WILEY-BLACKWELL

Cladistics 37 (2021) 375–401

Cladistics

10.1111/cla.12447

Total evidence and sensitivity phylogenetic analyses of egg-brooding frogs (Anura: Hemiphractidae)

Lourdes Y. Echevarría*[a,b], Ignacio De la Riva[c], Pablo J. Venegas[b], Fernando J.M. Rojas-Runjaic[d], Iuri R. Dias[e] and Santiago Castroviejo-Fisher[a,f]

[a]*Laboratório de Sistemática de Vertebrados, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Av. Ipiranga 6681, Porto Alegre, RS, 90619-900, Brazil;* [b]*División de Herpetología-Centro de Ornitología y Biodiversidad (CORBIDI), Urb. Huertos de San Antonio, Santa Rita No. 105 Of. 202, Surco, Lima, Perú;* [c]*Museo Nacional de Ciencias Naturales-CSIC, C/José Gutiérrez Abascal 2, Madrid, 28006, Spain;* [d]*Fundación La Salle de Ciencias Naturales, Museo de Historia Natural La Salle, Caracas, 1050, Venezuela;* [e]*Graduate Program in Zoology, Universidade Estadual de Santa Cruz, Rodovia Jorge Amado, km 16, Ilhéus, Bahia, 45662-900, Brazil;* [f]*Department of Herpetology, American Museum of Natural History, New York, NY, 10024, USA*

Accepted 10 November 2020

Abstract

We study the phylogenetic relationships of egg-brooding frogs, a group of 118 neotropical species, unique among anurans by having embryos with large bell-shaped gills and females carrying their eggs on the dorsum, exposed or inside a pouch. We assembled a total evidence dataset of published and newly generated data containing 51 phenotypic characters and DNA sequences of 20 loci for 143 hemiphractids and 127 outgroup terminals. We performed six analytical strategies combining different optimality criteria (parsimony and maximum likelihood), alignment methods (tree- and similarity-alignment), and three different indel coding schemes (fifth character state, unknown nucleotide, and presence/absence characters matrix). Furthermore, we analyzed a subset of the total evidence dataset to evaluate the impact of phenotypic characters on hemiphractid phylogenetic relationships. Our main results include: (i) monophyly of Hemiphractidae and its six genera for all our analyses, novel relationships among hemiphractid genera, and non-monophyly of Hemiphractinae according to our preferred phylogenetic hypothesis; (ii) non-monophyly of current supraspecific taxonomies of *Gastrotheca*, an updated taxonomy is provided; (iii) previous differences among studies were mainly caused by differences in analytical factors, not by differences in character/taxon sampling; (iv) optimality criteria, alignment method, and indel coding caused differences among optimal topologies, in that order of degree; (v) in most cases, parsimony analyses are more sensitive to the addition of phenotypic data than maximum likelihood analyses; (vi) adding phenotypic data resulted in an increase of shared clades for most analyses.

Introduction

Egg-brooding frogs (Hemiphractidae) are non-aquatic Neotropical anurans, with 118 described species, grouped in six genera: *Cryptobatrachus*, *Flectonotus*, *Fritziana*, *Gastrotheca*, *Hemiphractus*, and *Stefania* (Frost, 2020). Their distribution includes a wide variety of habitats, from Neotropical lowlands or montane rainforests to humid high-elevation Andean grasslands and the tepuis of the Guiana Shield (Castroviejo-Fisher et al., 2015; Duellman, 2015). The reproductive biology of hemiphractids is not only unique but also diverse. For instance, their common name derives from the fact that females carry their eggs on their backs, either inside a pouch or exposed on the dorsum (Del Pino, 1980). A variety of developmental modes are present among the members of this family, from direct development in species of *Cryptobatrachus*, *Hemiphractus*, and *Stefania*, to development with a free-living tadpole phase in species of *Flectonotus* and *Fritziana*, and some species presenting direct

*Corresponding author:
E-mail address: lourdese.20@gmail.com

development and other aquatic larvae in *Gastrotheca* (Duellman, 2015).

In 2015, two independent studies revisited the evolutionary relationships of egg-brooding frogs (Castroviejo-Fisher et al., 2015; Duellman, 2015). Both studies included a broad, although different, sampling of hemiphractids (more than 70% of the species), but a markedly different outgroup sampling. While Duellman (2015) included only three nobleobatrachian outgroup terminals, Castroviejo-Fisher et al. (2015) included 127 outgroup terminals with representatives of all nobleobatrachian families recognized at the time. The analytical approaches of both studies were also broadly different. The phylogeny presented by Castroviejo-Fisher et al. (2015) was based on a total-evidence parsimony analysis under tree-alignment of DNA sequences (up to 20 markers, including mitochondrial and nuclear genes) combined with 51 phenotypic characters. The phylogeny presented by Duellman (2015) was based on maximum likelihood and Bayesian analyses of a fixed-alignment of DNA sequences of two mitochondrial (16S and ND1) and two nuclear (RAG-1 and POMC) markers. Both studies agreed on the monophyly of Hemiphractidae and its six genera (although Duellman, 2015 only included one terminal of *Cryptobatrachus*), but differed most noticeably on the sister relationship of the family, relationships among genera, and among infrageneric taxa of *Gastrotheca* (Fig. 1). This later conflict implies that several subgenera of Duellman (2015) were non-monophyletic within the paradigm of Castroviejo-Fisher et al. (2015). Although Duellman (2015) presented an additional phylogeny of combined molecular and osteological data, information on which characters were included and how they were coded was missing so that repeatability and replicability were not possible. More recently, Kok et al. (2017) published a detailed study on the biogeography of *Stefania* including a more comprehensive taxon sampling of the genus than either Castroviejo-Fisher et al. (2015) or Duellman (2015). Also, Walker et al. (2018a) published a comprehensive study about the genetic diversity and phylogenetic relationships of *Fritziana* greatly increasing taxon sampling within this genus.

Conflicting phylogenetic hypothesis, as in hemiphractids, deserve detailed comparisons to evaluate the qualitative and quantitative impact of different analytical factors (e.g., optimality criteria, taxon sampling, intensity of tree searches). Adding to the difficulty created by the variety of results that stem from using different characters and combinations of analytical factors is the hierarchical nature of phylogenetic hypotheses, which leads to marked differences in taxon sampling among studies depending on which taxon is considered the ingroup and how the outgroup is chosen (Nixon and Carpenter, 1993; Grant, 2019). The continuous reevaluation of phylogenetic hypotheses and the necessity of repeatability, replicability, and maximization of explanatory power leads to "total evidence" analysis (Kluge, 1989, 1997, 1998; Eernisse and Kluge, 1993)—"supermatrix-analysis", "simultaneous-analysis", "combined-analysis" or "concatenated-analysis" can be performed in a total evidence frame (de Queiroz and Gatesy, 2007). Thus, revisionary approaches that combine legacy empirical data derived from independent studies of partially overlapping datasets are crucial to summarize our state of understanding of the phylogenetic relationships among any group of organisms (Driskell et al., 2004; Gatesy et al., 2004; Frost et al., 2006; Pyron and Wiens, 2011; Padial et al., 2014; Castroviejo-Fisher et al., 2015; Goicoechea et al., 2016; Peloso et al., 2016). Nonetheless, combining characters and terminals from independent studies often result in datasets with large amounts of missing data (often >60%; Simmons and Goloboff, 2014).

Although a phylogenetic total evidence analysis allows evaluation of differences among previous competing hypotheses due to character and taxon sampling, it is mute about the effect that other analytical factors may have played in selecting incompatible optimal topologies. Sensitivity analysis, in which each parameter is varied while all others are fixed, to isolate the effects of different parameters, is a powerful experimental approach to attempting to understand disparate conflicting results. Arguably, most attention has been directed to differences caused by choice of optimality criterion, on both simulated and empirical studies (e.g., Felsenstein, 1978; Huelsenbeck and Hillis, 1993; Huelsenbeck, 1995, 1997; Yang, 1996a; Siddall and Kluge, 1997, 1999; Siddall, 1998; Wiens and Servedio, 1998; Siddall and Whiting, 1999; Huelsenbeck et al., 2001; Pol and Siddall, 2001; Swofford et al., 2001; Leaché and Reeder, 2002; Huelsenbeck and Lander, 2003; Kolaczkowski and Thorton, 2004; Yu et al., 2008; Puttick et al., 2017). However, a number of other factors could be equally influential on the inferred phylogenetic hypotheses, such as alignment of sequence data, indel coding, character weighting, model selection, heuristic tree-search strategies, and representation of results such as optimal trees versus a variety of consensus trees (Chippindale and Wiens, 1994; Simon et al., 1994; Yang et al., 1994; Milinkovitch et al., 1996; Yang, 1996b; Morrison and Ellis, 1997; Simmons and Ochoterena, 2000; Simmons et al., 2001; Brandley et al., 2005; Ogden and Rosenberg, 2006; Zwickl, 2006; Brown and Lemmon, 2007; Kumar and Filipski, 2007; McGuire et al., 2007; Simmons et al., 2007; Li et al., 2008; Wong et al., 2008; Ward et al., 2010; Sanderson et al., 2011; Goloboff, 2014; Cabra-García and Hormiga, 2020). Furthermore, in the context of total evidence analyses, combined matrices of molecular and non-molecular data have

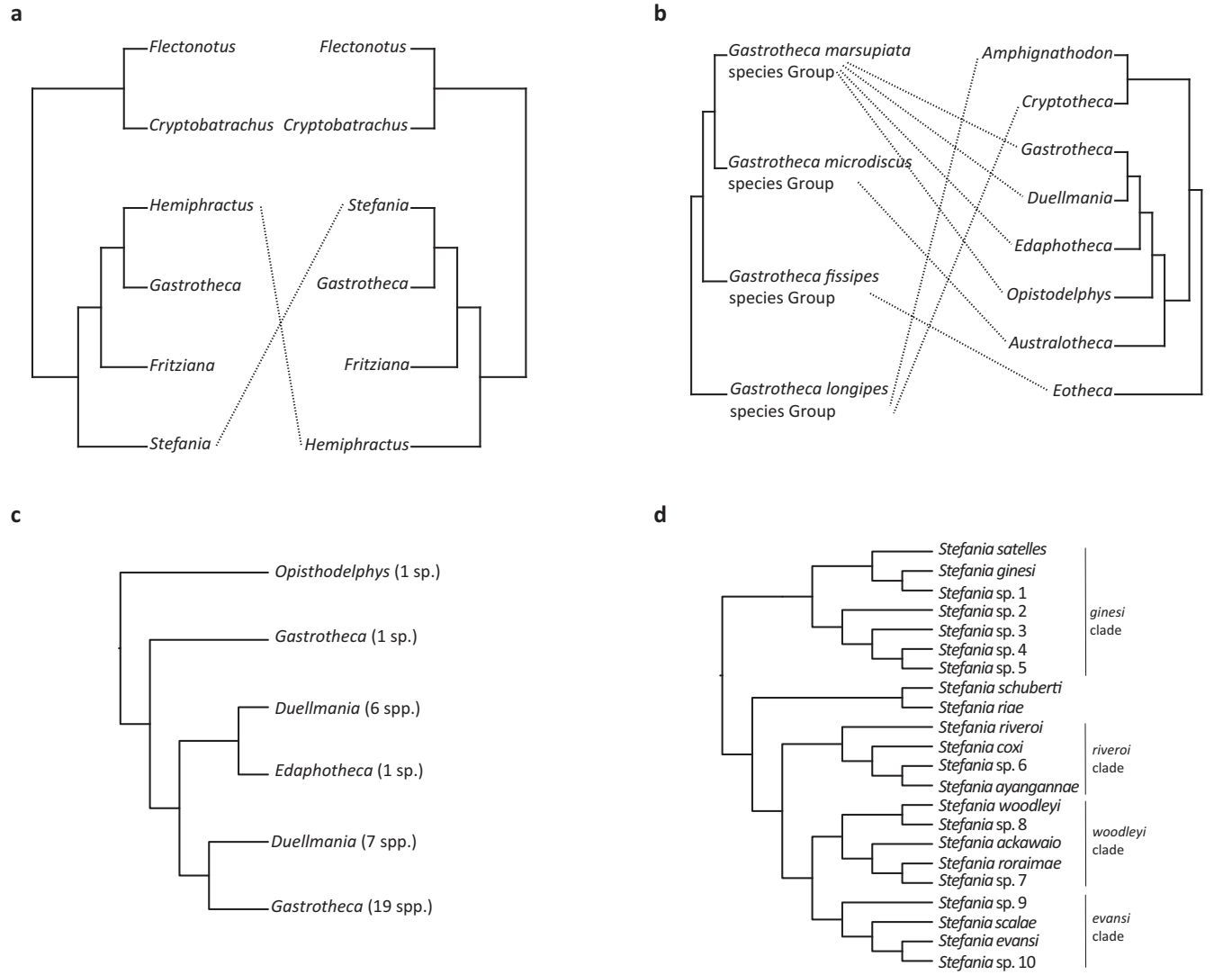

Figure 1. (a) Differences in relationships among genera of Hemiphractidae between Castroviejo-Fisher et al. (2015) (left), and Duellman (2015) (right); (b) relative position of species groups (Castroviejo-Fisher et al., 2015) (left) and subgenera (Duellman, 2015) (right) within *Gastrotheca*; (c) non-monophyly of subgenera *Duellmania* and *Gastrotheca* of Duellman (2015) according to the results of Castroviejo-Fisher et al. (2015); (d) simplified topology showing the phylogenetic relationships among *Stefania* clades proposed by Kok et al. (2017).

attracted discussion, with emphasis on the role of phenotypic characters in large molecular datasets (Giribet, 2015; Lee and Palci, 2015; Chakrabarty et al., 2017; Koch and Gauthier, 2018; Martin et al., 2018). A recent discussion about the suitability of different optimization criteria to analyze morphological data revealed that this type of data fits poorly the Mkv model (Goloboff et al., 2019, Puttick et al., 2017, 2019; Smith, 2019; Wright and Hillis, 2014). This opens an interesting question related to the total evidence analysis of DNA sequences and phenotypic data. Perhaps the very different ways in which parsimony and probabilistic methods, using the Mkv model, deal with phenotypic data are behind some of the incompatible evolutionary histories selected by each approach, even in cases when the number of phenotypic characters is much smaller than nucleotides.

The objective of this study was to identify and discuss the importance of the effect of different alignment methods, optimality criteria, indel codification, and the inclusion of phenotypic data on the phylogenetic relationships of Hemiphractidae, as well as comparing the effects of increasing taxa and characters with regard to the results of the most recent and comprehensive studies (Castroviejo-Fisher et al., 2015; Duellman, 2015; Kok et al., 2017; Walker et al., 2018a). To this end, we analyzed a single dataset of all available and newly generated evidence for Hemiphractidae and relevant outgroup taxa through six strategies that combined all the analytical factors of interest. Furthermore, we used

a subset of this dataset to evaluate the impact of including phenotypic data.

Materials and methods

Taxon sampling

We used the dataset of Castroviejo-Fisher et al. (2015) as an initial scaffold. We selected this study because it constitutes the most complete dataset when both terminals and characters are considered simultaneously. We did not modify the outgroup of Castroviejo-Fisher et al. (2015), which includes 121 terminals representing 20 nobleobatrachian families, and six non-nobleobatrachian terminals (Appendix S1), so that differences are only attributable to changes in hemiphractid taxon and character sampling. The ingroup was modified to include as many species as possible. We list all species and specimens used in the analysis in Appendix S1.

We updated the name or identification of several terminals to reflect taxonomic changes since 2015 and re-identifications of specimens (Table S1). When data from different individuals of the same species resulted in sister relationships in exploratory phylogenetic analysis (i.e., similarity-alignment and equal weights parsimony, results not shown), we calculated the corresponding uncorrected proportional genetic distances (p-distances) between shared 16S ribosomal RNA (rRNA) gene fragments (between 414 and 1157 bp). If genetic distances were <1%, we assigned sequences from different specimens to a single composite terminal, which occurred in 38 ingroup terminals (Appendix S1), in order to reduce the number of missing entries per terminal. The only exception was the terminal *Gastrotheca ovifera*, which combines DNA data from KU 185758 and MHNLS 20979. These two specimens lack overlapping sequences and we based the decision upon morphological similarity of the voucher specimens with the species original description. On the other hand, we included more than one terminal for 15 species (Appendix S1). We excluded *Gastrotheca riobambae* UIMNH 94580 because it corresponds to *G. pseustes*, as pointed out by Castroviejo-Fisher et al. (2015) and confirmed by Carvajal-Endara et al. (2019), a species represented by other specimens.

The total evidence dataset was composed by 143 terminals of egg-brooding frogs, including nominal and undescribed species. The outgroup was represented by 121 nobleobatrachian and six non-nobleobatrachian terminals of Castroviejo-Fisher et al. (2015). *Heleophryne purcelli* was used to root all trees.

Molecular dataset

We included sequences of 20 genes: tRNAPhe, 12S rRNA, tRNA-Val, two non-overlapping fragments of the 16S rRNA, tRNALeu, NADH dehydrogenase subunit 1 (*ND1*), cytochrome oxidase I (COI), cytochrome b (*cytb*), 28S rRNA, proto-oncogene cellular myelocytomatosis exons 2 (*C-MYC 2*) and 3 (*C-MYC 3*), chemokine receptor 4 (*CXCR4*), histone (*H3a*), proopiomelanocortin A (*POMC*), two non-overlapping fragments of the recombination activating gene 1 (*RAG1*), rhodopsin exon 1 (*Rho*), seven in absentia homolog 1 (*SIAH*), solute carrier family 8 member 1 (*SLC8A1*), solute carrier family 8 member 3 (*SLC8A3*), and tyrosinase (*Tyr*). A total of 218 new sequences were generated, representing ten genes (two fragments of *RAG1*) from 55 specimens of egg-brooding frogs (Appendix S2). In addition, 194 sequences, representing 13 genes from 52 hemiphractid terminals were downloaded from GenBank (Appendix S1) prior to April 30, 2018. A new cytochrome b sequence was downloaded for *Acris crepitans*. Genomic DNA was extracted from ethanol-preserved tissues, using the Wizard® Genomic DNA Purification Kit from Promega. Amplification, sequencing, and editing protocols followed those of Guayasamin et al. (2008) and Castroviejo-Fisher et al. (2015).

