed from split frequencies are also affected by large island bias. With the two largest, and most topologically similar (Fig. 2B; Serra Silva & Wilkinson, 2021a, fig. 3d), islands making up $\approx 80\%$ of all Pardo et al.’s (2017) MPTs, minor increments in split frequency threshold result in major changes to the consensus networks. With the 20% threshold network displaying a tree-like topology (Fig. 2B), while the 10% threshold consensus consists almost entirely of cycles/3-cubes, resembling an actual net (Fig. 2C). Thus, in the presence of large island bias consensus networks, and any reticulation network inferred from them, can be uninformative and, if the goal is to explore the different evolutionary hypotheses supported by the islands, networks are not an effective alternative to partitioned-by-island consensus trees approaches. While the networks computed for this study might not be representative of the data/island structure underpinning the (multi)sets of trees commonly summarized with consensus networks, it may be beneficial to, prior to any splits-based network analysis, use RBIC-like methods or MDS to ascertain whether island structure is likely to be present.

Tree islands and its effects aside, the Pardo et al. (2017) dataset is particularly interesting due to the myriad ways it can be analysed, even when restricted to heuristic parsimony tree searches (Table 2). Because the data matrix includes inapplicable data and polymorphic taxa it raises the still open questions of how best to analyse these types of character scores (e.g. Brazeau et al., 2019; Nixon & Davis, 1991). Even without addressing the inconsistent use of conventional coding in character construction/descriptions (Supplemental material Table S7), the very different sets of MPTs recovered, when the various ways of parametrizing inapplicable data and polymorphism available in the software used above are compared, show that the existing matrix does not yield robust tree inferences and it should not be used to make definitive statements about lissamphibian relationships. Other algorithms for analyses of and/or codings of polymorphism might yield different results from the tree

lengthening found here; see Nixon and Davis (1991) for a detailed discussion of ways to code for polymorphism. This lack of robust inferences from broadly sampled morphological matrices is not restricted to Pardo et al.'s (2017) data matrix, with Marjanović and Laurin (2019) finding that minor analytical changes to searches on a carefully curated matrix (all characters and their scores revised) for early limbed vertebrates does not yield robust topologies.

The Bayesian analyses of Pardo et al.'s (2017) matrix at first appeared to display less topological variation than the parsimony analyses, with multiple MRCs displaying the extended Lissamphibia hypothesis. However, a closer look at the results of the analysis on the explicitly partitioned data matrix revealed that, much like in the parsimony analyses, Bayesian analyses can recover more than one 'major' topology from the Pardo et al. (2017) data. Additionally, morphological analyses in MrBayes can only be run with the Mk or Mk_v models (Lewis, 2001) and always treat polymorphism as uncertainty (see pp. 11–12 and 85–87 in the manual), which restricts the analyses that can be performed to explore how polymorphic taxa and inapplicable data affect the inferred trees, much as was done in the parsimony context. Also, the performance of Bayesian inference vs parsimony debate (e.g. Goloboff et al., 2018; O'Reilly et al., 2016; Puttick et al., 2017; Wright & Hillis, 2014) aside, recent works have contended that the Mk_v model is not adequate for analyses of morphological datasets (e.g. Goloboff & Arias, 2019; Goloboff et al., 2019), which raises the question of whether Bayesian inference on morphological datasets amounts to an exercise in model misspecification. Yet, there are currently no other identifiable and easily implementable models for morphological evolution available. Recently, Tarasov (2019, 2023) proposed combining knowledge of ontology and developmental processes with structured (SMM: Nodelman et al., 2002) and hidden Markov models (HMM: Beaulieu & O'Meara, 2014) to address the problems of inapplicable data and character construction subjectivity during tree inference. However, it is unclear how scalable this SMM-based approach is, how differently the explicit Mk models and SMMs that condense into a multistate Mk model behave, and how the models behave with empirical data. Thus, there are still many questions surrounding model-based inference of morphological data.

As for character coding, while I looked exclusively at logic violations in the scoring of dependent characters (Supplemental material Table S7), a more thorough review of character construction and coding (akin to the ones seen in Marjanović & Laurin [2019]) may yield yet another set of MPTs and supported lissamphibian

relationships. Along with the ever present possibility of typographical errors (Gee, 2022; Kligman et al., 2023) and the sometimes uncertain homology resulting from bone loss/fusion (e.g. Maddin et al., 2016; Schultze et al., 2008) and nomenclatural variation across taxa (e.g. Abel & Werneburg, 2021; Schultze et al., 2008), Schoch et al.'s (2020) and Marjanović et al.'s (2024) recoding of some character scores, and their reasons, suggests that there are at least some characters whose scoring is not agreed upon by all workers. Also, it has been shown that palaeontologists and neontologists approach character scoring differently (e.g. Harris, 2005; M. Wilkinson, pers. comm., 16 October 2018), which may again influence character scoring and tree inference, given that both extinct and extant taxa are present in the Pardo et al. (2017) and Schoch et al. (2020) matrices. Additionally, one of the extant taxa sampled in these data matrices, *Hynobius japonicus*, is not a valid taxon (Frost, 2023), with no apparent record of this binomial outside of these data matrices, and thus of the source of the character scoring for this 'taxon'. As such, a thorough revision of character construction and scoring of both matrices, based on biological and logical criteria, may be warranted. This need is further supported by Gee's (2022) extensive review of the history of, and the multitude of coding practices and errors accumulated in, the matrices used for the present study. In fact, Kligman et al.'s (2023) completely revised matrix only found support for the restricted Lissamphibia within Dissorophoidea, although it cannot be discounted that increasing/changing taxonomic sampling in this revised matrix might resurrect the extended Lissamphibia hypothesis.

Thus, particular care must be taken when interpreting the results of phylogenetic analyses on taxa like Amphibia, where recalcitrant branches and uncertain relationships abound. From the tree (multi)set heterogeneity angle, the effect of tree islands, particularly large island bias, on tree summaries and instability analyses (and possibly branch support, although further work is needed to confirm this) is an important consideration when exploring (multi)sets of trees and the evolutionary histories they display. And, more fundamentally, the drastic change in island structure/size seen between the various inferences on Pardo et al.'s (2017) matrix is a reminder that phylogenetic analyses are only as good as their underlying data and that analytical settings are not one size fits all.

Acknowledgements

I would like to thank Mark Wilkinson, Mike Benton and Davide Pisani for their supervision, and, alongside

Marc Jones, for their critical commentary of earlier versions of this work. I would further like to thank Mark Wilkinson for enlightening conversations on the large island bias phenomenon, and, along with Marc Jones, for fruitful discussions on amphibian taxonomy and character coding. I would like to thank the Associate Editor, an anonymous reviewer and Bryan Gee for constructive and detailed criticism of the submitted manuscript. This work was supported by the Natural Environment Research Council [NE/L002434/1].

Disclosure statement

No potential conflict of interest was reported by the author(s).

Supplemental material

Supplemental material for this article can be accessed here: <https://doi.org/10.1080/14772019.2024.2321620>. Data is also available from the Dryad Digital Repository: <https://doi.org/10.5068/D1RT1J>.

ORCID

Ana Serra Silva  <http://orcid.org/0000-0001-8020-3227>

References

- Abel, P., & Werneburg, I. (2021). Morphology of the temporal skull region in tetrapods: research history, functional explanations, and a new comprehensive classification scheme. *Biological Reviews*, 96(5), 2229–2257. <https://doi.org/10.1111/brv.12751>