Phenotypic data

We complemented the phenotypic dataset of Castroviejo-Fisher et al. (2015), which is an updated version of Mendelson et al. (2000), with data for the following terminals: *Fritziana* cf. *fissilis* 1 MNRJ 62845, *F.* cf. *fissilis* 2 CFBH 28886, *F. goeldii* Go III CFBH 10910, *F. ohausi* Oh III CFBH 7611, *F.* sp. CS 1 I CFBH 24810, *F.* sp. CS 1 II MZUFV 11721, *F.* sp. CS 1 III CFBH 30747, *F.* sp CS 2 MCNAM 12341, *F. mitus* CFBH 8273, *F. tonimi* MNRJ 34921, *F. ulei* MNRJ 44622, *Gastrotheca aguaruna* KU 212026, *G. aratia* KU 212056, *G. flamma* (MZUESC 21989 + MZUESC 21990), *G. griswoldi* CORBIDI 16066, *G. oresbios* CORBIDI 11076, *G. rebeccae* CORBIDI 10821, *G.* sp. G CORBIDI 16614, *G. pulchra* (MZUESC 14541 + MZUESC 21991), *G. spectabilis* CORBIDI 11790, and *G.* sp. G CORBIDI 16614. We took pertinent information from the literature (Juncá and Nunes, 2008; Duellman et al., 2014; Folly et al., 2014; Duellman, 2015; Duellman and Venegas, 2016; Walker et al., 2016) or from direct observations of preserved specimens. Also, we coded skull characters for *Hemiphractus fasciatus* ZSM 36/0, *H. elioti* MVUP 1927, and *H. panamensis* (EVACC 061 + CHP 6670) from high-resolution computed tomography reconstructions (Hill et al., 2018), available at MorphoSource. We assigned the phenotypic data from *Cryptobatrachus furhmanni* KU 169378, coded by Mendelson et al. (2000) and modified by Castroviejo-Fisher et al. (2015), to both terminals *Cryptobatrachus fuhrmanni* JDL 14865 and *C. fuhrmanni* (KU 204891 + TNHC-GDC 451). These two terminals have 4.4% p-distances between an 809 bp of 16S fragment and supposedly originate from the same locality (Colombia: Santander: Municipio San Gil: 7 km by road SW San Gil). Because we were not able to study the specimens, we considered it better to assign the same phenotypic data to both terminals.

We assigned the data of *Fritziana ohausi* KU 92226 (Brasil: Rio de Janeiro: Teresópolis) only to the terminal *F. ohausi* Oh I CFBH 16287 (Brasil: Rio de Janeiro: Teresópolis: Parque Nacional da Serra dos Orgãos) because terminals from the other two *F. ohausi* lineages come from localities in São Paulo. We assigned the phenotypic data of *F. goeldii* KU 84721 to the terminals *F. goeldii* Go I (MNRJ 44592 + CFBH 30938) and *F. goeldii* Go II MNRJ 53758, both from Rio de Janeiro, while terminal *F. goeldii* Go III CFBH 10910 comes from São Paulo. We assigned the data of *Gastrotheca stictopleura* to the terminal nearest to the type locality, *G. stictopleura* CORBIDI 14563. We assigned the phenotypic data of *Gastrotheca testudinea* KU 163275 (Peru: Ayacucho: Tutumbaro: río Piene) to the terminal *G. testudinea* CORBIDI 8009 (Peru: Cusco: CC. NN. Aendoshiari) because both localities lie on southern Peru, approximately 500 km apart. We transferred the data from the re-identified terminals *Stefania ginesi* LM 1056 and *S. satelles* VUB 3755 to the terminals *S. ginesi* IRSNB 16736 and *S. satelles* IRSNB 16728.

Modifications on character 46 (state of development of young at hatching) from Mendelson et al. (2000) by Castroviejo-Fisher et al. (2015) included: the fusion of states 1 and 2 of Mendelson et al. (2000), inclusion of specific developmental stages in character states description, and removal of the feeding information (i.e., endotrophic vs. exotrophic) because it constitutes a different character (Sereno, 2007). Thus, character 46 was coded as follows:

- 0—hatch as an early "embryonic" tadpole, < state 30 of Gosner (1960)
- 1—hatch as a well-developed tadpole, ≥ state 30 of Gosner (1960)
- 2—hatch as a froglet, ≥ state 46 of Gosner (1960)

In summary, the final phenotypic dataset includes 243 terminals—of which 116 correspond to hemiphractids—, and 51 characters. However, characters are well represented for 34 terminals—mostly those of Mendelson et al. (2000)—while most other terminals are coded for just four behavioral and developmental characters (i.e., characters 41, 42, 46, and 47 of Mendelson et al., 2000). The terminals *Cryptobatrachus* sp., *H. fasciatus*, and *H. johnsoni* are only represented by phenotypic data.

Theoretical considerations

Regardless of optimality criterion, a phylogenetic hypothesis that is the optimal solution according to a criterion (parsimony or ML in this study) was considered supported if not contradicted by other, equally optimal hypotheses (i.e., evidence is ambiguous, such as when multiple most-parsimonious cladograms are obtained). We interpret frequency of clades based on resampling measures (i.e., jackknife and bootstrap) as a proxy for the relative amount of favorable and contradictory evidence for each group present in the optimal topology when ≥50% (Goloboff et al., 2003; Ramírez, 2005; Kopuchian and Ramírez, 2010). These indices are specific to each dataset and analytical assumptions; they should not be extrapolated between different datasets and analyses, used to judge the validity of a method or hypothesis, or to predict the stability of a clade. In other words, at their best they only convey information on the relative amount of supporting and contradicting evidence for each optimal clade obtained for a specific dataset under specific analytical conditions.

Tree-alignment + parsimony analysis (TAPa)

An assumption of tree-alignment analysis in POY is that length variation among DNA sequences is only due to insertions and/or deletions of nucleotides (Wheeler et al., 2006). Because we used many DNA sequences generated by different studies, which amplified fragments of different length within the same marker, the aforementioned assumption was violated. To solve this problem, prior to the TAPa analysis we performed independent similarity alignments for each marker so that we could divide them into putatively homologous blocks without missing data. We used Muscle (Edgar, 2004) with default parameters, as implemented in Aliview 1.17.1 (Larsson, 2014), to align independently the sequences of tRNAPhe, tRNAVal, tRNALeu, COI, cytb, C-MYC 2, C-MYC 3, CXCR4, H3a, ND1, POMC, the two RAG(1) fragments, Rho, SIAH, SLC8A1, SLC8A3, and Tyr. We aligned sequences of 12S, the two 16S fragments, and 28S using the FFT-NS-i strategy of MAFFT v7 online version (Kuraku et al., 2013; Katoh et al., 2017); all parameters were left at default values. The resulting multiple sequence alignments were visualized and edited in Aliview. We modified the alignments of each DNA marker to correct a clear artifact created by the alignment programs in which, sometimes, the end and/or the beginning of shorter fragments within a gene were placed at the corresponding end or beginning of the whole alignment. Those fragments were placed at the beginning or end of the corresponding sequence. As mentioned above, we divided the similarity alignment of each marker into putatively homologous blocks. For non-coding genes, block partitions were located in conserved regions (no gaps and few or no nucleotide substitutions) according to the multiple sequence alignments. For coding genes, block partitions were established considering groups of conserved codon triplets. Finally, for each putatively homologous block, we removed all gaps inferred by the similarity alignment. We analyzed the DNA sequence blocks (Data S1), along with the phenotypic matrix (Data S2), in POY 5.1.1 (Varón et al., 2010; Wheeler et al., 2015). The phenotypic matrix included 31 binary and 20 multistate characters. Seventeen multistate characters were coded as additive, according to Mendelson et al. (2000). Tree-alignment was performed under parsimony with equal weights for all transformations using direct optimization. For all POY analyses, we used a HP Proliant BL620C G7 server, with one 2.0 GHz processor Intel Xeon E7-2850 and 40 cores, of the cluster Amazonia—housed at the *Laboratório de Alto Desempenho* (LAD)—PUCRS. Tree searches were conducted using the command "search"—which implements an algorithm based on random addition sequence Wagner builds, tree bisection and reconnection (TBR) branch swapping (see Goloboff, 1996, 1999), parsimony ratcheting (Nixon, 1999), and tree fusing (Goloboff, 1999)—running consecutive rounds of searches within a specified run-time (24, 168, 72, and 96 h respectively), storing the shortest trees of each independent run and performing a final round of tree fusing on the pooled trees of each search (200–5000 iterations). The last searches did not improve tree cost, so the resulting optimal tree was submitted to a round of swap using iterative pass optimization (Wheeler, 2003b) in POY 4.1.2 (Varón et al., 2010). Finally, the optimal implied alignment from iterative pass optimization was converted to a data matrix (Wheeler, 2003a) and submitted to driven searches in TNT 1.5 (Goloboff et al., 2008; Goloboff and Catalano, 2016). The implied alignment of all DNA sequences is included in Data S3. We implemented all the New Technologies algorithms (Sectorial Search, Ratchet, Drift, Tree Fusing) in their default mode, with "Search" set for all taxa at level 70, the minimum length tree to be found set to 100 times, and random seed = 1.

We calculated jackknife frequencies (JK) in TNT using the implied alignment for 1000 pseudoreplicate searches under a Traditional Search analysis with 50 replicates and 50 trees saved per replication, gaps treated as fifth state, and removal probability 0.36 ($\sim e^{-1}$) to render bootstrap and JK values comparable (Farris et al., 1996). We acknowledge that jackknife frequencies are expected to be higher when resampling from an implied alignment than from a similarity alignment (Butler et al., 2010; Wheeler, 2012; Padial et al., 2014). We calculated Goodman-Bremer (GB) values (Goodman et al., 1982; Bremer, 1988; see Grant and Kluge, 2008) for each supported clade in TNT using the optimal tree-alignment matrix and the parameters specified in the bremer.run macro (available at http://www.zmuc.dk/public/phylogeny/tnt/), which begins by searching for trees N steps longer than the optimum (10 random addition sequence Wagner builds and TBR swapping saving two trees per replicate) and then using inverse constraints for each node of the most parsimonious tree. Swapping of each constrained search was limited to 20 min and constrained searches were repeated 100 times rather than the three times specified as default in bremer.run macro.

Similarity-alignment + parsimony analysis, gaps as a fifth state (SAP5th)

For all similarity-alignment analyses, we concatenated all blocks of each marker generated for the TAPa analysis, but including gaps implied by the multiple sequence alignment, into a single matrix using SequenceMatrix 1.7.8 (Vaidya et al., 2011). We used POY 5.1.1 to concatenate the DNA sequence matrix with the phenotypic matrix, and exported the resulting Nexus file in.tnt format from Mesquite 3.31 (Maddison and Maddison, 2017). We ran tree searches in TNT 1.5, all transformations were equally weighted, gaps treated as a fifth state, and phenotypic multistate characters were coded as additive or non-additive following Mendelson et al. (2000). Searches for the shortest trees, JK frequencies, and GB values were conducted in TNT as explained in the previous section. The analyzed data matrix is available as Data S4.

Similarity-alignment + parsimony analysis, gaps as binary characters (SAPg)

We used the molecular matrix analyzed in SAP5th to code indels as the longest possible fragments. We incorporated indels in the

dataset as a block of binary characters according to the Simple Indel Coding method (SIC) of Simmons and Ochoterena (2000) as implemented in SeqState 1.4.1 (Müller, 2005, 2006). We concatenated all datasets (i.e., DNA sequences, indels coded as binary characters, and phenotypic characters) using POY 5.1.1, and exported the resulting Nexus file as a.tnt file using Mesquite 3.31. We conducted tree searches, calculation of JK frequencies, and GB values in TNT 1.5 as explained above. We recoded gap characters ("–") of the molecular matrix as missing data ("?"). The analyzed data matrix is available as Data S5.

Similarity-alignment + parsimony analysis, gaps as missing data (SAPm)

For this analysis, we analyzed the same total evidence matrix as in the SAP5th but we set all searches to consider gaps as missing data. We executed tree searches, calculation of JK frequencies and GB values in TNT 1.5 as for the implied alignment of TAPa. The analyzed data matrix is available as Data S6.

Similarity-alignment + maximum likelihood analysis, gaps as binary characters (SALg)

We used PartitionFinder v 1.1.1 (Lanfear et al., 2012) to select the most optimal partition scheme and substitution models for the molecular dataset, under the corrected Akaike information criterion (AICc). Branch lengths were estimated as linked. A tree topology generated in TNT (Traditional Search of 1000 replicates, swapping algorithm TBR, and saving 10 trees per replication) was supplied to the PartitionFinder analysis. Three partition schemes were evaluated: (i) all data combined; (ii) mtDNA and nuDNA; (iii) a 20-partition scheme (each locus independently). To incorporate indel information, we used the SIC method as in SAPg analysis. We included the aligned DNA sequences, with gap characters ("–") recoded as missing data ("?"), indels as a block of binary characters, and binary and multistate additive and non-additive phenotypic characters in a single Nexus file. A Standard Variable model (Mkv) was assigned to the indel character block and the binary and non-additive phenotypic character block. A Standard Variable (Mkv) Ordered model was assigned to the additive phenotypic characters block. There are limitations of GARLI regarding the treatment of multistate characters under this model; however, our objective is to compare the results of parsimony to those of ML using the most commonly applied models for discrete morphological characters.

We conducted tree searches using GARLI 2.01 (Zwickl, 2006). A total of 500 independent searches were conducted using as a starting point a random topology, 100 000 generations without topology improvement required for termination (genthreshfortopoterm), treerejectionthreshold at 50, and the limsprrange parameter was set at 15. Finally, we selected the best tree among the 500 independent searches by comparing the likelihood scores.

We calculated 500 independent bootstrap pseudoreplicates, run under less rigorous parameters than tree searches (genthreshfortopoterm: 10 000; limsprrange: 6) to reduce execution time. We compiled the 500 pseudoreplicates using the R package Ape 4.1 (Paradis et al., 2004) and the bootstrap frequencies were assigned to the corresponding clades of the optimal tree using SumTrees 4.3.0 (Sukumaran and Holder, 2010a) of the DendroPy 4.3.0 package (Sukumaran and Holder, 2010b). The analyzed data matrix is available as Data S7.

We ran tree searches using the Amazonia cluster housed at the Laboratório de Alto Desempenho (LAD)—PUCRS, and the bootstrap pseudoreplicates using the Amazonia cluster and the CIPRES Science Gateway (Miller et al., 2010).

Similarity-alignment + maximum likelihood analysis, gaps as missing data (SALm)

We performed partitions and model selection, tree searches, selection of optimal trees, and calculation of bootstrap frequencies as explained for the SALg analysis. Gap characters ("–"), of the DNA sequences alignment remained unaltered, as GARLI automatically treats gaps as nucleotides of unknown identity. The analyzed data matrix is available as Data S8.

Comparisons and the effect of analytical factors

The six analyses were designed to single out, heuristics aside, the effect of each analytical factor of interest through alternative comparison of the results (Fig. 2). The comparison of the results of TAPa and SAP5th analyses allowed identifying the direct effects of using different alignment methods. The effect of using different optimality criteria can be more easily identified by comparing differences between the results of SAPm vs. SALm and SAPg vs. SALg. Finally, the effect of different treatments of indels becomes clear-cut when comparing the results among SAP5th, SAPg, and SAPm on the one hand and among SALm and SALg on the other. All other comparisons were cautiously done by connecting analyses (Fig. 2) and keeping in mind the potential additive effects that more than one varying analytical parameter could imply.

We quantified differences among topologies using the program ybara_sa.py of YBYRÁ (Machado, 2015), which implements an algorithm to calculate local and global distances between trees based on the Robinson-Foulds metric (Robinson and Foulds, 1981). Robinson-Foulds local distances are calculated by $1 - (S/U)$, where S is the number of shared clades or splits and U is the number of unique clades or splits in the compared trees. Local distances were calculated considering the number of shared clades among trees. Using TNT, we also calculated pairwise SPR distances (Goloboff, 2008) using 100 replicates and calculations stratified in 20 levels. Both RF and SPR distances were calculated among optimal trees of all six analyses. The prevalence of clades per tree file and sensitivity plots were also generated through a sensitivity analysis, using ybara_sa.py of YBYRÁ (Machado, 2015). Additionally, we calculated the reciprocal cost of optimal topologies of each analytical strategy under the conditions of all the other implemented strategies. For all parsimony analyses, we calculated the reciprocal cost of both one optimal tree. Costs in POY were calculated under IPO.

Wildcard taxa

We conducted analyses to search for potential wildcard taxa for each set of resulting trees of parsimony analyses using YBYRÁ. First, consensus and optimal tree files were pruned one terminal at a time using ybyra_sa.py. Then, pruned tree files were submitted to MSdist to calculate how each terminal affects the average matching split distances (MSD) among trees. Terminals resulting in the lowest MSD were identified as more likely to cause a decrease of resolution and considered potential wildcards.