- Aberer, A. J., Krompass, D., & Stamatakis, A. (2012). Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Systematic Biology*, 62(1), 162–166. <https://doi.org/10.1093/sysbio/sys078>
- Anderson, J. S., Reisz, R. R., Scott, D., Fröbisch, N. B., & Sumida, S. S. (2008). A stem batrachian from the Early Permian of Texas and the origin of frogs and salamanders. *Nature*, 453(7194), 515–518. <https://doi.org/10.1038/nature06865>
- Beaulieu, J. M., & O'Meara, B. C. (2014). Hidden Markov models for studying the evolution of binary morphological characters. In L. Z. Garamszegi (Ed.), *Modern phylogenetic comparative methods and their application in evolutionary biology* (pp. 395–408). Springer. <https://doi.org/10.1007/978-3-662-43550-2>
- Bolt, J. R. (1969). Lissamphibian origins: possible protolissamphibian from the Lower Permian of Oklahoma. *Science*, 166(3907), 888–891. <https://doi.org/10.1126/science.166.3907.888>
- Bolt, J. R., & Chatterjee, S. (2000). A new temnospondyl amphibian from the Late Triassic of Texas. *Journal of Paleontology*, 74(4), 670–683. (2000)074 < 0670:ANTAFT > 2.0.CO;2 <https://doi.org/10.1666/0022-3360>
- Boulenger, G. (1883). XXIX.—Description of a new genus of Cœcilæ. *Journal of Natural History*, 11(63), 202–203. <https://doi.org/10.1080/00222938309459129>
- Brazeau, M., & Desjardins, C. (2020). *Morphy* (Version 0.2) [Computer software]. <https://doi.org/10.5281/zenodo.3974824>
- Brazeau, M. D., Guillerme, T., & Smith, M. R. (2019). An algorithm for morphological phylogenetic analysis with inapplicable data. *Systematic Biology*, 68(4), 619–631. <https://doi.org/10.1093/sysbio/syy083>
- Burmeister, H. (1849). Die Labyrinthodonten aus dem bunten Sandstein von Bernburg. G. Reimer, Berlin, iv + 69 pp., 4 pls.
- Cai, L., Xi, Z., Lemmon, E. M., Lemmon, A. R., Mast, A., Buddenhagen, C. E., Liu, L., & Davis, C. C. (2021). The perfect storm: gene tree estimation error, incomplete lineage sorting, and ancient gene flow explain the most recalcitrant ancient angiosperm clade, Malpighiales. *Systematic Biology*, 70(3), 491–507. <https://doi.org/10.1093/sysbio/syaa083>
- Caparros, M., & Prat, S. (2021). A phylogenetic networks perspective on reticulate human evolution. *iScience* 24(4), 102359. <https://www.sciencedirect.com/science/article/pii/S2589004221003278> <https://doi.org/10.1016/j.isci.2021.102359>
- Carroll, R. L., & Gaskill, P. (1971). A captorhinomorph reptile from the Lower Permian of Europe. *Journal of Paleontology*, 45(3), 450–463.
- Chernomor, O., Minh, B. Q., & von Haeseler, A. (2015). Consequences of common topological rearrangements for partition trees in phylogenomic inference. *Journal of Computational Biology* 22(12), 1129–1142. <https://doi.org/10.1089/cmb.2015.0146>
- Colless, D. (1980). Congruence between morphometric and allozyme data for *Menidia* species: a reappraisal. *Systematic Zoology*, 29(3), 288–299. <https://doi.org/10.2307/2412663>
- Cope, E. D. (1868). Synopsis of the extinct Batrachia of North America. *Proceedings of the Academy of Natural Sciences of Philadelphia*, 20, 208–221.
- Cope, E. D. (1882). Third contribution to the history of the Vertebrata of the Permian formation of Texas. *Proceedings of the American Philosophical Society*, 20(112), 447–461.
- Cope, E. D. (1895). A batrachian armadillo. *American Naturalist*, 29(998), 1896a.
- De Laet, J. (2015). Parsimony analysis of unaligned sequence data: maximization of homology and minimization of homoplasy, not minimization of operationally defined total cost or minimization of equally weighted transformations. *Cladistics*, 31(5), 550–567. <https://doi.org/10.1111/cla.12098>
- Dias-Da-Silva, S., & Marsicano, C. (2006). *Sangaia*, a replacement generic name for the rhytidosteid temnospondyl *Cabralia*, a preoccupied name. *Journal of Vertebrate Paleontology*, 26(4), 1004–1004. (2006)26[1004:SARGNF]2.0.CO;2 <https://doi.org/10.1671/0272-4634>

- Dilkes, D. W.** (1990). A new trematopsid amphibian (Temnospondyli: Dissorophoidea) from the Lower Permian of Texas. *Journal of Vertebrate Paleontology*, 10(2), 222–243. <https://doi.org/10.1080/02724634.1990.10011809>
- Dos Santos, C. M., & Falaschi, R. L.** (2007). Missing data in phylogenetic analysis: comments on support measures. *Darwiniana*, 45(Sup), 25–26.
- Efremov, I.** (1937). On the stratification of continental Permian and Triassic the Soviet Union based on the terrestrial vertebrate fauna. *Doklady Akademii Nauk SSSR, Nov. Ser.*, 16(2), 125–132.
- Elworth, R. A. L., Ogilvie, H. A., Zhu, J., & Nakhleh, L.** (2019). Advances in computational methods for phylogenetic networks in the presence of hybridization. In T. Warnow (Ed.), *Bioinformatics and phylogenetics* (pp. 317–360). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-030-10837-3_13
- Farris, J. S., Albert, V. A., Källersjö, M., Lipscomb, D., & Kluge, A. G.** (1996). Parsimony jackknifing outperforms neighbor-joining. *Cladistics*, 12(2), 99–124. <https://doi.org/10.1111/j.1096-0031.1996.tb00196.x>
- Feller, A. E., & Hedges, S. B.** (1998). Molecular evidence for the early history of living amphibians. *Molecular Phylogenetics and Evolution*, 9(3), 509–516. <https://doi.org/10.1006/mpev.1998.0500>
- Felsenstein, J.** (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39(4), 783–791. <https://doi.org/10.2307/2408678>
- Fisher von Waldheim, G.** (1813), *Zoognosia tabulis synopticis illustrata: in usum praelectionum Academiae imperialis medico-chirurgicae mosquensis edita*, Typis Nicolai S. Vsevolozsky, Moscow, xiv + 465 pp.
- Fitzinger, L. J.** (1826). Neue Classification der Reptilien nach ihren Natürlichen Verwandtschaften: Nebst einer Verwandtschafts-Tafel und einem Verzeichnisse der Reptilien-Sammlung des K. K. Zoologischen Museums zu Wien. J. G. Heubner.
- Forey, P. L., & Kitching, I. J.** (2000). Experiments in coding multistate characters. In R. Scotland & R. T. Pennington (Eds.), *Homology and systematics: Coding characters for phylogenetic analyses* (pp. 54–80). CRC Press.
- Fox, R. C., & Naylor, B. G.** (1982). A reconsideration of the relationships of the fossil amphibian *Albanerpeton*. *Canadian Journal of Earth Sciences*, 19(1), 118–128. <https://doi.org/10.1139/e82-009>
- Fraas, E.** (1889). Die Labyrinthodonten der schwäbischen Trias. *Palaeontographica (1846–1933)*, 36, 1–158.
- Frost, D. R.** 2023. Amphibian species of the world: an online reference (Version 6.2) [Data set]. Retrieved November 18, 2023, from <https://amphibiansoftheworld.amnh.org/index.php>. American Museum of Natural History, New York, USA. <https://doi.org/10.5531/db.vz.0001>
- Frost, D. R., Grant, T., Faivovich, J., Bain, R. H., Haas, A., Haddad, C. F. B., De Sa, R. O., Channing, A., Wilkinson, M., Donnellan, S. C., Raxworthy, C. J., Campbell, J. A., Blotto, B. L., Moler, P., Drewes, R. C., Nussbaum, R. A., Lynch, J. D., Green, D. M., & Wheeler, W. C.** (2006). The amphibian tree of life. *Bulletin of the American Museum of Natural History*, 297, 1–291. [https://doi.org/10.1206/0003-0090\(2006\)297\[0001:TATOL\]2.0.CO;2](https://doi.org/10.1206/0003-0090(2006)297[0001:TATOL]2.0.CO;2)