Impact of phenotypic data

To determine the impact of phenotypic data, we assembled a reduced dataset. We included only those terminals represented by ≥41 phenotypic characters but excluded all terminals represented only by phenotypic characters. This yielded a dataset of five hylids and 31 hemiphractids, with representatives of all six genera. Besides the phenotypic characters, the terminals were represented by up to 20 mitochondrial and nuclear genes. A total of 12 matrices were

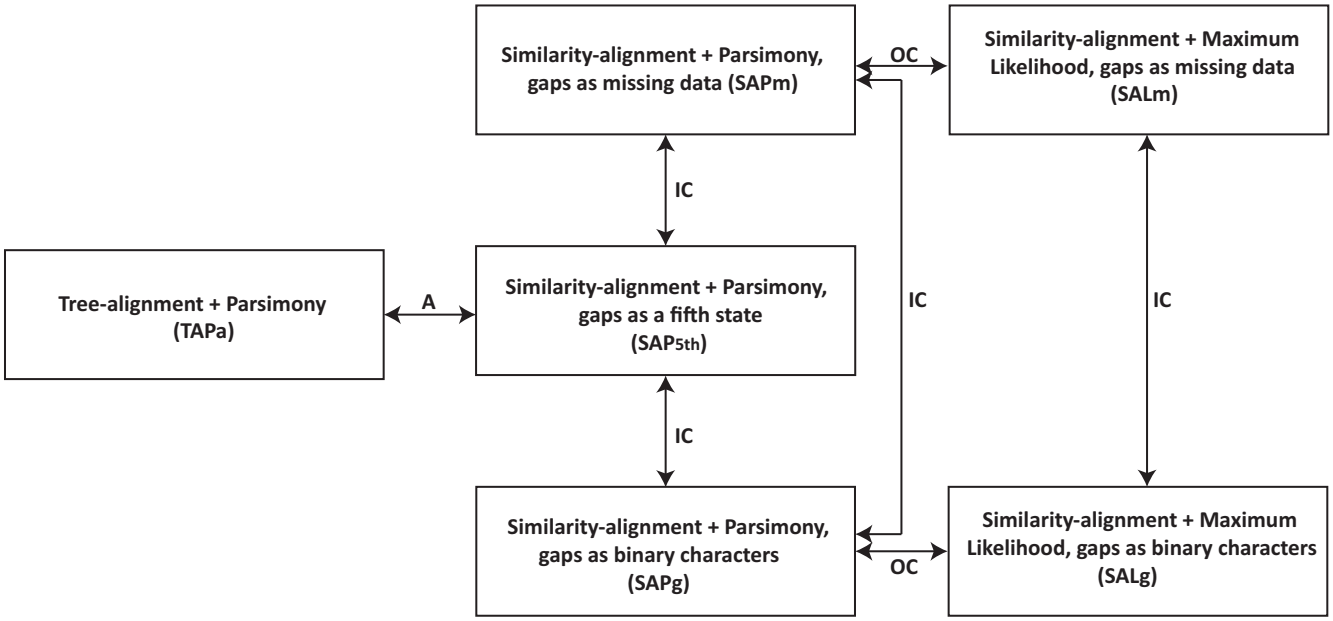

Figure 2. The six phylogenetic analyses performed in this study to evaluate the effect of optimality criterion (OC), alignment (A), and indel coding (IC). Arrows show direct pathways between analyses where just the single analytical factor indicated differs (heuristics aside).

assembled for the combined (phenotypic + molecular data) and molecular-only analyses. We performed all analyses as for the complete matrices, only for TAPa tree searches were reduced to two consecutive search rounds of 48 and 24 h, remaining commands were not modified and TNT analyses were carried out as for the complete matrix. We quantified the number of shared clades between pairs of topologies using the program ybara_sa.py of YBYRÁ. For these comparisons, the five outgroup terminals were removed from all trees.

Results

Detailed information for each of the analyzed matrices is in Table 1. Costs of optimal tree(s) of all six analyses are included in Table 2. The models selected for each locus of both SALg and SALm analyses are in Table S2.

At the outgroup level, half of the analyses recovered a monophyletic Athesphatanura and a non-monophyletic Brachycephaloidea (Fig. 3, Figs. S1–S6). In the strict consensus of SAP5th, the non-monophyly of Athesphatanura is determined by the position of *Pseudopaludicola falcipes* as sister of all other nobleobatrachians, and the non-sister relationships of four main athesphatanuran clades (Fig. S2). In the SALg optimal tree, Dendrobatoidea (Aromobatidae + Dendrobatidae) is sister of a clade including all other athesphatanurans and Hemiphractidae (Fig. S5). In the SALm optimal tree, Dendrobatoidea is sister of all other nobleobatrachians (Fig. S6). The position of *Ceuthomantis smaragdinus* as sister of all nobleobatrachians causes the non-monophyly of Brachycephaloidea in TAPa, SAPg, and SAPm analyses. Two analyses, SALm and SAP5th, recover Hemiphractidae as sister to Brachycephaloidea, the other four analyses recover Athesphatanura, or all athesphatanurans excluding Dendrobatoidea, as sister to Hemiphractidae (Fig. 3; Figs. S1–S6). Of these three sister relationship alternatives, only Athesphatanura + Hemiphractidae inferred from TAPa have resampling support values above 60 (JK = 99, Table 4). GB values vary between 15 (SAP5th) and 58 (TAPa).

At the ingroup level, we do not consider detailed comparisons among the relative position of each terminal due to their large number. We focus on the differences with respect to the monophyly and relationships among currently recognized supraspecific taxa of hemiphractids, such as subfamilies, genera, subgenera, species groups, and other supraspecific clades

Table 1
Summary of the six analyzed matrices detailing the total number of aligned characters, and missing data and gap cells

Matrix	Aligned characters	Missing data cells	Gap cells
TAPa*	17 852	2 705 026	715 223
SAP5th	14 315	2 378 915	141 124
SAPg	15 736	2 718 617	1421
SAPm	14 315	2 520 039	–
SALg	15 736	2 718 617	1421
SALm	14 315	2 520 039	–

Total number of terminals = 270.
*Implied alignment.

Table 2
Costs of all optimal trees (in bold) and reciprocal cost of single optimal tree inferred by the six different analyses, this later value is followed by percentage of decrease in optimal cost

Tree from/analytical conditions from	TAPa	SAP_{5th}	SAPm	SAPg	SALg	SALm
TAPa	**84 885**	90 768/0.09%	83 851/0.06%	87 724/0.06%	−369 449.19/0.24%	−349 434.25/0.3%
SAP_{5th}	85 281/0.47%	**90 684**	83 897/0.12%	87 769/0.11%	−369 443.03/0.24%	−349 231.27/0.24%
SAPm	85 258/0.44%	90 756/0.08%	**83 798**	87 673/0.03%	−369 729.38/0.32%	−349 436.53/0.3%
SAPg	85 262/0.44%	90 750/0.07%	83 801/0.004%	**87 670**	−369 656.08/0.3%	−349 581.02/0.34%
SALg	85 481/0.7%	90 979/0.33%	84 132/0.4%	87 988/0.36%	**−368 553.75**	−348 416.14/0.008%
SALm	85 457/0.67%	90 995/0.34%	84 099/0.36%	87 979/0.35%	−368 585.44/0.009%	**−348 389.07**

Each column defines analytical conditions. For parsimony analyses, the cost of optimal trees was calculated using the first of all optimal trees of each analysis.

(Castroviejo-Fisher et al., 2015; Duellman, 2015; Kok et al., 2017; Walker et al., 2018a). We present a summary of the main phylogenetic relationships among supraspecific groups and their support values in Fig. 4 and Table 4. We also include visual comparisons, within each genus of Hemiphractidae, of the relative position of each terminal among the results of all analyses in Figs. S7–S21. All analyses recovered the monophyly of Hemiphractidae, although with different support values (Table 4). The monophyly of *Flectonotus*, *Fritziana*, *Gastrotheca*, *Hemiphractus*, and *Stefania* was recovered by all analyses, with JK or BS ≥ 90 except for *Gastrotheca* in SALg (BS = 57; Table 4). However, *Cryptobatrachus* is not monophyletic with respect to *Flectonotus*, as they are part of a polytomy according to TAPa, SAPg, and SAPm analyses (Fig. 3, Figs. S1, S3, S4). Hemiphractinae, as previously defined, is recovered only by SALg and SALm results (Figs. S5 and S6, respectively). In all ML analyses, relationships within Hemiphractinae are pectinate as follows (*Hemiphractus*(*Stefania*(*Fritziana*, *Gastrotheca*))). All parsimony analyses place *Fritziana* and *Stefania* as more closely related to *Cryptobatrachus* and *Flectonotus*, while *Gastrotheca* is sister of *Hemiphractus*, rendering a non-monophyletic Hemiphractinae. All analyses, although with differences regarding the relationships among terminals, recovered the monophyly of all the supraspecific clades proposed by Kok et al. (2017) for *Stefania*. No analyses recovered as monophyletic all the supraspecific taxa of *Gastrotheca* proposed by either Castroviejo-Fisher et al. (2015) or Duellman (2015).

Comparison of trees

The costs of optimal trees under the conditions of all alternative analytical strategies are not better than the best cost of each analytical strategy (Table 2). Among parsimony analyses, the best costs correspond to the SAPm and TAPa analyses, the SAPm analysis obtained shorter trees than TAPa, a result explained by the exclusion of indels from the SAPm analysis. Similarly, the SALm analysis resulted in a better likelihood score than SALg.

Under TAPa conditions, the optimal tree with the closest tree length to TAPa is SAPm (373 steps longer), followed by SAPg (377 steps longer), SAP5th (454 steps longer), SALm (572 steps longer), and SALg (596 steps longer). The optimal tree with the closest cost to SALg is SALm, followed by the parsimony optimal trees of SAP5th, TAPa, SAPg, and SAPm. The optimal tree with the closest cost to SALm (excluding SALg) is SAP5th followed by TAPa, SAPm, and SAPg optimal trees.

The total number of clades for each final topology per analysis and the number of shared clades among

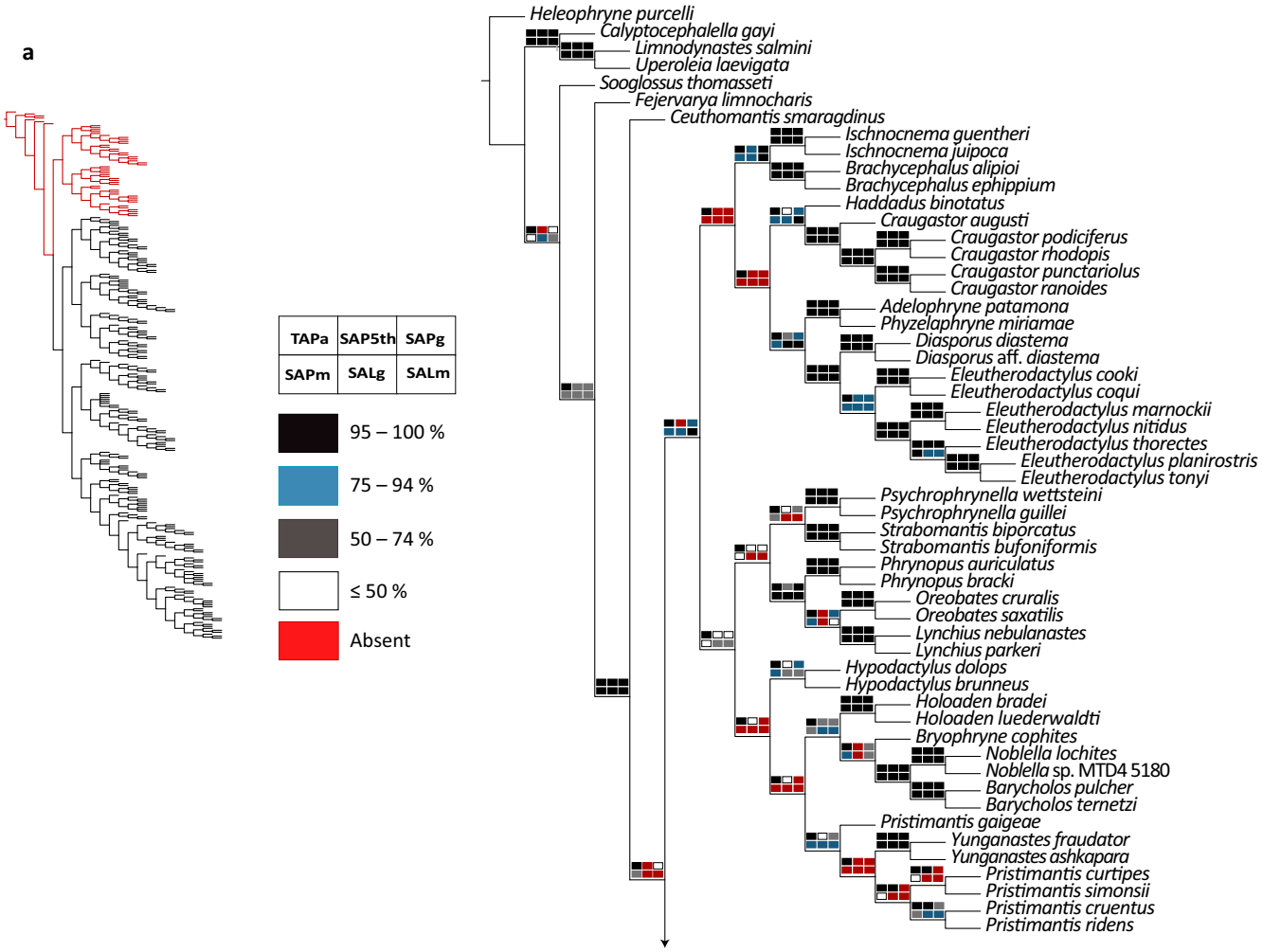

Figure 3. Phylogenetic relationships among egg-brooding frogs and outgroups inferred from a tree-alignment + parsimony analysis (TAPa) of phenotypic characters and DNA sequences. This topology reflects the strict consensus of the 393 most parsimonious trees (tree costs = 84 885 steps). Colored boxes over branches indicate monophyly (all colors but red) and frequency of clades in jackknife or bootstrap resampling measures in parsimony and maximum likelihood analyses, respectively, as inferred from each of the six analyses performed in this study: tree-alignment + parsimony (TAPa); similarity-alignment + parsimony, gaps as binary characters (SAPg); similarity-alignment + parsimony, gaps as missing data (SAPm); similarity-alignment + parsimony, gaps as a fifth state (SAP5th); similarity-alignment + maximum likelihood analysis, gaps as missing data (SALm); and similarity-alignment + maximum likelihood, gaps as binary characters (SALg). A skeletal topology on the left shows in red the illustrated section. [Colour figure can be viewed at wileyonlinelibrary.com]

them are presented in Table 3. The optimal trees of the SALm and SALg analyses have the highest number of clades (269), followed by the strict consensuses of TAPa (256), SAP5th (255), SAPm (247), and SAPg (225). The topologies with the greatest number of shared clades are SALm and SALg (256), despite the different treatment of indels. The topologies with the least number of shared clades, the consensus tree of SAPg and SALg optimal tree (188), differ in just optimality criteria. Among the parsimony analyses, the topologies with the greatest number of shared clades are the consensus trees of SAPm and SAPg (222), both just differing in the treatment of indels. The topologies with the least number of shared clades, from analyses with different treatment of indels, are the consensus trees of SAP5th and SAPg (197). The parsimony tree sharing more clades with SALg is the consensus tree of SAP5th (199), and the one sharing the least number of clades is the consensus tree of SAPg (188). The parsimony consensus trees sharing more clades with SALm are SAP5th and SAPm (198), and the one sharing the least number of clades is the consensus tree of SAPg (191).