- Gardner, J. D.** (2001). Monophyly and affinities of albanerpetontid amphibians (Temnospondyli; Lissamphibia). *Zoological Journal of the Linnean Society*, 131(3), 309–352. <https://doi.org/10.1111/j.1096-3642.2001.tb02240.x>
- Gee, B. M.** (2022). The disadvantage of derivation: conserved systematic flaws in primary data have repeatedly biased the phylogenetic inference of Temnospondyli (Tetrapoda, Amphibia). *bioRxiv preprint*. <https://doi.org/10.1101/2022.06.22.496729>
- Goloboff, P. A., & Arias, J. S.** (2019). Likelihood approximations of implied weights parsimony can be selected over the Mk model by the Akaike information criterion. *Cladistics*, 35(6), 695–716. <https://doi.org/10.1111/cla.12380>
- Goloboff, P. A., De Laet, J., Ríos-Tamayo, D., & Szumik, C. A.** (2021). A reconsideration of inapplicable characters, and an approximation with step-matrix recoding. *Cladistics* 37(5), 596–629. <https://doi.org/10.1111/cla.12456>
- Goloboff, P. A., Pittman, M., Pol, D., & Xu, X.** (2019). Morphological data sets fit a common mechanism much more poorly than DNA sequences and call into question the Mk model. *Systematic Biology*, 68(3), 494–504. <https://doi.org/10.1093/sysbio/syy077>
- Goloboff, P. A., Torres, A., & Arias, J. S.** (2018). Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. *Cladistics*, 34(4), 407–437. <https://doi.org/10.1111/cla.12205>
- Goloboff, P. A., Torres Galvis, A., & Arias, J. S.** (2018). Parsimony and model-based phylogenetic methods for morphological data: Comments on O'Reilly et al. *Palaeontology*, 61(1), 625–630. <https://doi.org/10.1111/pala.12353>
- Gray, J. E.** (1825). A synopsis of the genera of reptiles and Amphibia, with a description of some new species. *Annals of Philosophy*, 10, 193–217.
- Haeckel, E.** (1866). *Generelle Morphologie der Organismen (Volume 2)*. G. Reimer, Berlin, clx + 462 pp.
- Harris, S. R.** (2005). *Character construction in morphological phylogenetics and the affinities of turtles* [PhD thesis]. University of Bristol.
- Hawkins, J. A., Hughes, C. E., & Scotland, R. W.** (1997). Primary homology assessment, characters and character states. *Cladistics*, 13(3), 275–283. <https://doi.org/10.1111/j.1096-0031.1997.tb00320.x>
- Hendy, M. D., Steel, M. A., Penny, D., & Henderson, I. M.** (1988). Families of trees and consensus. In H. H. Bock (Ed.), *Classification and related methods of data analysis* (pp. 355–362). Elsevier.
- Hime, P. M., Lemmon, A. R., Lemmon, E. C. M., Prendini, E., Brown, J. M., Thomson, R. C., Kratovil, J. D., Noonan, B. P., Pyron, R. A., Peloso, P. L. V., Kortyna, M. L., Keogh, J. S., Donnellan, S. C., Mueller, R. L., Raxworthy, C. J., Kunte, K., Ron, S. R., Das, S., Gaitonde, N., ..., Weisrock, D. W.** (2021). Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. *Systematic Biology*, 70(1), 49–66. <https://doi.org/10.1093/sysbio/syaa034>
- Höhna, S., & Drummond, A. J.** (2011). Guided tree topology proposals for Bayesian phylogenetic inference. *Systematic Biology*, 61(1), 1–11. <https://doi.org/10.1093/sysbio/syr074>
- Holland, B., & Moulton, V.** (2003). Consensus networks: a method for visualising incompatibilities in collections of trees. In G. Benson & R. D. M. Page (Eds.), *International*