Considering the total number of optimal trees obtained for each parsimony analyses, both Robinson-Foulds (RF) local distance values and pairwise SPR distances were calculated for three optimal trees (0, 150, 300) of TAPa and SAP5th and for six optimal

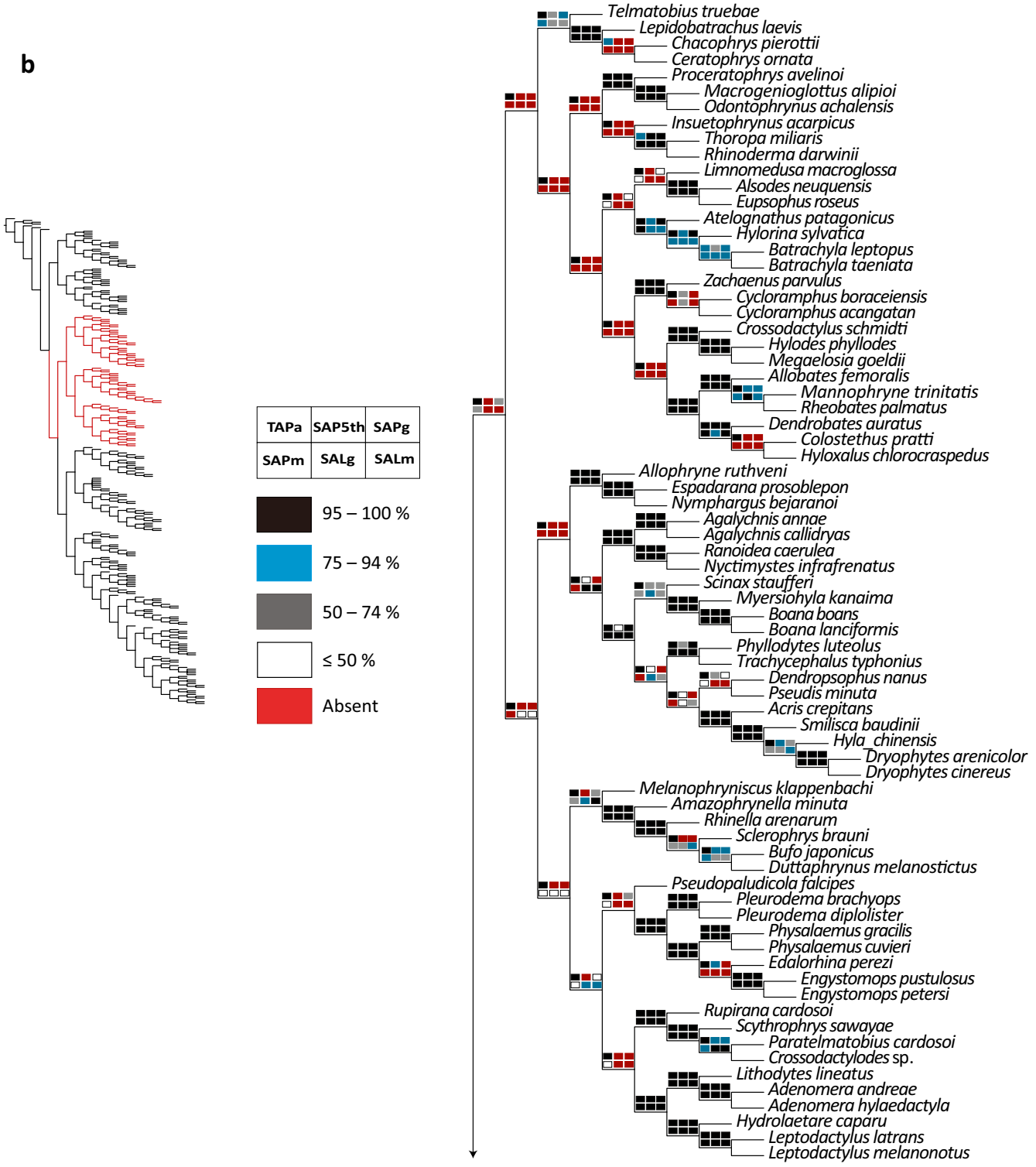

Figure 3. Continued

trees (0, 50, 150, 300, 1000, 1100) of SAPg and SAPm. The mean Robinson-Foulds (RF) local distance values (Table S3) and mean pairwise SPR distances (Table S4) are higher for comparisons between SAL topologies and trees from all parsimony analyses. Also, optimal trees from TAPa are more distant to SAP5th optimal trees than to trees from SAPg and SAPm analyses. Both measures of topological distance

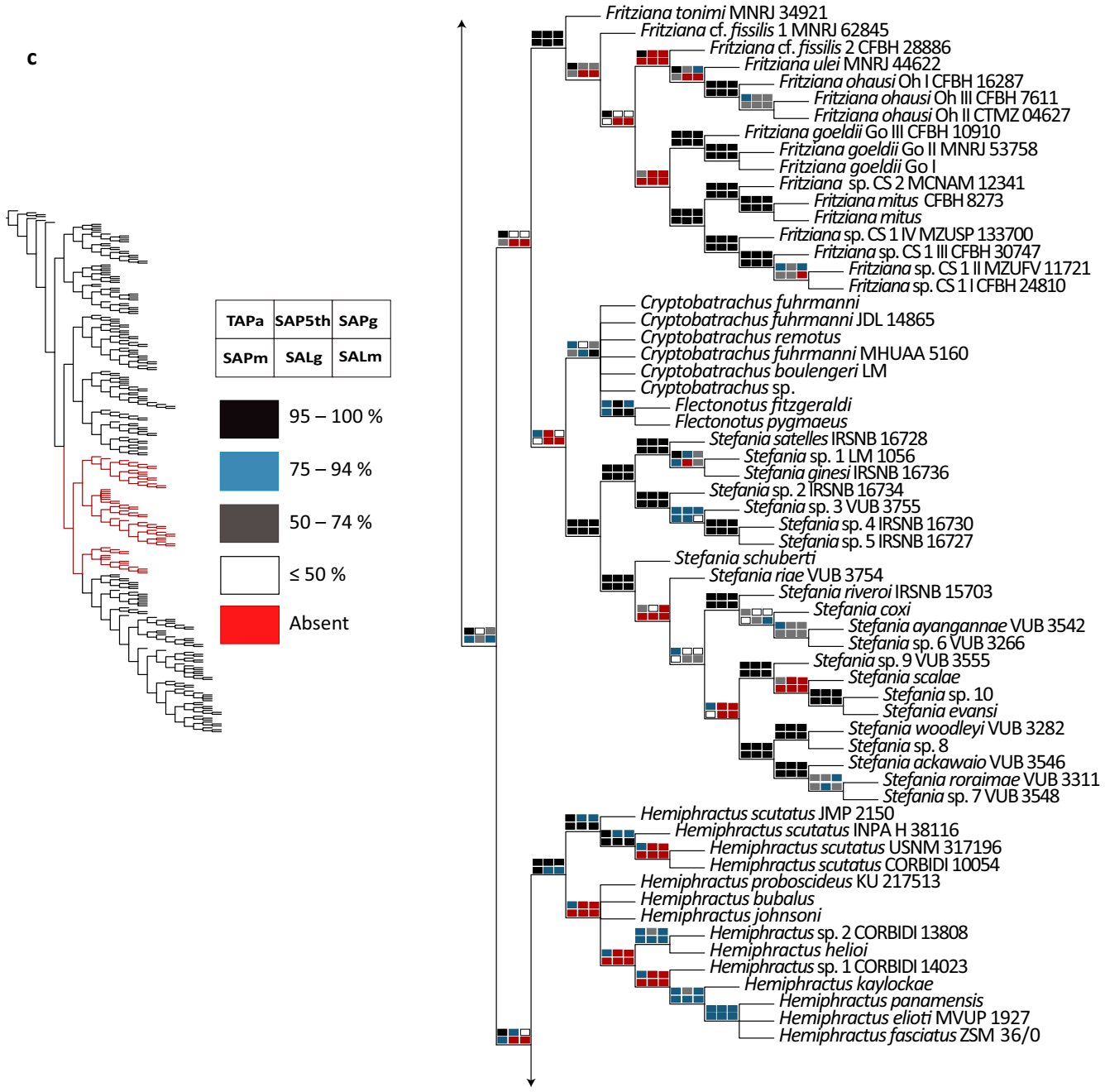

Figure 3. Continued

are congruent with the number of clades shared between trees of the six analyses, the higher the number of shared clades, the smaller the topological distance values (Table 3, Tables S3 and S4).

The effect of alignment method: tree-alignment vs. similarity-alignment

The strict consensuses of TAPa and SAP5th analyses, of 393 and 309 trees respectively, are almost equally resolved (256 vs. 255 clades; Table 3), and share 80% of its clades.

In SAP5th and TAPa results, Hemiphractidae have different sister groups (Brachycephaloidea and Athesphatanura, respectively). Both analyses' results have unresolved clades only at the ingroup level (Fig. 3, Figs. S1 and S2), within *Cryptobatrachus*, *Gastrotheca* and *Hemiphractus*, and within *Stefania* only in SAP5th.

Both TAPa and SAP5th analyses recover the clade *Gastrotheca* + *Hemiphractus* as sister of a clade

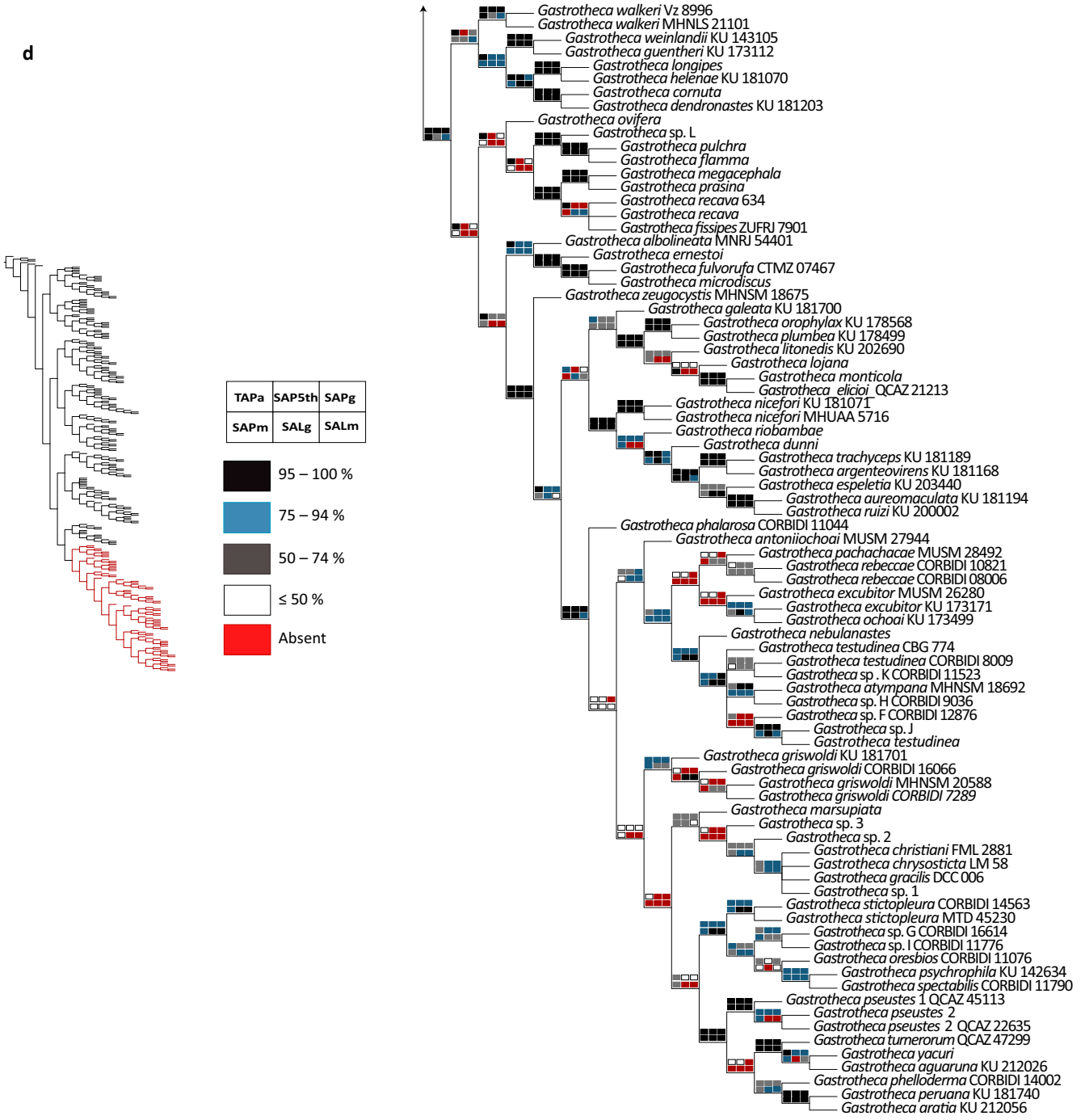

Figure 3. Continued

including the other four genera. However, only SAP5th recovers a sister relationship between *Fritziana* and *Stefania*, while in TAPa *Stefania* is sister of the clade that contains *Cryptobatrachus* and *Flectonotus*. Although both topologies have a polytomy among *Cryptobatrachus* terminals, only in the strict consensus of TAPa the position of *Flectonotus* is unresolved as part of a polytomy with *Cryptobatrachus*. Other noteworthy differences include the relationship between the clades of *Stefania evansi* and *S. woodleyi* (sister clades in TAPa), the relative relationships and level of resolution among the terminals of *Hemiphractus*, the position of *G. ovifera*, and the non-monophyly of the *G. longipes* (*Amphignathodon*) species group. Visual detailed comparisons between TAPa and SAP5th results are presented in Fig. S7.

The effect of optimality criteria: parsimony vs. maximum likelihood

This comparison was made by contrasting the results of SAPg vs. SALg and of SAPm vs. SALm because, heuristics aside, these two sets of analyses only differ in optimality criterion.

The SAPg strict consensus of 1402 optimal trees is less-resolved than SALg optimal tree (225 vs. 269 clades; Table 2). Similarly, the SAPm strict consensus of 1132 trees resulted in 247 clades, and the SALm optimal solution in 269 clades. The results of SALg (Fig. S5) and SAPg (Fig. S3) analyses share 188 clades, 84% of the 225 clades present in the SAPg strict consensus. The strict consensus of the SAPm analysis shares 80% of its clades with the SALm optimal tree.

At the ingroup level, the most striking differences of SAL analyses results, compared to SAPg and SAPm results, are the presence of Cryptobatrachinae and Hemiphractinae, in the former, and the sister relationship of *Gastrotheca* and *Fritziana* (*Gastrotheca* + *Hemiphractus* in SAPg and SAPm results). The position of *G. ovifera* and the polyphyly of the *G. fissipes* species group (*Eotheca*) are the main differences within *Gastrotheca*. Within *Fritziana* and *Hemiphractus* the relative position of several terminals is different and in *Stefania*, the SAPm analysis recovered a sister relationship between the *Stefania evansi* and *S. woodleyi* clades, whereas the SALm analysis found the *S. riveroi* clade as sister of the *S. woodleyi* clade.

Detailed visual comparisons between SAPg and SALg, and of SAPm and SALm are presented in Figs. S17 and S20, respectively.

The effect of indel coding: gaps as unknown nucleotides, as fifth state or as binary characters

Within parsimony, the three indel coding strategies were easily investigated by comparing the results of SAP5th (Fig. S2), SAPg (Fig. S3), and SAPm (Fig. S4). Within a likelihood context, two coding options (i.e., indels as unknown nucleotides and as binary characters) were compared through the results of SALg (Fig. S5) and SALm (Fig. S6).

Among SAP strict consensus trees, the SAP5th analysis has the following unique results: (i) the monophyly of Brachycephaloidea; (ii) the non-monophyly of Athesphatanura; (iii) the sister relationship between Brachycephaloidea and Hemiphractidae; and (iv) the sister relationship of *Fritziana* and *Stefania*. Also, *Cryptobatrachus* is monophyletic with respect to *Flectonotus* in the SAP5th strict consensus tree, opposite to other parsimony results.

The results of SAPg and SAPm agree on the non-monophyly of Brachycephaloidea, the monophyly of Athesphatanura and its sister relationship to Hemiphractidae, and the sister relationship of *Stefania* with the clade of *Cryptobatrachus* + *Flectonotus*. All SAP analyses agree on recovering *Hemiphractus* as sister to *Gastrotheca*.

The strict consensus of the SAP5th analysis is more resolved than the SAPm strict consensus (255 vs. 247 clades; Table 3). These are the third pair of parsimony analyses sharing the highest number of clades (203). The strict consensus of SAPm shared 82% of its clades with the SAP5th strict consensus. Differences between the results of SAP5th and SAPm involved the relationships of few terminals and clades within *Stefania*, and most important differences involved *Gastrotheca* and *Hemiphractus* relationships (Figs. S2, S4, and S13).

The SAPm strict consensus was more resolved than the SAPg strict consensus (247 vs. 225 clades; Table 3). These analyses shared the highest number of shared clades (222) among the results of parsimony analyses; the SAPg strict consensus tree shared all but three clades with the SAPm strict consensus tree. Both SAPg and SAPm recovered the same relationships among hemiphractid genera. The main differences between SAPg and SAPm involved relationships within *Gastrotheca* and *Stefania* (Figs. S3, S4 and S16).

The strict consensus of the SAP5th analysis was better resolved than the SAPg strict consensus tree (255 vs. 225 clades, Table 3). These analyses shared the least number of clades (197) among the results of parsimony analyses. The SAPg strict consensus tree shared 88% of its clades with the SAP5th strict consensus. The results of SAPg and SAP5th had few differences within *Stefania*, with most important differences within *Gastrotheca* and *Hemiphractus* (Figs. S2, S3 and S12).

Both SALg and SALm optimal trees had 269 fully resolved clades. These trees shared 256 clades (95%), the highest number of shared clades among all analyses. Within the outgroup, both analyses recovered the non-monophyly of Athesphatanura and the monophyly of Brachycephaloidea. However, different clades were recovered as sister to Hemiphractidae, the largest clade of Athesphatanura and Brachycephaloidea for SALg and SALm, respectively. The SALg and SALm analyses recovered the same relationships among hemiphractid genera. The most important differences between the SALg and SALm results involved relationships within *Hemiphractus*, followed by differences within the *Gastrotheca marsupiata* species group (restricted to the *Gastrotheca* subgenus clade), and few differences within *Fritziana* and *Stefania* (Figs. S5, S6 and S21).

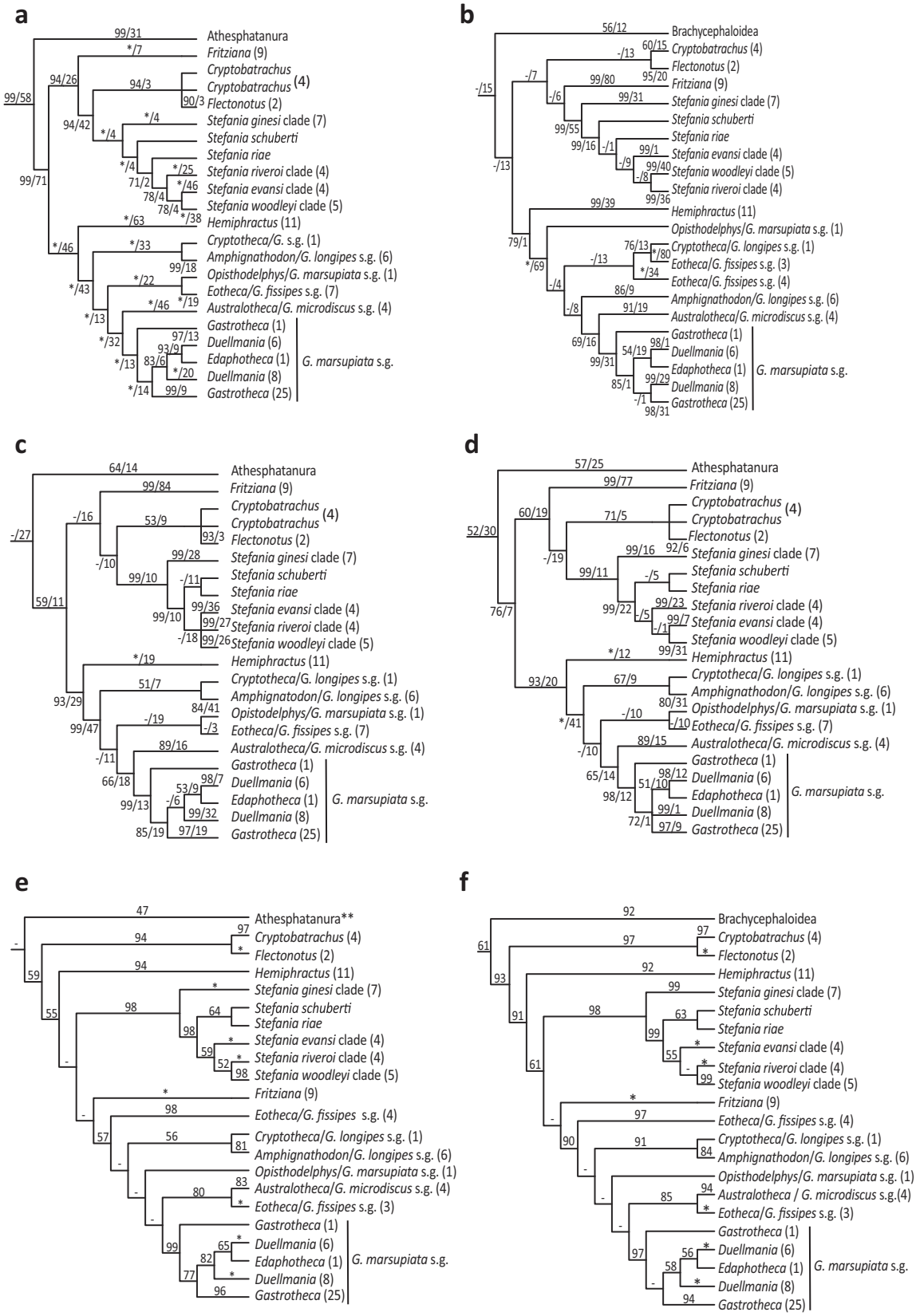
a
Athesphatanura
Fritziana (9)
Cryptobatrachus (4)
Flectonotus (2)
Stefania ginesi clade (7)
Stefania schuberti
Stefania riae
Stefania riveroi clade (4)
Stefania evansi clade (4)
Stefania woodleyi clade (5)
Hemiphractus (11)
Cryptotheca/G. s.g. (1)
Amphignathodon/G. longipes s.g. (6)
Opisthodelphys/G. marsupiata s.g. (1)
Eotheca/G. fissipes s.g. (7)
Australotheca/G. microdiscus s.g. (4)
Gastrotheca (1)
Duellmania (6)
Edaphotheca (1)
Duellmania (8)
Gastrotheca (25)
G. marsupiata s.g.
b
Brachycephaloidea
Cryptobatrachus (4)
Flectonotus (2)
Fritziana (9)
Stefania ginesi clade (7)
Stefania schuberti
Stefania riae
Stefania evansi clade (4)
Stefania woodleyi clade (5)
Stefania riveroi clade (4)
Hemiphractus (11)
Opisthodelphys/G. marsupiata s.g. (1)
Cryptotheca/G. longipes s.g. (1)
Eotheca/G. fissipes s.g. (3)
Eotheca/G. fissipes s.g. (4)
Amphignathodon/G. longipes s.g. (6)
Australotheca/G. microdiscus s.g. (4)
Gastrotheca (1)
Duellmania (6)
Edaphotheca (1)
Duellmania (8)
Gastrotheca (25)
G. marsupiata s.g.
c
Athesphatanura
Fritziana (9)
Cryptobatrachus (4)
Flectonotus (2)
Stefania ginesi clade (7)
Stefania schuberti
Stefania riae
Stefania evansi clade (4)
Stefania riveroi clade (4)
Stefania woodleyi clade (5)
Hemiphractus (11)
Cryptotheca/G. longipes s.g. (1)
Amphignatodon/G. longipes s.g. (6)
Opistodelphys/G. marsupiata s.g. (1)
Eotheca/G. fissipes s.g. (7)
Australotheca/G. microdiscus s.g. (4)
Gastrotheca (1)
Duellmania (6)
Edaphotheca (1)
Duellmania (8)
Gastrotheca (25)
G. marsupiata s.g.
d
Athesphatanura
Fritziana (9)
Cryptobatrachus (4)
Flectonotus (2)
Stefania ginesi clade (7)
Stefania schuberti
Stefania riae
Stefania riveroi clade (4)
Stefania evansi clade (4)
Stefania woodleyi clade (5)
Hemiphractus (11)
Cryptotheca/G. longipes s.g. (1)
Amphignathodon/G. longipes s.g. (6)
Opisthodelphys/G. marsupiata s.g. (1)
Eotheca/G. fissipes s.g. (7)
Australotheca/G. microdiscus s.g. (4)
Gastrotheca (1)
Duellmania (6)
Edaphotheca (1)
Duellmania (8)
Gastrotheca (25)
G. marsupiata s.g.
e
Athesphatanura**
Cryptobatrachus (4)
Flectonotus (2)
Hemiphractus (11)
Stefania ginesi clade (7)
Stefania schuberti
Stefania riae
Stefania evansi clade (4)
Stefania riveroi clade (4)
Stefania woodleyi clade (5)
Fritziana (9)
Eotheca/G. fissipes s.g. (4)
Cryptotheca/G. longipes s.g. (1)
Amphignathodon/G. longipes s.g. (6)
Opisthodelphys/G. marsupiata s.g. (1)
Australotheca/G. microdiscus s.g. (4)
Eotheca/G. fissipes s.g. (3)
Gastrotheca (1)
Duellmania (6)
Edaphotheca (1)
Duellmania (8)
Gastrotheca (25)
G. marsupiata s.g.
f
Brachycephaloidea
Cryptobatrachus (4)
Flectonotus (2)
Hemiphractus (11)
Stefania ginesi clade (7)
Stefania schuberti
Stefania riae
Stefania evansi clade (4)
Stefania riveroi clade (4)
Stefania woodleyi clade (5)
Fritziana (9)
Eotheca/G. fissipes s.g. (4)
Cryptotheca/G. longipes s.g. (1)
Amphignathodon/G. longipes s.g. (6)
Opisthodelphys/G. marsupiata s.g. (1)
Australotheca / G. microdiscus s.g.(4)
Eotheca/G. fissipes s.g. (3)
Gastrotheca (1)
Duellmania (6)
Edaphotheca (1)
Duellmania (8)
Gastrotheca (25)
G. marsupiata s.g.

Figure 4. Summarized topologies of (a) tree alignment + parsimony (TAPa), (b) similarity-alignment + parsimony, gaps as a fifth state (SAP5th), (c) similarity-alignment + parsimony, gaps as binary characters (SAPg), (d) similarity-alignment + parsimony, gaps as missing data (SAPm) strict consensus trees, (e) similarity-alignment + maximum likelihood, gaps as binary characters (SALg) and (f) similarity-alignment + maximum likelihood, gaps as missing data (SALm) optimal trees, showing the main phylogenetic relationships among genera and the supraspecific groupings proposed by Castroviejo-Fisher et al. (2015) and Duellman (2015) for *Gastrotheca*, and by Kok et al. (2017) for *Stefania*. Numbers on branches are jackknife or bootstrap proportions (left) in parsimony and maximum likelihood results, respectively, and Goodman-Bremer values (right) in parsimony results. Numbers in parentheses after taxon names indicate number of sampled species. An asterisk denotes jackknife or bootstrap of 100, two asterisks a non-monophyletic Athesphatanura, excluding Dendrobatoidea, and an en dash denotes bootstrap or jackknife < 50.

Wildcard taxa

Searches for wildcard taxa were performed for trees of all parsimony analyses. *Cryptobatrachus* sp. and *Gastrotheca* sp. 1 are among the first ten terminals with the greater potential to cause resolution decrease in the strict consensus of all four parsimony analyses. For SAPg and SAPm, *Gastrotheca excubitor* MUSM 26280, a terminal poorly represented both in the molecular and phenotypic datasets, has the highest potential to behave as a wildcard. For SAP5th and TAPa, *Cryptobatrachus* sp. and *Hemiphractus johnsoni*, respectively, are the terminals with the highest potential to have unstable behavior and both are represented by phenotypic characters only. A detailed list of taxa ordered by decreasing potential of wildcard behavior, for each parsimony analysis, is presented in Tables S5–S8.

Impact of phenotypic data

For SAL analyses, the inclusion of the phenotypic dataset hardly had an effect on the inferred intergeneric relationships within Hemiphractidae except for a few minor arrangements within *Hemiphractus* and *Gastrotheca* (Figs. S26, S27 and S32, S33). Also relevant is that adding the phenotypic dataset modified the optimal topology in the exact same way regardless of how we coded indels (Table S9). The results are very different within parsimony. For example, for all parsimony analyses except SAP5th, the inferred relationships among genera and among species of *Hemiphractus* and *Gastrotheca* are different when comparing the optimal topologies of the combined dataset with those from the molecular-only one for each of the four types of analyses (Figs. S22–S25 and S28–S31). Also, the inclusion of phenotypic characters increased the resolution of optimal trees in SAP5th and SAPm—from six trees and five polytomies to a single tree with no polytomies and from nine trees and seven polytomies to two trees and one polytomy, respectively—, had no effect with each dataset yielding a single tree with no polytomies in TAPa, and decrease their resolution in the total evidence dataset with eight trees and seven polytomies to eight trees and eight polytomies SAPg results. Finally, the addition of phenotypic characters to the molecular dataset caused an increase in the number of shared clades in 80% of the comparisons, a decrease in 13%, while it remained constant for the two likelihood analyses. Of the eight comparisons between likelihood and parsimony, the addition of phenotypic characters to the molecular dataset caused an increase in the number of shared clades in all instances.

Discussion

More taxa and more characters: considerations concerning the studies of Castroviejo-Fisher et al. (2015) and Duellman (2015)

Our TAPa analysis applied the same analytical strategy as Castroviejo-Fisher et al. (2015). However, our TAPa dataset has 46% more hemiphractid terminals, 359 more DNA sequences for up to 13 genes (mitochondrial and nuclear), additional data for external morphology and behavioral characters of 21 terminals (*Fritziana* and *Gastrotheca*), and for osteological characters of three *Hemiphractus* terminals. Observed differences between our TAPa results and those of

Table 3
Number of clades (in parenthesis) and clades shared by parsimony strict consensus trees and optimal maximum likelihood trees

	TAPa (256)	SAP5th (255)	SAPm (247)	SAPg (225)	SALg (269)	SALm (269)
SAP5th	206					
SAPm	215	203				
SAPg	209	197	222			
SALg	192	199	193	188		
SALm	194	198	198	191	256	-

Total number of terminals = 270.

Castroviejo-Fisher et al. (2015) can only be attributed to the increase in evidence.

The results of Castroviejo-Fisher et al. (2015) and ours agree on recovering Brachycephaloidea as non-monophyletic and a monophyletic Athesphatanura as sister to Hemiphractidae. Hemiphractinae is not recovered in our TAPa analysis. Considering the relationships among Hemiphractidae genera, the only common result between our TAPa analysis and Castroviejo-Fisher et al. (2015) is the sister relationship of *Gastrotheca* and *Hemiphractus*.

Our TAPa analysis recovered a monophyletic *Flectonotus* within a polytomy of *Cryptobatrachus* terminals. The non-monophyly of *Cryptobatrachus* is caused by *Cryptobatrachus* sp.—a specimen only represented by phenotypic characters—as it resulted as sister of *Flectonotus* in 20% of the TAPa optimal trees. *Cryptobatrachus* sp. is the second most probable terminal to decrease resolution in the strict consensus of TAP results. To figure out why *Cryptobatrachus* sp. is collapsing the monophyly of *Cryptobatrachus* in our study but not in Castroviejo-Fisher et al. (2015), we performed a simple analysis. Since *Cryptobatrachus* sp. is represented only by phenotypic characters, one can logically conclude that only terminals with phenotypic characters can influence its position within the tree. Therefore, we analyzed a matrix of 39 terminals and the 51 phenotypic characters—we only included all the well-represented terminals for the phenotypic characters—in TNT under Traditional Search (results not shown). If we forced *Fritziana* as sister of *Cryptobatrachus* and *Flectonotus* such as in our results of TAPa, SAPm, and SAPg the position of *Cryptobatrachus* sp. as sister to the two *Flectonotus* terminals is as optimal as both genera being sisters and reciprocally monophyletic. However, if *Fritziana* is sister of *Stefania* and this clade is sister of *Cryptobatrachus* and *Flectonotus*, the optimal solution is the monophyly of *Cryptobatrachus* (i.e., the position of *Cryptobatrachus* sp. as sister to the two *Flectonotus* terminals adds one step to the tree length). We conclude that adding additional sequences and terminals of *Fritziana* and *Stefania* caused changes in their relative positions within Hemiphractidae with respect to the results of Castroviejo-Fisher et al. (2015), which caused changes in the polarity and number of transformations of the phenotypic characters of *Cryptobatrachus* sp. This result carries a valuable lesson: the addition of characters for a subset of terminals can affect the relationships of other taxa that are not even scored for those data (de Sá et al., 2014).

Differences between Castroviejo-Fishers et al's. (2015) dataset and ours are mainly related to taxa and DNA sequences of *Fritziana* and *Stefania*. We added a relatively fewer number of taxa and characters for *Gastrotheca* and *Hemiphractus*. Regarding *Fritziana* relationships, the only common result is the position

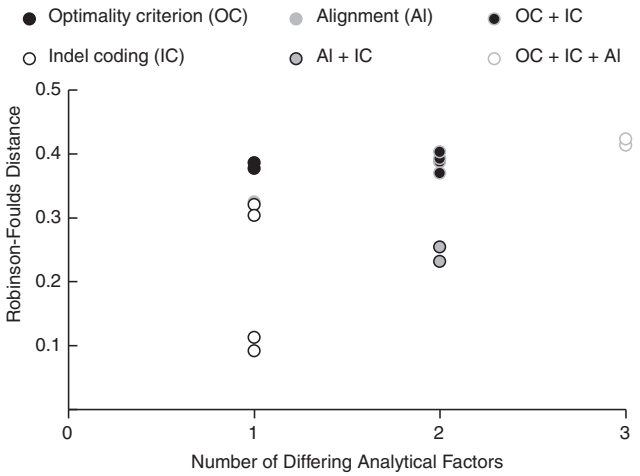

Figure 5. Robinson-Foulds distances between pairs of trees ordered according to the number of differing analytical factors (optimality criterion, alignment, and indel coding) between phylogenetic analyses. There is a positive additive effect in all cases (i.e., an increase in differing analytical factors implies an increase in the differences between optimal topologies) except for two cases in parsimony. Coding indels as unknown nucleotides or as an additional presence/absence dataset following the Simple Indel Coding method (SIC) of Simmons and Ochoterena (2000) instead of a fifth character makes tree-alignment converge with similarity-alignment.

of *F. tonimi* as sister of all other *Fritziana*. In contrast, the relationships among *Stefania* clades are stable, and only the position of *S. riae* is different. Within *Gastrotheca*, relationships among species groups are very similar. However, *G. ovifera* is sister of the *G. fissipes* species group from the Atlantic Forest (Fig. 3d) and not of the Andean species of the *G. marsupiata* species group as in Castroviejo-Fishers et al. (2015). Finally, within *Hemiphractus*, the most important change is the position of a clade of all *H. scutatus* terminals as sister of all other *Hemiphractus* in our TAPa optimal trees and a decrease in resolution regarding the relationships of six terminals (Fig. 3c).