- workshop on algorithms in bioinformatics* (pp. 165–176). Springer.
- Hopkins, M. J., & St. John, K. (2021).** Incorporating hierarchical characters into phylogenetic analysis. *Systematic Biology*, 70(6), 1163–1180. <https://doi.org/10.1093/sysbio/syab005>
- Huson, D. H., & Bryant, D. (2006).** Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2), 254–267. <https://doi.org/10.1093/molbev/msj030>
- Huson, D. H., & Klopper, T. H. (2007).** Beyond galled trees – decomposition and computation of galled networks. In T. Speed & H. Huang (Eds.), *Annual international conference on research in computational molecular biology RECOMB 2007* (pp. 211–249), Springer.
- Huson, D. H., Klöpper, T., Lockhart, P. J., & Steel, M. A. (2005).** Reconstruction of reticulate networks from gene trees. p. 233–249. In: S. Miyano, J. Merisov, S. Kasif, S. Istrail, P. A. Pevzner, & M. Waterman (Eds.). *Annual international conference on research in computational molecular biology RECOMB 2005*, Springer.
- Huson, D. H., Rupp, R., & Scornavacca, C. (2010).** *Phylogenetic networks: concepts, algorithms and applications*. Cambridge University Press.
- Ivachnenko, M. (1978a).** Urodeles from the Triassic and Jurassic of Soviet Central Asia. *Paleontological Journal*, 12, 362–368.
- Ivakhnenko, M. (1978b).** Caudates from the Triassic and Jurassic of Middle Asia. *Paleontologicheskii Zhurnal*, 3, 84–89.
- Jäger, G. F. (1828).** Über die Fossile Reptilien, welche in Württemberg aufgefunden worden sind. J. B. Metzler.
- Jenkins Jr, P. A., & Walsh, D. M. (1993).** An Early Jurassic caecilian with limbs. *Nature* 365(6443), 246–250. <https://doi.org/10.1038/365246a0>
- Kitching, J. (1957).** A new small stereospondylous labyrinthodont from the Triassic beds of South Africa. *Palaeontology Africana*, 5, 67–82.
- Kligman, B. T., Gee, B. M., Marsh, A. D., Nesbitt, S. J., Smith, M. E., Parker, W. G., & Stocker, M. R. (2023).** Triassic stem caecilian supports dissorophoid origin of living amphibians. *Nature*, 614, 1–6. <https://doi.org/10.1038/s41586-022-05646-5>
- Lakner, C., van der Mark, P., Huelsenbeck, J. P., Larget, B., & Ronquist, F. (2008).** Efficiency of Markov chain Monte Carlo tree proposals in Bayesian phylogenetics. *Systematic Biology*, 57(1), 86–103. <https://doi.org/10.1080/10635150801886156>
- Lanyon, S. M. (1985).** Detecting internal inconsistencies in distance data. *Systematic Zoology*, 34(4), 397–403. <https://doi.org/10.2307/2413204>
- Larson, A., & Dimmick, W. W. (1993).** Phylogenetic relationships of the salamander families: an analysis of congruence among morphological and molecular characters. *Herpetological Monographs*, 7, 77–93. <https://doi.org/10.2307/1466953>
- Latreille, P. A. (1800).** Histoire naturelle des salamandres de France: précédée d'un tableau méthodique des autres reptiles indigènes. Chez Villier.
- Laurin, M., & Soler-Gijón, R. (2001).** The oldest stegocephalian from the Iberian Peninsula: evidence that temnospondyls were euryhaline. *Comptes Rendus de l'Académie des Sciences-Series III-Sciences de la Vie*, 324(5), 495–501. [https://doi.org/10.1016/s0764-4469\(01\)01318-x](https://doi.org/10.1016/s0764-4469(01)01318-x)
- Leuckart, F. (1821).** Einiges über die fischartigen Amphibien. *Isis von Oken*, 9, 257–265.
- Lewis, P. O. (2001).** A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, 50(6), 913–925. <https://doi.org/10.1080/106351501753462876>
- Maddin, H. C., Piekarski, N., Sefton, E. M., & Hanken, J. (2016).** Homology of the cranial vault in birds: new insights based on embryonic fate-mapping and character analysis. *Royal Society Open Science*, 3(8), 160356. <https://doi.org/10.1098/rsos.160356>
- Maddison, D. R. (1991).** The discovery and importance of multiple islands of most-parsimonious trees. *Systematic Biology*, 40(3), 315–328. <https://doi.org/10.1093/sysbio/40.3.315>
- Maddison, W. P. (1993).** Missing data versus missing characters in phylogenetic analysis. *Systematic Biology*, 42(4), 576–581. <https://doi.org/10.1093/sysbio/42.4.576>
- Margush, T., & McMorris, F. (1981).** Consensus n-trees. *Bulletin of Mathematical Biology*, 43(2), 239–244. [https://doi.org/10.1016/S0092-8240\(81\)90019-7](https://doi.org/10.1016/S0092-8240(81)90019-7)
- Marjanović, D., & Laurin, M. (2019).** Phylogeny of Paleozoic limbed vertebrates reassessed through revision and expansion of the largest published relevant data matrix. *PeerJ*, 6, e5565. <https://doi.org/10.7717/peerj.5565>
- Marjanović, D., Maddin, H. C., Olori, J. C., & Laurin, M. (2024).** The new problem of *Chinlestegophis* and the origin of caecilians (Amphibia, Gymnophionomorpha) is highly sensitive to old problems of sampling and character construction. *Fossil Record*, 27(1), 55–94. <https://doi.org/10.3897/fr.27.e109555>
- Müller, J. (1831).** Beiträge zur Anatomie und Naturgeschichte der Amphibien. *Zeitschrift für Physiologie*, 4, 190–275.
- Nixon, K. C., & Davis, J. I. (1991).** Polymorphic taxa, missing values and cladistic analysis. *Cladistics*, 7(3), 233–241. <https://doi.org/10.1111/j.1096-0031.1991.tb00036.x>
- Nodelman, U., Shelton, C. R., & Koller, D. (2002).** Continuous time Bayesian networks. In A. Darwiche & N. Friedman (Eds.), *UAI02: Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence* (pp. 378–387). Morgan Kaufmann Publishers Inc.
- Olmstead, R. G., Bremer, B., Scott, K. M., & Palmer, J. D. (1993).** A parsimony analysis of the Asteridae *sensu lato* based on rbcL sequences. *Annals of the Missouri Botanical Garden*, 80(3), 700–722. <https://doi.org/10.2307/2399855>
- O'Reilly, J. E., Puttick, M. N., Parry, L., Tanner, A. R., Tarver, J. E., Fleming, J., Pisani, D., & Donoghue, P. C. (2016).** Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biology Letters*, 12(4), 20160081. <https://doi.org/10.1098/rsbl.2016.0081>
- Pardo, J. D., Small, B. J., & Huttenlocker, A. K. (2017).** Stem caecilian from the Triassic of Colorado sheds light on the origins of Lissamphibia. *Proceedings of the National Academy of Sciences*, 114(27), E5389–E5395. <https://doi.org/10.1073/pnas.1706752114>
- Pattengale, N., Aberer, A., Swenson, K., Stamatakis, A., & Moret, B. (2011).** Uncovering hidden phylogenetic consensus in large data sets. *IEEE/ACM Transactions on*