Our SALm analysis is similar to the analytical strategy of Duellman (2015). However, the present study has 86% more terminals of egg-brooding frogs, 124 more outgroup terminals, up to 51 phenotypic characters for 243 terminals, and up to 9 580 more bp from 16 nuclear and mitochondrial genes. Both studies agreed on recovering Hemiphractidae as sister of a monophyletic Brachycephaloidea and the monophyly of Cryptobatrachinae, Hemiphractinae, and all their genera. However, within Hemiphractinae, *Fritziana* is sister of *Gastrotheca*, instead of *Stefania* as in Duellman (2015). The limited number of terminals of Duellman (2015) dataset prevented comparisons of the relationships within *Cryptobatrachus*, *Fritziana*, and *Hemiphractus*. Regarding *Gastrotheca*, of the five non-monotypic subgenera proposed by Duellman (2015) we only recover as monophyletic *Amphignathodon* and *Australotheca* (*Cryptotheca* is only represented by

Table 4
Bootstrap (BS), jackknife (JK), and Goodman Bremer (GB) values for the main clades considered in this study

	TAPa (JK/GB)	SAP_{5th} (JK/GB)	SAPg (JK/GB)	SAPm (JK/GB)	SALg (BS)	SALm (BS)
Athesphatanura	99/31	–	64/14	57/25	–	–
Brachycephaloidea	–	56/12	–	–	65	92
Hemiphractidae	99/71	**/13	59/11	76/7	59	93
Athesphatanura + Hemiphractidae	99/58	–	**/27	52/30		–
Brachycephaloidea + Hemiphractidae	–	**/15	–	–	–	61
Cryptobatrachinae Frost et al. (2006)	94/3	**/13	53/9	71/5	94	97
Hemiphractinae Peters (1862)	–	–	–	–	55	91
Cryptobatrachus	–	60/15	–	–	97	97
Flectonotus	90/3	95/20	93/3	92/6	100	100
Fritziana	100/7	99/80	99/84	99/77	100	100
Gastrotheca	100/43	100/69	99/47	100/41	57	90
Species groups (Castroviejo-Fisher et al., 2015)						
G. fissipes species group	100/19	–	**/3	**/10	–	–
G. longipes species group	100/33	–	51/7	67/9	56	91
G. marsupiata species group	–	–	–	–	–	–
G. microdiscus species group	100/46	91/19	89/16	89/15	83	94
Subgenera (Duellman, 2015)						
Amphignathodon	99/18	86/9	84/41	80/31	81	84
Australotheca	100/46	91/19	89/16	89/15	83	94
Duellmania	–	–	–	–	–	–
Eotheca	100/19	–	**/3	**/10	–	–
Gastrotheca	–	–	–	–	–	–
Hemiphractus	100/63	99/39	100/19	100/12	94	92
Stefania	100/4	99/55	99/10	99/11	98	98
Stefania clades (Kok et al., 2017)						
evansi clade	100/46	99/1	99/36	99/7	100	100
ginesi clade	100/4	99/31	99/28	99/16	100	99
riveroi clade	100/25	99/36	99/27	99/23	100	100
woodleyi clade	100/38	99/40	99/26	99/31	98	99

One asterisk refers to Athesphatanura without Dendrobatoidea, two asterisks denote jackknife < 50, while an en dash indicates non-monophyly in the optimal tree(s).

G. walkeri). The position of the monotypic *Opisthodelphys* (*G. ovifera*) also differs from that inferred by Duellman (2015).

We demonstrate that both Castroviejo-Fisher et al's. (2015) and Duellman's (2015) results are sensitive to the addition of taxa and characters, even when the increase of evidence is relatively limited as in the case of the dataset of Castroviejo-Fisher et al. (2015). If we consider that the amount of missing data in our dataset is still large (Table 1) and that 27 described species are not represented in our analyses, we can expect further changes among the relationships of hemiphractids, even if analyses remain constant. Another relevant conclusion is that the divergence observed between the results of Castroviejo-Fisher et al. (2015) and Duellman (2015) has not decreased despite identical taxon and character sampling. Below, we explore other analytical factors as potential causes for the observed differences.

The impact of optimality criteria

The comparison of SAPg vs. SALg, SAPm vs. SALm, and TAPa vs. the two SAL strategies allowed the identification of qualitatively important differences attributable to the use of ML or parsimony (e.g. the sister relationship of *Fritziana* and *Gastrotheca* vs. the sister relationship of *Hemiphractus* and *Gastrotheca*, and the monophyly vs. the non-monophyly of Hemiphractinae). When evaluated individually or in addition to indel coding and/or alignment, optimality criterion produces the strongest differences (Fig. 5, Tables S3 and S4). We want to highlight that the discussed clades differing among analyses have resampling support values <75 in at least some of our analyses (Fig. 4, S1–S6; Table 4). For example, Hemiphractidae and its sister clade, the sister relationship of *Fritziana* or *Stefania*, and the monophyly of the *Gastrotheca fissipes* and *G. longipes* species groups. Nonetheless, some clades that received JK or BS < 50 in some of the analyses are shared by the results of all of the analyses, such as Cryptobatrachinae and Hemiphractidae.

Besides these differences, other patterns emerge. First, even in the more distant results, what dominates is congruence among the content of clades (Table 4), a pattern already brought up by Rindal and Brower (2010) and Cabra-García and Hormiga (2020). Second,

some important qualitative differences occur within the same optimality criterion, such as *Fritziana* as sister of *Stefania* or *Cryptobatrachus* + *Flectonotus*, a monophyletic Brachycephaloidea or Atesphatanura sister of Hemiphractidae, and the monophyly or not of *Cryptobatrachus* (Fig. 4a–d). This indicates that analytical factors other than optimality criterion have an important impact on our understanding of evolutionary history due to the asymmetry between quantitative and qualitative differences, here illustrated by the impact in the supraspecific taxonomy. For example, although the results between TAPa and SAP5th are quantitatively more similar than between them with any of the SAL results, the taxonomies of SAP5th and SAL share a monophyletic Brachycephaloidea sister of Hemiphractidae, a monophyletic *Cryptobatrachus*, and a non-monophyletic *Gastrotheca fissipes* species group.

A different issue, with more pragmatic implications, is the limitation of current parametric phylogenetic software to report more than a single optimal tree, due to implementation issues or computational difficulties (Simmons, 2012; Simmons and Kessenich, 2019). This masks lack of or ambiguous evidence in datasets so clades recovered as a polytomy by parsimony analyses but completely resolved by ML analyses must be interpreted cautiously, particularly in the context of missing data (Lemmon et al., 2009; Simmons, 2012, 2014; Simmons and Goloboff, 2013). As mentioned before, the completely resolved relationships among *Cryptobatrachus* terminals recovered by both SALm and SALg analyses (Figs. S5 and S6), as opposed to a complete polytomy in all parsimony strict consensuses (Figs. S1–S4; Fig. 4), showed that available evidence is not enough to unambiguously resolve relationships within *Cryptobatrachus*. The terminal *Cryptobatrachus* sp. is represented only by 48 out of 51 phenotypic characters, and the only other terminal within the genus with more than three phenotypic characters is *C. fuhrmanni* JDL 14865 (49/51 phenotypic characters). Furthermore, those three shared phenotypic characters are not variable among *Cryptobatrachus* species. In addition, among terminals with DNA data, no molecular marker is shared among all *Cryptobatrachus* terminals. Simply put, the available evidence is not adequate to resolve the relationships within that clade and any sensible implementation of an optimality criterion should reflect this absence of evidence. Padial et al. (2014) provided a similar empirical example with *Eleutherodactylus* frogs. Solutions to these artifacts have been proposed (e.g., Nguyen et al., 2014; Chernomor et al., 2016; Biczok et al., 2018) but it is unclear if these approaches are able to correct empirical cases such as the one related here for *Cryptobatrachus*.

The impact of alignment strategies

We expected the optimal topologies of SAP5th and TAPa to be more similar despite the different nucleotide homology hypotheses, inferred through similarity and tree-alignment, respectively. However, the impact of alignment strategies as measured by Robinson-Foulds and pairwise SPR distances is nearly as strong as optimality criterion and stronger than differences derived by indel coding within parsimony or ML (Tables S3 and S4, Figs. 4 and 5). The explanation is that while similarity-alignment searches for pairwise minimum phenetic distances among sequences, tree-alignment searches for the best topology-specific statements of homology as judged by an optimality criterion that allows transformations through time (e.g., parsimony and maximum likelihood). Also, through direct optimization indels appear as transformations linking ancestral and descent nucleotide sequences, instead as patterns implied by MSA (Wheeler, 1996; Wheeler et al., 2006). This is also clearly shown by calculating the costs of SAP and SAL optimal trees under IPO in POY. In all cases (Table 2), the results demonstrate that there are more optimal alignments, as judged by equally weighted parsimony, for those topologies.

We are only aware of three previous studies that have quantitatively evaluated if optimality criterion affects relationships more or less than alignment strategy (tree- vs. similarity-alignment; Padial et al., 2014; Goicoechea et al., 2016; Cabra-García and Hormiga, 2020). Unfortunately, the results of those studies are only partially comparable to ours. For example, Goicoechea et al. (2016) only reported shared clades to compare topologies and Padial et al. (2014) and Cabra-García and Hormiga (2020), when using parsimony with similarity alignment, coded gaps as a fifth character state and did not evaluate gaps as unknown nucleotides. In general, our results are similar to those of Padial et al. (2014) and Cabra-García and Hormiga (2020), with a slightly smaller impact of alignment strategy over topological distances when compared to optimality criterion. Goicoechea et al. (2016), in a study of teioid lizards, reported that the two most similar topologies were those resulting from their SALm and TAPa, while the most different results were obtained by their SAP5th and SAPm. Their results have important implications because they question if similarity-alignment is appropriate for parsimony analysis and if tree-alignment under parsimony is more compatible with similarity-alignment under ML (Goicoechea et al., 2016). Our results add to this discussion. On the one hand, our results of SAP5th analysis are the most divergent within parsimony, which is compatible with Goicoechea et al's. (2016) results; on

the other hand, our TAPa and SAL results are the most different topologies, which is not.

In summary, our results are more compatible with Padial et al. (2014) and Cabra-García and Hormiga (2020), although those studies did not evaluate SAPm. However, sample size is still too small (only four studies, including ours) and differences among the performed comparisons are important. We conclude that the data are still too scarce to provide a general pattern except that using tree alignment or similarity alignment under parsimony with gaps coded as a fifth character state retrieves trees that are about as different as using similarity alignment with parsimony or likelihood.

The impact of indel coding in Hemiphractidae phylogenetics

Within SAP analyses, the SAP5th strict consensus is the most resolved, followed by those of SAPm and SAPg analyses (Table 3). The strict consensus of SAPg includes the same unresolved clades as SAPm plus some additional ones (Figs. S3 and S4). One expects that the diminished resolution of the SAPg strict consensus is caused by the reduction of information from the complete similarity-alignment to 1421 SIC characters (Simmons et al., 2007). However, this cannot explain that the strict consensus of SAPm, inferred from a dataset with no indel information, includes fewer polytomies than that of SAPg. We suspect that the 1421 SIC characters of SAPg retained a higher relative proportion of contradictory information than the complete similarity-alignment.

In addition, there are important topological differences among the results of the three SAP analyses. The strict consensus of the SAP5th strategy has several discordances with the strict consensus trees of SAPg and SAPm, the latter two analyses recovered more similar topologies (Figs. 4–5, S2–S4; Table 3). In other words, parsimony analyses resulted in more similar topologies when dismissing the pattern of indels inferred through MSA or by coding gaps as separate characters. Within Hemiphractidae, the support values of the three SAP strategies were very similar and overall high. These results indicate that the different treatments of indel information did not have any effect over the resampling measure.

The two SAL analyses recovered the most similar trees among all comparisons and are similar to the distances found between SAPm and SAPg trees (Figs. 4, S5 and S6; Table 3). The different treatment of indels in ML has more obvious effect on BS. Those of the SALm analysis are in general higher within Hemiphractidae (Table 4). State-of-the-art parametric methods in phylogenetics recode each gap as an unknown nucleotide and the role of different treatments of indels in a parametric framework has been poorly explored (e.g., Egan and Crandall, 2008; Nagy et al., 2012; Luan et al., 2013; Boutte et al., 2019). Although the quantitative effects of coding indels differently in our ML analysis are the lowest when compared to the other analytical factors evaluated in our study, they are relevant and highlight the importance of improving computational features of methods that incorporate indel information into ML and Bayesian phylogenetic software (Fleissner et al., 2005; Lunter et al., 2005; Redelings and Suchard, 2005; Suchard and Redelings, 2006; Novák et al., 2008; Westesson et al., 2012). For example, the latest version of POY allows the inclusion of indels as a fifth character state in ML analysis, both in static and dynamic contexts (Wheeler et al., 2015). However, relative large datasets such as ours remain computationally intractable[1].

In summary, the quantitatively relevant phylogenetic information of indel characters changed the topology, resolution, and resampling measures of clade support in our analyses. Although the impact of different indel coding strategies varied, it was never negligible (Table 3, Fig. 5).

Additive effects

One could expect that an increment in the differences of analytical factors implies an increase in the differences between results due to an additive effect. Our results offer a mixed signal for such an additive effect (Fig. 5). When analyses differ in both optimality criterion and indel coding, the Robinson-Foulds distances always increase with respect to comparisons between analyses that just differ in one of those two factors. This is also the case when analyses differ in alignment strategy, optimality criterion, and indel coding. However, within parsimony the differences caused by alignment strategy decrease when indels are not coded as a fifth state. It seems that the pattern of indels inferred by similarity-alignment contains such different phylogenetic information from that of tree-alignment, that decreasing their weight either partially (SAPg) or completely (SAPm) increases similarity among optimal topologies.

The impact of phenotypic data

Our results show that the addition of a few phenotypic characters to a three orders of magnitude larger molecular dataset can impact the results in different

[1] A ML search, coding indels as a fifth state, of part of our dataset (252 terminals and similarity alignment of 12S and 16S sequences) did not finish after 20 days running on one processor (adding processors decreased performance).

ways. Regarding the resolution of optimal trees, we found a positive or neutral effect in five out of six analyses (two and three, respectively), while it decreased tree resolution only in SAPg analysis. This generally positive or neutral effect is commonly reported in the literature (Nylander et al., 2004; de Sá et al., 2014; Chakrabarty et al., 2017; Mirande, 2017; Koch and Gauthier, 2018; Martin et al., 2018; Sánchez-Pacheco et al., 2018; Cabra-García and Hormiga, 2020). Another positive result (at least in our view) is that the addition of the phenotypic characters increased congruence up to 12% among the optimal topologies of the different analyses in 80% of the comparisons. Taken together, these results flag phenotypic characters as a promising source of information to improve our understanding of egg-brooding frogs' evolutionary history.

Goloboff et al. (2019) found that the common mechanism assumption of the Mkv model is rarely met by empirical phenotypic data. This violation could then cause important divergences between the optimal topologies selected by methods using this model and those that do not. When we compare the results of SAPm and SAPg with those of SALm and SALg, with and without the phenotypic dataset, we found an interesting result. The optimal topologies of SAPm and SALm and those of SAPg and SALg became 8% and 11% more similar (i.e., more shared clades), respectively, with the addition of the phenotypic dataset. Furthermore, both SAPm and SALg as well as SAPg and SALm yield optimal trees 11% and 8% more similar, respectively, when we added the phenotypic dataset. The effects of analyzing phenotypic characters, as well as indels, under the Mkv model should be carefully investigated. For example, the predictions of Goloboff et al. (2019) regarding branch lengths, parsimony, and the common mechanism of the Mkv model could be applied to indels coded as binary characters.

The supraspecific taxonomy of Hemiphractidae

Our six analytical strategies recover the monophyly of Hemiphractidae in accordance with the two previously most extensive phylogenies of egg-brooding frogs (Castroviejo-Fisher et al., 2015; Duellman, 2015). To discuss the supraspecific taxonomy of Hemiphractidae, we use the results of the strict consensus resulting from the TAPa analysis because it is the only analytical strategy that tested nucleotide homologies dynamically, using indels as characters, applying a logically consistent optimality criterion between alignment and tree searches, while simultaneously optimizing the characters of the phenotypic matrix, on multiple alternative trees. This strategy maximized the explanation of the observed differences, while simultaneously interpreting evidence more conservatively and thoroughly exploring tree space. A list of phenotypic synapomorphies is presented in Appendix S3.

We find novel phylogenetic relationships among and within genera and the monophyly of Hemiphractinae *sensu* Castroviejo-Fisher et al. (2015) is not recovered. Although the TAPa analysis recovers a non-monophyletic *Cryptobatrachus*, we abstain from making nomenclatural changes because a single terminal coded for just 48 phenotypic characters is causing the polytomy. A better character and taxon sampling of this genus is needed. The TAPa analysis recovers a monophyletic *Flectonotus* with the same phenotypic synapomorphies found by Castroviejo-Fisher et al. (2015) (Appendix S3). We recover a monophyletic *Fritziana* in accordance with previous results (Duellman et al., 2011; Schmid et al., 2012; Blackburn and Duellman, 2013; Castroviejo-Fisher et al., 2015; Walker et al., 2016, 2018a). Our analysis did not recover any of the phenotypic synapomorphies found by Castroviejo-Fisher et al. (2015), the only phenotypic synapomorphy for *Fritziana* is the pouch type (Appendix S3). With respect to *Stefania*, all the clades proposed by Kok et al. (2017) are monophyletic. As in Castroviejo-Fisher et al. (2015), no phenotypic synapomorphies were recovered for *Stefania*. Regarding the relationships within *Hemiphractus*, there are no nomenclatural implications because it does not include supraspecific taxa, notwithstanding the much needed description of all the putative new species reported herein. Our TAPa analysis recovers a subset of the phenotypic synapomophies recovered by Castroviejo-Fisher et al. (2015) (Appendix S3).

For *Gastrotheca* the same synapomorphies found by Castroviejo-Fisher et al. (2015) were recovered. According to our results, current taxonomies are incompatible with the principle of monophyly. The monophyly of the supraspecific taxonomy of Castroviejo-Fisher et al. (2015) is easier to achieve (i.e., requires fewer nomenclatural changes). Consequently, we provide an updated taxonomy following the species groups of *Gastrotheca* suggested by Castroviejo-Fisher et al. (2015). For those interested in using the subgenera of Duellman (2015), we also suggest a new taxonomy based on the principle of monophyly that renders the content of both taxonomies equivalent, so that users are referring to the same entities regardless of the name (Appendix 1).

Acknowledgments

We are grateful to Juan Carlos Chaparro (MUBI), Mirco Solé, Victor Dill Orrico, Caio Vinicius De Mira-Mendes (UESC), and Giussepe Gagliardi-Urrutia for collaborating with tissue samples and

specimens. We thank Lourdes Alcaraz (MNCN-CSIC) for helping with the molecular lab. We are also grateful to Andrew J. Crawford, Juan M. Guayasamin, and Martín Pereyra for their valuable comments on an earlier version of this manuscript. Ward Wheeler and Taran Grant kindly answered our questions regarding maximum likelihood analysis in POY. Two reviewers and the Associate Editor helped us to improve previous versions of the manuscript with their constructive criticism. Funding for this study was provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil to LYE and SC-F (scholarship numbers 830432/1999-0 and 312744/2017-0), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil to LYE (scholarship number 88887.179352/2018-00) and projects CGL2014-53523-P and PGC2018-097421-B-100 of the Spanish Government (PI, Ignacio De la Riva).