- Computational Biology and Bioinformatics*, 8(4), 902–911. <https://doi.org/10.1109/TCBB.2011.28>
- Platnick, N. I., Griswold, C. E., & Coddington, J. A.** (1991). On missing entries in cladistic analysis. *Cladistics*, 7(4), 337–343. <https://doi.org/10.1111/j.1096-0031.1991.tb00042.x>
- Puttick, M. N., O'Reilly, J. E., Tanner, A. R., Fleming, J. F., Clark, J., Holloway, L., Lozano-Fernandez, J., Parry, L. A., Tarver, J. E., Pisani, D., & Donoghue, P. C. J.** (2017). Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proceedings of the Royal Society B: Biological Sciences*, 284(1846), 20162290. <https://doi.org/10.1098/rspb.2016.2290>
- Pyron, R. A., & Wiens, J. J.** (2011). A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution*, 61(2), 543–583. <https://doi.org/10.1016/j.ympev.2011.06.012>
- R Core Team (2021). *R: a language and environment for statistical computing*, R Foundation for Statistical Computing. <https://www.R-project.org/>
- Robinson, D. F., & Foulds, L. R.** (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*, 53(1–2), 131–147. [https://doi.org/10.1016/0025-5564\(81\)90043-2](https://doi.org/10.1016/0025-5564(81)90043-2)
- Romer, A.** (1969). A temnospondylous amphibian from the Lower Carboniferous. *Kirtlandia*, 6, 1–20.
- Romer, A. S.** (1970). A new anthracosaurian labyrinthodont, *Proterogyrinus scheelei*, from the Lower Carboniferous. *Kirtlandia*, 10, 1–16.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P.** (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), 539–542. <https://doi.org/10.1093/sysbio/sys029>
- Ruta, M., & Coates, M. I.** (2007). Dates, nodes and character conflict: addressing the lissamphibian origin problem. *Journal of Systematic Palaeontology*, 5(1), 69–122. <https://doi.org/10.1017/S1477201906002008>
- San Mauro, D., Gower, D. J., Müller, H., Loader, S. P., Zardoya, R., Nussbaum, R. A., & Wilkinson, M.** (2014). Life-history evolution and mitogenomic phylogeny of caecilian amphibians. *Molecular Phylogenetics and Evolution*, 73, 177–189. <https://doi.org/10.1016/j.ympev.2014.01.009>
- Sanderson, M. J., McMahon, M. M., & Steel, M.** (2011). Terraces in phylogenetic tree space. *Science*, 333(6041), 448–450. <https://doi.org/10.1126/science.1206357>
- Santos, R. O., Laurin, M., & Zaher, H.** 2020. A review of the fossil record of caecilians (Lissamphibia: Gymnophionomorpha) with comments on its use to calibrate molecular timetrees. *Biological Journal of the Linnean Society*, 131(4), 737–755. <https://doi.org/10.1093/biolinnean/blaa148>
- Schoch, R. R.** (2006). A complete trematosaurid amphibian from the Middle Triassic of Germany. *Journal of Vertebrate Paleontology*, 26(1), 29–43. [https://doi.org/10.1671/0272-4634\(2006\)26\[29:ACTAFT\]2.0.CO;2](https://doi.org/10.1671/0272-4634(2006)26[29:ACTAFT]2.0.CO;2)
- Schoch, R. R., Werneburg, R., & Voigt, S.** (2020). A Triassic stem salamander from Kyrgyzstan and the origin of salamanders. *Proceedings of the National Academy of Sciences*, 117(21), 11584–11588. <https://doi.org/10.1073/pnas.2001424117>
- Schultze, H.-P., Arratia, G., & Wilson, M. V. H.** (2008). Nomenclature and homologization of cranial bones in actinopterygians. In G. Arratia, H.-P. Schultze & V. H. Wilson (Eds.), *Mesozoic fishes 4: homology and phylogeny* (pp. 23–48). Verlag Dr. Friedrich Pfeil.
- Schultze, H.-P., & Foreman, B.** (1981). A new gymnarthrid microsauro from the Lower Permian of Kansas with a review of the tuditanomorph microsaur (Amphibia). *Occasional Papers of the Museum of Natural History, the University of Kansas*, 91, 1–25.
- Serra Silva, A.** (2022). *Post-processing of phylogenetic trees: On islands, clumps and (non-) effective overlap* [PhD thesis]. University of Bristol. <https://hdl.handle.net/1983/1e17b53f-107a-4ee8-9cea-9fe8ac2589e5>
- Serra Silva, A., & Wilkinson, M.** (2021a). On defining and finding islands of trees and mitigating large island bias. *Systematic Biology*, 70(6), 1282–1294. <https://doi.org/10.1093/sysbio/syab015>
- Serra Silva, A., & Wilkinson, M.** (2021b). *On defining and finding islands of trees and mitigating large island bias* [Data set]. Dryad. <https://doi.org/10.5068/D14X10>
- Sharkey, M. J., & Leathers, J. W.** (2001). Majority does not rule: the trouble with majority-rule consensus trees. *Cladistics*, 17(3), 282–284. <https://doi.org/10.1006/clad.2001.0174>
- Shishkin, M., & Rubidge, B.** (2000). A relict rhinesuchid (Amphibia: Temnospondyli) from the Lower Triassic of South Africa. *Palaeontology*, 43(4), 653–670. <https://doi.org/10.1111/1475-4983.00144>
- Simões, T. R., Vernygora, O. V., de Medeiros, B. A. S., & Wright, A. M.** (2023). Handling logical character dependency in phylogenetic inference: extensive performance testing of assumptions and solutions using simulated and empirical data. *Systematic Biology*, 72(3), 662–680. <https://doi.org/10.1093/sysbio/syad006>
- Siu-Ting, K., Pisani, D., Creevey, C. J., & Wilkinson, M.** (2015). Concatabominations: identifying unstable taxa in morphological phylogenetics using a heuristic extension to safe taxonomic reduction. *Systematic Biology*, 64(1), 137–143. <https://doi.org/10.1093/sysbio/syu066>
- Siu-Ting, K., Torres-Sánchez, M., San Mauro, D., Wilcockson, D., Wilkinson, M., Pisani, D., O'Connell, M. J., & Creevey, C. J.** (2019). Inadvertent paralog inclusion drives artifactual topologies and timetree estimates in phylogenomics. *Molecular Biology and Evolution*, 36(6), 1344–1356. <https://doi.org/10.1093/molbev/msz067>
- Sokal, R. R., & Rohlf, F. J.** (1981). Taxonomic congruence in the Leptopodomorpha re-examined. *Systematic Zoology*, 30(3), 309–325. <https://doi.org/10.2307/2413252>
- Sumrall, C. D., Brochu, C. A., & Merck, J. W.** (2001). Global lability, regional resolution, and majority-rule consensus bias. *Paleobiology*, 27(2), 254–261. (2001)0272.0.CO;2 <https://doi.org/10.1666/0094-8373>
- Suvorov, A., Kim, B. Y., Wang, J., Armstrong, E. E., Peede, D., D'Agostino, E. R. R., Price, D. K., Waddell, P. J., Lang, M., Courtier-Orgogozo, V., David, J. R., Petrov, D., Matute, D. R., Schrider, D. R., & Comeault, A. A.** (2022). Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *Current Biology*, 32(1), 111–123. <https://doi.org/10.1016/j.cub.2021.10.052>
- Swofford, D. L.** (2003). *PAUP*: phylogenetic analysis using parsimony (*and other methods)* (Version 4.0 a165). Sinauer Associates.

- Tarasov, S. (2019).** Integration of anatomy ontologies and evo-devo using structured Markov models suggests a new framework for modelling discrete phenotypic traits. *Systematic Biology*, 68(5), 698–716. <https://doi.org/10.1093/sysbio/syz005>
- Tarasov, S. (2023).** New phylogenetic Markov models for inapplicable morphological characters. *Systematic Biology*, 72(3), 681–693. <https://doi.org/10.1093/sysbio/syad005>
- Thorley, J. L. (2000).** *Cladistic information, leaf stability and supertree construction* [PhD thesis]. University of Bristol.
- Thorley, J. L., & Wilkinson, M. (1999).** Testing the phylogenetic stability of early tetrapods. *Journal of Theoretical Biology*, 200(3), 343. <https://doi.org/10.1006/jtbi.1999.0999>
- Tschudi, J. J. (1838).** Classification der Batrachier: mit Berücksichtigung der Fossilen Thiere dieser Abtheilung der Reptilien. Petitpierre.
- van Hoepen, E. (1915).** Stegocephalia of Senekal, O.F.S. *Annals of the Transvaal Museum*, 5(2), 125–149.
- Vaughn, P. P. (1969).** Further evidence of close relationship of the trematopsid and dissorophoid labyrinthodont amphibians with a description of a new genus and new species. *Bulletin, Southern California Academy of Sciences*, 68(3), 121–130.
- von Meyer, H. (1844).** Briefliche mittheilung an Prof. Bronn. Neues Jahrbuch für Mineralogie, Geognosie, Geologie und Petrefaktenkunde, 1844, 336–337.
- von Zittel, K. A. (1887).** Handbuch der Paläontologie, Paläozoologie, Vertebrata (Pisces, Amphibia, Reptilia, Aves) (Volume 3). R. Oldenbourg.
- Wagler, J. (1827).** Untitled footnote *Isis von Oken*, 20, 726.
- Warren, A., & Hutchinson, M. (1990).** *Lapillopsis*, a new genus of temnospondyl amphibians from the Early Triassic of Queensland. *Alcheringa*, 14(2), 149–158. <https://doi.org/10.1080/03115519008527816>
- Watson, D. M. S. (1958).** A new labyrinthodont (*Paracyclotosaurus*) from the Upper Trias of New South Wales. *Bulletin of the British Museum of Natural History, London (Geology)*, 3, 233–263.
- Watson, D. M. S. (1962).** The evolution of the labyrinthodonts. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 245(723), 219–265.
- Wilkinson, M. (1995a).** Coping with abundant missing entries in phylogenetic inference using parsimony. *Systematic Biology*, 44(4), 501–514. <https://doi.org/10.2307/2413657>
- Wilkinson, M. (1995b).** More on reduced consensus methods. *Systematic Biology*, 44(3), 435–439. <https://doi.org/10.2307/2413604>
- Wilkinson, M. (1995c).** A comparison of two methods of character construction. *Cladistics*, 11(3), 297–308. [https://doi.org/10.1016/0748-3007\(95\)90017-9](https://doi.org/10.1016/0748-3007(95)90017-9)
- Wilkinson, M. (2003).** Missing entries and multiple trees: instability, relationships, and support in parsimony analysis. *Journal of Vertebrate Paleontology*, 23(2), 311–323. [https://doi.org/10.1671/0272-4634\(2003\)023\[0311:MEAMTI\]2.0.CO;2](https://doi.org/10.1671/0272-4634(2003)023[0311:MEAMTI]2.0.CO;2)
- Wilkinson, M. (2006).** Identifying stable reference taxa for phylogenetic nomenclature. *Zoologica Scripta*, 35(1), 109–112. <https://doi.org/10.1111/j.1463-6409.2005.00213.x>
- Wilkinson, M., & Crotti, M. (2017).** Comments on detecting rogue taxa using RogueNaRok. *Systematics and Biodiversity*, 15(4), 291–295. <https://doi.org/10.1080/14772000.2016.1252440>
- Williston, S. (1910).** *Cacops desmospondylus*; new genera of Permian vertebrates. *Bulletin of the Geological Society of America*, 21(1), 249–284. <https://doi.org/10.1130/GSAB-21-249>
- Williston, S. W. (1914).** *Broiliellus*, a new genus of amphibians from the Permian of Texas. *The Journal of Geology*, 22(1), 49–56. <https://doi.org/10.1086/622132>
- Wright, A. M., & Hillis, D. M. (2014).** Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS One*, 9(10), e109210. <https://doi.org/10.1371/journal.pone.0109210>

Associate Editor: Jennifer Olori