Conflict of interest

None declared.

References

Biczok, R., Bozsoky, P., Eisenmann, P., Ernst, J., Ribizel, T., Scholz, F., Trefzer, A., Weber, F., Hamann, M. and Stamatakis, A., 2018. Two C++ libraries for counting trees on a phylogenetic terrace. Bioinformatics 34, 3399–3401.

Blackburn, D.C. and Duellman, W.E., 2013. Brazilian marsupial frogs are diphyletic (Anura: Hemiphractidae: *Gastrotheca*). Mol. Phylogenet. Evol. 68, 709–714.

Boutte, J., Fishbein, M., Liston, A. and Straub, S.C., 2019. NGS-Indel Coder: a pipeline to code indel characters in phylogenomic data with an example of its application in milkweeds (*Asclepias*). Mol. Phylogenet. Evol. 139, 106534.

Brandley, M.C., Schmitz, A. and Reeder, T.W., 2005. Partitioned Bayesian analysis, partition choice, and the phylogenetic relationships of scincid lizards. Syst. Biol. 54, 373–390.

Bremer, K., 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42, 795–803.

Brown, J.M. and Lemmon, A.R., 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst. Biol. 56, 643–655.

Butler, A.D., Edgecombe, G.D., Ball, A.D. and Giribet, G., 2010. Resolving the phylogenetic position of enigmatic New Guines and Seychelles Scutigeromorpha (Chilopoda): a molecular and morphological assessment of Ballonemini. Invertebr. Syst. 24, 539–559.

Cabra-García, J. and Hormiga, G., 2020. Exploring the impact of morphology, multiple sequence alignment and choice of optimality criteria in phylogenetic inference: a case study with the Neotropical orb-weaving spider genus *Wagneriana* (Araneae: Araneidae). Zool. J. Linnean Soc. 188, 976–1151.

Carvajal-Endara, S., Coloma, L.A., Morales-Mite, M.A., Guayasamin, J.M., Szekely, P. and Duellman, W.E., 2019. Phylogenetic systematics, ecology, and conservation of marsupial frogs (Anura: Hemiphractidae) from the Andes of southern Ecuador, with descriptions of four new biphasic species. Zootaxa 4562, 1–102.

Castroviejo-Fisher, S., Padial, J.M., De la Riva, I., Pombal, J.P. Jr, Da Silva, H.R., Rojas-Runjaic, F.J.M., Medina-Méndez, E. and Frost, D.R., 2015. Phylogenetic systematics of egg-brooding frogs (Anura: Hemiphractidae) and the evolution of direct development. Zootaxa 4004, 1–75.

Chakrabarty, P., Faircloth, B.C., Alda, F., Ludt, W.B., Mcmahan, C.D., Near, T.J., Dornburg, A., Albert, J.S., Arroyave, J., Stiassny, M.L.J., Sorenson, L. and Alfaro, M.E., 2017. Phylogenomic systematics of ostariophysan fishes: ultraconserved elements support the surprising non-monophyly of characiformes. Syst. Biol. 66, 881–895.

Chernomor, O., von Haeseler, A. and Minh, B.Q., 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Syst. Biol. 65, 997–1008.

Chippindale, P.T. and Wiens, J.J., 1994. Weighting, partitioning, and combining characters in phylogenetic analysis. Syst. Biol. 43, 278–287.

de Queiroz, A. and Gatesy, J., 2007. The supermatrix approach to systematics. Trends Ecol. Evol. 22, 34–41.

de Sá, R.O., Grant, T., Camargo, A., Heyer, W.R., Ponssa, M.L. and Stanley, E., 2014. Systematics of the neotropical genus *Leptodactylus* Fitzinger, 1826 (Anura: Leptodactylidae): phylogeny, the relevance of non-molecular evidence, and species accounts. S. Am. J. Herpetol. 9, S1–S128.

Del Pino, E.M., 1980. Morphology of the pouch and incubatory integument in marsupial frogs (Hylidae). Copeia 10–17.

Driskell, A.C., Ané, C., Burleigh, J.G., McMahon, M.M., O'Meara, B.C. and Sanderson, M.J., 2004. Prospects for building the tree of life from large sequence databases. Science 306, 1172–1174.

Duellman, W.E., 1980. A new species of marsupial frog (Hylidae: *Gastrotheca*) from Venezuela. Occas. Pap. Mus. Zool. Univ. Michigan 690, 1–7.

Duellman, W.E., 2015. Marsupial Frogs *Gastrotheca* and Allied Genera. Baltimore, Maryland: Johns Hopkins University Press, 408 pp. + XVIII.

Duellman, W.E., Jungfer, K.H. and Blackburn, D.C., 2011. The phylogenetic relationship of geographically separated "*Flectonotus*" (Anura: Hemiphractidae), as revealed by molecular, and morphological data. Phyllomedusa 10, 29–43.

Duellman, W.E., Barley, A.J. and Venegas, P.J., 2014. Cryptic species diversity in marsupial frogs (Anura: Hemiphractidae: *Gastrotheca*) in the Andes of northern Peru. Zootaxa 3768, 159–177.

Duellman, W.E. and Venegas, P.J., 2016. Diversity of marsupial frogs (Anura: Hemiphractidae: *Gastrotheca*) in the northern Cordillera Central, Peru, with the descriptions of two new species. Phyllomedusa 15, 103–117.

Edgar, R.C., 2004. MUSCLE: multiple alignment with high accuracy and high throughput. Nuclei Acids Res. 32, 1792–1797.

Eernisse, D.J. and Kluge, A.G., 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. Mol. Biol. Evol. 10, 1170–1195.

Egan, A.N. and Crandall, K.A., 2008. Incorporating gaps as phylogenetic characters across eight DNA regions: ramifications for North American Psoraleeae (Leguminosae). Mol. Phyl. Evol. 46, 532–546.

Farris, J.S., Albert, V.A., Källersjö, M., Lipscomb, D. and Kluge, A.G., 1996. Parsimony jackknifing outperforms neighbor-joining. Cladistics 12, 99–124.

Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27, 401–410.

Fleissner, R., Metzler, D. and Von Haeseler, A., 2005. Simultaneous statistical multiple alignment and phylogeny reconstruction. Syst. Biol. 54, 548–561.

Folly, M., Hepp, F., Carvalho-e-Silva, S.P. & Duellman, W.E., 2014. Taxonomic status and redescription of *Flectonotus ulei* (Anura: Hemiphractidae), with a key for the species of *Fritziana*. Zoologia (Curitiba) 31, 393–399.

Frost, D.R., Grant, T., Faivovich, J., Bain, R., Haas, A., Haddad, C.F.B., de Sá, 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. and Wheeler, W.C., 2006. The amphibian tree of life. Bull. Am. Mus. Nat. Hist. 297, 1–370.

Frost, D.R., 2020. Amphibian Species of the World: an Online Reference. Version 6.0 (August 2019). Electronic Database accessible at http://research.amnh.org/herpetology/amphibia/index.html. American Museum of Natural History, New York, USA.
Gatesy, J., Baker, R.H. and Hayashi, C., 2004. Inconsistencies in arguments for the supertree approach: supermatrices versus supertrees of Crocodylia. Syst. Biol. 53, 342–55.
Giribet, G., 2015. Morphology should not be forgotten in the era of genomics – a phylogenetic perspective. Zool. Anz. 256, 96–103.
Goicoechea, N., Frost, D.R., De la Riva, I., Pellegrino, K.C.M., Sites, J. Jr, Rodrigues, M.T. and Padial, J.M., 2016. Molecular systematics of teioid lizards (Teioidea/Gymnophthalmoidea: Squamata) based on the analysis of 48 loci under tree-alignment and similarity-alignment. Cladistics 32, 624–671.
Goloboff, P.A., 1996. Methods for faster parsimony analysis. Cladistics 12, 199–220.
Goloboff, P.A., 1999. Analyzing large data sets in reasonable times: solutions for composite optima. Cladistics 15, 415–428.
Goloboff, P.A., 2008. Calculating SPR distance between trees. Cladistics 24, 591–597.
Goloboff, P.A., 2014. Hide and vanish: data sets where the most parsimonious tree is known but hard to find, and their implications for tree search methods. Mol. Phyl. Evol. 79, 118–131.
Goloboff, P.A., Farris, J.S., Källersjö, M., Oxelman, B., Ramírez, M.J. and Szumik, C., 2003. Improvements to resampling measures of group support. Cladistics 19, 324–332.
Goloboff, P.A., Farris, J. and Nixon, K., 2008. TNT, a free program for phylogenetic analysis. Cladistics 24, 774–786.
Goloboff, P.A. and Catalano, S., 2016. TNT, version 1.5, with a full implementation of phylogenetic morphometrics. Cladistics 32, 221–238.
Goloboff, P.A., Pittman, M., Pol, D. and Xing, X., 2019. Morphological data sets fit a common mechanism much more poorly than DNA sequences and call into question the Mkv model. Syst. Biol. 68, 494–504.
Gosner, K.L. 1960. A simplified table for starting anuran embryos and larvae, with notes on identification. Herpetoloy, 16, 183–190.
Goodman, M., Olson, C.B., Beeber, J.E. and Czelusniak, J., 1982. New perspectives in the molecular biological analysis of mammalian phylogeny. Acta Zool. Fenn. 169, 19–35.
Grant, T., 2019. Outgroup sampling in phylogenetics: severity of test and successive outgroup expansion. J. Zool. Syst. Evol. Res. 57, 748–763.
Grant, T. and Kluge, A.G., 2008. Credit where credit is due: the Goodman-Bremer support metric. Mol. Phylogenet. Evol. 49, 405–406.
Guayasamin, J.M., Castroviejo-Fisher, S., Trueb, L., Ayarzagüena, J., Trueb, L. and Vilà, C., 2008. Phylogenetic relationships of glassfrogs (Centrolenidae) based on mitochondrial and nuclear genes. Mol. Phylogenet. Evol. 48, 574–595.
Hill, R.L., Martin, K.G., Stanley, E. and Mendelson, J.R.I., 2018. A taxonomic review of the genus *Hemiphractus* (Anura: Hemiphractidae) in Panama: Description of two new species, resurrection of *Hemiphractus panamensis* (Stejneger, 1917), and discussion of *Hemiphractus fasciatus* Peters, 1862. Zootaxa 4429, 495–512.
Huelsenbeck, J.P., 1995. Performance of phylogenetic methods in simulation. Syst. Biol. 44, 17–48.
Huelsenbeck, J.P., 1997. Is the Felsenstein zone a fly trap? Syst. Biol. 46, 69–74.
Huelsenbeck, J.P. and Hillis, D.M., 1993. Success of phylogenetic methods in the Four-Taxon case. Syst. Biol. 42, 247–264.
Huelsenbeck, J.P., Ronquist, F., Nielsen, R. and Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.
Huelsenbeck, J.P. and Lander, K.M., 2003. Frequent inconsistency of parsimony under a simple model of cladogenesis. Syst. Biol. 52, 641–648.
Juncá, F.A. and Nunes, I., 2008. A new species of marsupial frog of the genus *Gastrotheca* Fitzinger (Anura: Amphignatodontidae) from the state of Bahia, northeastern Brazil. Zootaxa 1907, 61–68.
Katoh, K., Rozewicki, J. and Yamada, K.D., 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. 20, 1160–1166.
Kluge, A.G., 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Syst. Zool. 38, 7–25.
Kluge, A.G., 1997. Testability and the refutation and corroboration of cladistics hypothesis. Cladistics 13, 81–96.
Kluge, A.G., 1998. Total evidence or taxonomic congruence: cladistics or consensus classification. Cladistics 14, 151–158.
Koch, N.M. and Gauthier, J.A., 2018. Noise and biases in genomic data may underlie radically different hypotheses for the position of Iguania within Squamata. PLoS One 13, e0202729.
Kok, P.J.R., Russo, V.G., Ratz, S., Means, D.B., MacCulloch, R.D., Lathrop, A., Aubret, F. and Bossuyt, F., 2017. Evolution in the South American 'Lost World': insights from multilocus phylogeography of stefanias (Anura, Hemiphractidae, *Stefania*). J. Biogeogr. 44, 170–181.
Kolaczkowski, B. and Thorton, J.W., 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431, 980–984.
Kopuchian, C. and Ramírez, M.J., 2010. Behaviour of resampling methods under different weighting schemes, measures and variable resampling strengths. Cladistics 26, 86–97.
Kumar, S. and Filipski, A., 2007. Multiple sequence alignment: in pursuit of homologous DNA positions. Genome Res. 17, 127–135.
Kuraku, S., Zmasek, C.M., Nishimura, O. and Katoh, K., 2013. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res. 41, W22–W28.
Lanfear, R., Calcott, B., Ho, S.Y.W. & Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Phylogenet. Evol. 29, 1695–1701.
Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large data sets. Bioinformatics 30, 3276–3278.
Leaché, A.D. and Reeder, T.W., 2002. Molecular systematic of the eastern fence lizard (*Scleropus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. Syst. Biol. 51, 44–68.
Lee, M.S.Y. and Palci, A., 2015. Morphological phylogenetics in the genomic age. Curr. Biol. 25, R922–R929.
Lemmon, A.R., Brown, J.M., Stanger-Hall, K. and Lemmon, E.M., 2009. The effect of ambiguous data on phylogenetic estimates obtained by Maximum Likelihood and Bayesian inference. Syst. Biol. 58, 130–145.
Li, C., Lu, G. and Ortí, G., 2008. Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. Syst. Biol. 57, 519–539.
Luan, P., Ryder, O.A., Davis, H., Zhang, Y. and Yu, L., 2013. Incorporating indels as phylogenetic characters: impact for interfamilial relationships within Arctoidea (Mammalia: Carnivora). Mol. Phylogenet. Evol. 66, 748–756.
Lunter, G., Miklós, I., Drummond, A., Jensen, J.L. and Hein, J., 2005. Bayesian coestimation of phylogeny and sequence alignment. BMC Bioinformatics, 6, 83.
Machado, D.J., 2015. YBYRÁ facilitates comparison of large phylogenetic trees. BMC Bioinformatics, 16, 204.
Maddison, W.P. and Maddison, D.R., 2017. Mesquite: a modular system for evolutionary analysis. Version 3.31. http://mesquite project.org.
Martin, R.P., Olson, E.E., Girard, M.G., Smith, W.L. and Davis, M.P., 2018. Light in the darkness: new perspective on lanternfish relationships and classification using genomic and morphological data. Mol. Phylogenet. Evol. 121, 71–85.
McGuire, J.A., Witt, C.C., Altshuler, D.L. and Remsen, J.V. Jr, 2007. Phylogentic systematics and biogeography of hummingbirds: Bayesian and Maximum Likelihood analysis of

partitioned data and selection of an appropriate partitioning strategy. Syst. Biol. 56, 837–856.
Mendelson, J.R. III, da Silva, H.R. and Maglia, A.M., 2000. Phylogenetic relationships among marsupial frog genera (Anura: Hylidae: Hemiphractinae) based on evidence from morphology and natural history. Zool. J. Linn. Soc. 128, 125–148.
Milinkovitch, M.C., LeDuc, R.G., Adachi, J., Farnir, F., Georges, M. and Hasegawa, M., 1996. Effects of character weighting and species sampling on phylogeny reconstruction: a case study based on DNA sequence data in cetaceans. Genetics 144, 1817–1833.
Miller, M.A., Pfeiffer, W. & Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA, pp. 1–8.
Mirande, J.M., 2017. Combined phylogeny of ray-finned fishes (Actinopterygii) and the use of morphological characters in large-scale analyses. Cladistics 33, 333–350.
Morrison, D.A. and Ellis, J.T., 1997. Effects of nucleotide sequence alignment on phylogeny estimation, a case study of 18S rDNAs of Apicomplexa. Mol. Biol. Evol. 14, 428–441.
Müller, K., 2005. SeqState – primer design and sequence statistics for phylogenetic DNA data sets. Appl. Bioinform. 4, 65–69.
Müller, K., 2006. Incorporating information from length-mutational events into phylogenetic analysis. Mol. Phylogenet. Evol. 38, 667–676.
Nagy, L.G., Kocsubé, S., Csanádi, Z., Kovács, G.M., Petkovits, T., Vágvölgyi, C. and Papp, T., 2012. Re-mind the gap! Insertion-deletion data reveal neglected phylogenetic potential of the nuclear ribosomal internal transcribed spacer (ITS) of fungi. PLoS One 7, e49794.
Nguyen, L.T., Schmidt, H.A., von Haeseler, A. and Minh, B.Q., 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.
Nixon, K.C., 1999. The parsimony ratchet, a new method for rapid parsimony analysis. Cladistics 15, 407–414.
Nixon, K.C. and Carpenter, J.M., 1993. On outgroups. Cladistics 9, 413–426.
Novák, A., Miklós, I., Lyngsø, R. and Hein, J., 2008. StatAlign: an extendable software package for joint Bayesian estimation of alignments and evolutionary trees. Bioinformatics 24, 2403–2404.
Nylander, J.A., Ronquist, F., Huelsenbeck, J.P. and Nieves-Aldrey, J., 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53, 47–67.
Ogden, T.H. and Rosenberg, M.S., 2006. Multiple sequence alignment accuracy and phylogenetic inference. Syst. Biol. 55, 314–328.
Padial, J.M., Grant, T. and Frost, D.R., 2014. Molecular systematic and biogeography of terraranas (Anura: Brachycephaloidea): a new taxonomy and an assessment of the effects of alignment and optimality criteria. Zootaxa 3825, 1–132.
Paradis, E., Claude, J. and Strimmer, K., 2004. APE: analysis of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.
Peloso, P.L.V., Frost, D.R., Richards, S.J., Rodrigues, M.T., Donellan, S., Matsiu, M., Raxworthy, C.J., Biju, S.D., Moriaty-Lemon, E., Lemmon, A.R. and Wheeler, W.C., 2016. The impact of anchored phylogenomics and taxon sampling on phylogenetic inference in narrow-mouthed frogs (Anura, Microhylidae). Cladistics 32, 113–140.
Pol, D. and Siddall, M.E., 2001. Biases in maximum likelihood and parsimony: a simulation approach to the 10-taxon case. Cladistics 17, 266–281.
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. Proc. R. Soc. B. 284, 20162290.
Puttick, M.N., O´Reilly, J.E., Pisani, D. & Donoghue, P. C. J., 2019. Probabilistic methods outperform parsimony in the phylogenetic analysis of data simulated without a probabilistic model. Paleontology 62, 1–17.
Pyron, R.A. and Wiens, J.J., 2011. A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of advanced frogs, salamanders, and caecilians. Mol. Phylogenet. Evol. 61, 543–583.
Ramírez, M.J., 2005. Resampling measures of group support: a reply to Grant and Kluge. Cladistics 21, 83–89.
Redelings, B.D. and Suchard, M.A., 2005. Joint Bayesian estimation of alignment and phylogeny. Syst. Biol. 54, 401–418.
Rindal, E. and Brower, A.V.Z., 2010. Do model-based phylogenetic analyses perform better than parsimony? A test with empirical data. Cladistics 27, 331–334.
Robinson, D.F. and Foulds, L.R., 1981. Comparison of phylogenetic trees. Math. Biosci. 53, 131–147.
Sánchez-Pacheco, S.J., Torres-Carvajal, O., Aguirre-Peñafiel, V., Nunes, P.M., Verrastro, L., Rivas, G.A., Rodrigues, M.T., Grant, T. and Murphy, R.W., 2018. Phylogeny of *Riama* (Squamata: Gymnophthalmidae), impact of phenotypic evidence on molecular datasets, and the origin of the Sierra Nevada de Santa Marta endemic fauna. Cladistics 34, 260–291.
Sanderson, M.J., McMahon, M.M. and Steel, M., 2011. Terraces in phylogenetic tree space. Science 333, 448–450.
Schmid, M., Steinlein, C., Bogart, J.P., Feichtinger, W., Haaf, T., Nanda, I., del Pino, E.M., Duellman, W.E. and Hedges, S.B., 2012. The hemiphractid frogs – phylogeny, embryology, life history, and cytogenetics. Cytogenet. Genome Res. 138, 2–4.
Sereno, P.C., 2007. Logical basis for morphological characters in phylogenetics. Cladistics 23, 565–587.
Siddall, M.E., 1998. Success of parsimony in the four-taxon case: long-branch repulsion by likelihood in the Farris zone. Cladistics 14, 209–220.
Siddall, M.E. and Kluge, A.G., 1997. Probabilism and phylogenetic inference. Cladistics 13, 313–336.
Siddall, M.E. and Kluge, A.G., 1999. Letter to the editor. Cladistics 15, 439–440.
Siddall, M.E. and Whiting, M.F., 1999. Long-branch abstractions. Cladistics 15, 9–24.
Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. and Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87, 651–701.
Simmons, M.P., 2012. Misleading results of likelihood-based phylogenetic analysis in the presence of missing data. Cladistics 28, 208–222.
Simmons, M.P., 2014. A confounding effect of missing data on character conflict in maximum likelihood and Bayesian MCMC phylogenetic analyses. Mol. Phylogenet. Evol. 80, 267–280.
Simmons, M.P. and Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analysis. Syst. Biol. 49, 369–381.
Simmons, M.P., Ochoterena, H. & Carr, T.G., 2001. Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analysis. Syst. Biol. 50, 454–462.
Simmons, M.P., Müller, K. and Norton, A.P., 2007. The relative performance of indel coding methods in simulations. Mol. Phylogenet. Evol. 44, 724–740.
Simmons, M.P. and Goloboff, P.A., 2013. An artifact caused by undersampling optimal trees in supermatrix analysis of locally sampled characters. Mol. Phylogenet. Evol. 69, 265–275.
Simmons, M.P. and Goloboff, P.A., 2014. Dubious resolution and support from published sparse supermatrices: the importance of through tree searches. Mol. Phyl. Evol. 78, 334–348.
Simmons, M.P. and Kessenich, J., 2019. Divergence and support among slightly suboptimal likelihood gene trees. Cladistics 36, 322–340. https://doi.org/10.1111/cla.12404.
Smith, M.R., 2019. Bayesian and parsimony approaches reconstruct informative trees from simulated morphological datasets. Biol. Lett. 15, 20180632.
Suchard, M.A. and Redelings, B.D., 2006. Bali-Phy: simultaneous Bayesian inference of alignment and phylogeny. Bioinformatics 22, 2047–2048.

Sukumaran, J. and Holder, M.T., 2010a. DendroPy: a python library for phylogenetic computing. Bioinformatics 26, 1569–1571.
Sukumaran, J. and Holder, M.T., 2010b. SumTrees: phylogenetic tree sumarization. 4.3.0. Available at https://github.com/jeetsukumaran/DendroPy.
Swofford, D.L., Wadell, P., Huelsenbeck, J.P., Foster, P., Lewis, P. and Rogers, J., 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Syst. Biol. 50, 525–539.
Vaidya, G., Lohman, D.J. and Meier, R., 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27, 171–180.
Varón, A., Vinh, L.S. & Wheeler, W.C., 2010. POY version 4: phylogenetic analysis using dynamic homologies. Cladistics 26, 72–85.
Walker, M., Gasparini, J.L. and Haddad, C.F.B., 2016. A new polymorphic species of egg-brooding frog of the genus *Fritziana* from southeastern Brazil (Anura:Hemiphractidae). Salamandra 52, 221–229.
Walker, M., Lyra, M.L. and Haddad, C.F., 2018a. Phylogenetic relationships and cryptic species diversity in the Brazilian egg-brooding tree frog, genus *Fritziana* Mello-Leitão 1937 (Anura: Hemiphractidae). Mol. Phylogenet. Evol. 123, 59–72.
Ward, P.S., Brady, S.G., Fisher, B.L. and Schultz, T.R., 2010. Phylogeny and biogeography of dolichoderine ants: effects of data partitioning and relict taxa on historical inference. Syst. Biol. 59, 342–362.
Westesson, O., Barquist, L. and Holmes, I., 2012. HandAlign: Bayesian multiple sequence alignment, phylogeny and ancestral reconstruction. Bioinformatics 28, 1170–1171.
Wheeler, W.C., 1996. Optimization alignment, the end of multiple sequence alignment in phylogenetics? Cladistics 12, 1–10.
Wheeler, W.C., 2003a. Implied alignment, a synapomorphy-based multiple sequence alignment method and its use in cladogram search. Cladistics 19, 261–268.
Wheeler, W.C., 2003b. Iterative pass optimization of sequence data. Cladistics 19, 254–260.
Wheeler, W.C., 2012. Systematics, A Course of Lectures. Wiley-Blackwell, Hoboken, NJ.
Wheeler, W.C., Arango, C.P., Grant, T., Janies, D., Varón, A., Aagesen, L., Faivovich, J., D'Haese, C., Smith, W.L. and Giribet, G., 2006. Dynamic Homology and Phylogenetic Systematics, A Unified Approach Using POY. American Museum of Natural History, New York, NY.
Wheeler, W.C., Lucaroni, N., Hong, L., Crowley, L.M. & Varón, A., 2015. POY version 5: phylogenetic analysis using dynamic homologies under multiple optimality criteria. Cladistics 31, 189–196.
Wiens, J.J. and Servedio, M.R., 1998. Phylogenetic analysis and intraespecific variation: performance of parsimony, likelihood, and distance methods. Syst. Biol. 47, 228–253.
Wong, K.M., Suchard, M.A. and Huelsenbeck, J.P., 2008. Alignment uncertainty and genomic analysis. Science 319, 473–476.
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, e109210.
Yang, Z., 1996a. Phylogenetic analysis using parsimony and likelihood methods. J. Mol. Evol. 42, 294–307.
Yang, Z., 1996b. Among-site rate variation and its impact on phylogenetic analysis. Trends Ecol. Evol. 11, 367–372.
Yang, Z., Goldman, N. and Friday, A., 1994. Comparison of models for nucleotide substitution used in Maximum-Likelihood phylogenetic estimation. Mol. Biol. Evol. 11, 316–324.
Yu, G., Rao, D., Yang, J. and Zhang, M., 2008. Phylogenetic relationships among Racophorinae (Racophoridae, Anura, Amphibia), with an emphasis on the Chinese species. Zool. J. Linn. Soc. 153, 733–749.
Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Results of the tree-alignment + parsimony (TAPa) analysis: strict consensus of the 393 most parsimonious trees of 84 885 steps inferred from nuclear and mitochondrial DNA sequences and phenotypic data.

Fig. S2. Results of the similarity-alignment + parsimony, with gaps as fifth state (SAP5th), analysis: strict consensus of the 309 most parsimonious trees of 90 684 steps inferred from nuclear and mitochondrial DNA sequences and phenotypic data.

Fig. S3. Results of the similarity-alignment + parsimony, with gaps as binary characters (SAPg), analysis: strict consensus of the 1402 most parsimonious trees of 87 670 steps inferred from nuclear and mitochondrial DNA sequences and phenotypic data.

Fig. S4. Results of the similarity-alignment + parsimony, with gaps as missing data (SAPm), analysis: strict consensus of the 1132 most parsimonious trees of 83 798 steps inferred from nuclear and mitochondrial DNA sequences and phenotypic data.

Fig. S5. Results of the similarity-alignment + maximum likelihood, with gaps as binary characters (SALg), analysis: optimal tree (log likelihood = −368553.75) inferred from nuclear and mitochondrial DNA sequences and phenotypic data, numbers on branches are bootstrap percentages.

Fig. S6. Results of the similarity-alignment + maximum likelihood, with gaps as missing data (SALm), analysis: optimal tree (log likelihood = −348389.07) inferred from nuclear and mitochondrial DNA sequences and phenotypic data, numbers on branches are Bootstrap percentages.

Fig. S7. Comparisons, within each genus of Hemiphractidae, of the strict consensuses resulting from SAP5th (left) and TAPa (right) analysis.

Fig. S8. Comparisons, within each genus of Hemiphractidae, of the strict consensuses resulting from SAPg (left) and TAPa (right) analysis.

Fig. S9. Comparisons, within each genus of Hemiphractidae, of the strict consensuses resulting from SAPm (left) and TAPa (right) analysis.

Fig. S10. Comparisons, within each genus of Hemiphractidae, of the optimal tree and the strict

consensus resulting from SALg (left) and TAPa (right) analysis, respectively.

Fig. S11. Comparisons, within each genus of Hemiphractidae, of the optimal tree and the strict consensus resulting from SALm (left) and TAPa (right) analysis, respectively.

Fig. S12. Comparisons, within each genus of Hemiphractidae, of the strict consensuses resulting from SAP5th (left) and SAPg (right) analysis.

Fig. S13. Comparisons, within each genus of Hemiphractidae, of the strict consensuses resulting from SAP5th (left) and SAPm (right) analysis.

Fig. S14. Comparisons, within each genus of Hemiphractidae, of the optimal tree and the strict consensus resulting from SALg (left) and SAP5th (right) analysis, respectively.

Fig. S15. Comparisons, within each genus of Hemiphractidae, of the optimal tree and the strict consensus resulting from SALm (left) and SAP5th (right) analysis, respectively.

Fig. S16. Comparisons, within each genus of Hemiphractidae, of the strict consensuses resulting from SAPg (left) and SAPm (right) analysis.

Fig. S17. Comparisons, within each genus of Hemiphractidae, of the optimal tree and the strict consensus resulting from SALg (left) and SAPg (right) analysis, respectively.

Fig. S18. Comparisons, within each genus of Hemiphractidae, of the optimal tree and the strict consensus resulting from SALm (left) and SAPg (right) analysis, respectively.

Fig. S19. Comparisons, within each genus of Hemiphractidae, of the optimal tree and the strict consensus resulting from SALg (left) and SAPm (right) analysis, respectively.

Fig. S20. Comparisons, within each genus of Hemiphractidae, of the optimal tree and the strict consensus resulting from SALm (left) and SAPm (right) analysis, respectively.

Fig. S21. Comparisons, within each genus of Hemiphractidae, of the optimal trees resulting from SALg (left) and SALm (right) analysis.

Fig. S22. Tree-alignment + parsimony (TAPa) analysis on a 36 terminals dataset, showing ingroup relationships: optimal tree of 12 236 steps inferred from nuclear and mitochondrial DNA sequences and phenotypic data.

Fig. S23. Similarity-alignment + parsimony, gaps as fifth state (SAP5th) analysis on a 36 terminals dataset, showing ingroup relationships: optimal tree of 13 024 steps inferred from nuclear and mitochondrial DNA sequences and phenotypic data.

Fig. S24. Similarity-alignment + parsimony, gaps as binary characters (SAPg) analysis on a 36 terminals dataset, showing ingroup relationships: one of eight optimal trees of 12 340 steps inferred from nuclear and mitochondrial DNA sequences and phenotypic data.

Fig. S25. Similarity-alignment + parsimony, gaps as missing data (SAPm) analysis on a 36 terminals dataset, showing ingroup relationships: one of two optimal trees of 12 006 steps inferred from nuclear and mitochondrial DNA sequences and phenotypic data.

Fig. S26. Similarity-alignment + maximum likelihood, gaps as binary characters (SALg) analysis on a 36 terminals dataset, showing ingroup relationships: optimal tree (log likelihood = −68555.403517) inferred from nuclear and mitochondrial DNA sequences and phenotypic data.

Fig. S27. Similarity-alignment + maximum likelihood, gaps as missing data (SALm) analysis on a 36 terminals dataset, showing ingroup relationships: optimal tree (log likelihood = −67203.063422) inferred from nuclear and mitochondrial DNA sequences and phenotypic data.

Fig. S28. Tree-alignment + parsimony (TAPa) analysis on a 36 terminals dataset, showing ingroup relationships: optimal tree of 11 947 steps inferred from nuclear and mitochondrial DNA sequences.

Fig. S29. Similarity-alignment + parsimony, gaps as fifth state (SAP5th) analysis on a 36 terminals dataset, showing ingroup relationships: one of six optimal trees of 12 725 steps inferred from nuclear and mitochondrial DNA sequences.

Fig. S30. Similarity-alignment + parsimony, gaps as binary characters (SAPg) analysis on a 36 terminals dataset, showing ingroup relationships: one of eight optimal trees of 12 050 steps inferred from nuclear and mitochondrial DNA sequences.

Fig. S31. Similarity-alignment + parsimony, gaps as missing data (SAPm) analysis on a 36 terminals dataset, showing ingroup relationships: one of nine optimal trees of 11 718 steps inferred from nuclear and mitochondrial DNA sequences.

Fig. S32. Similarity-alignment + maximum likelihood, gaps as binary characters (SALg) analysis on a 36 terminals dataset, showing ingroup relationships: optimal tree (log likelihood = −67563.887800) inferred from nuclear and mitochondrial DNA sequences.

Fig. S33. Similarity-alignment + maximum likelihood, gaps as missing data (SALm) analysis on a 36 terminals dataset, showing ingroup relationships: optimal tree (log likelihood = −66210.248945) inferred from nuclear and mitochondrial DNA sequences.

Table S1. Updated identification of ingroup and outgroup samples used in previous studies.

Table S2. Partition schemes evaluated in PartitionFinder and resulting scores for the corrected Akaike Information Criterion (AICc).

Table S3. Mean Robinson-Foulds local distance values calculated among optimal maximum likelihood

trees and a subsample of optimal trees from parsimony analysis.

Table S4. Mean number of SPR moves calculated from pairwise SPR distance comparisons among optimal maximum likelihood trees and a subsample of optimal trees from parsimony analyses.

Table S5. Terminals ordered from minimum to maximum average match split distances (MSD) among TAPa optimal trees.

Table S6. Terminals ordered from minimum to maximum average match split distances (MSD) among SAP5th optimal trees.

Table S7. Terminals ordered from minimum to maximum average match split distances (MSD) among SAPg optimal trees.

Table S8. Terminals ordered from minimum to maximum average match split distances (MSD) among SAPm optimal trees.

Table S9. Number of clades shared by optimal trees of analyses performed using combined and molecular-only datasets.

Appendix S1. Terminals, voucher codes (for ingroup taxa), and GenBank accession numbers of DNA sequences used in this study.

Appendix S2. Specimens, voucher/field number codes, localities, and list of sequences generated in this study.

Appendix S3. List of phenotypic synapomorphies common to the 393 optimal trees of tree-alignment + parsimony (TAPa) analysis.

Data S1. DNA sequence fragments, of 20 mitochondrial and nuclear loci, analyzed jointly with a phenotypic matrix in the tree-alignment + parsimony (TAPa) analysis.

Data S2. Phenotypic matrix, of 51 characters, analyzed jointly with DNA sequence fragments in the tree-alignment + parsimony (TAPa) analysis.

Data S3. Implied alignment of DNA sequences resulting from the tree-alignment + parsimony (TAPa) analysis.

Data S4. Total-evidence data matrix analyzed in the similarity-alignment + parsimony, with gaps as fifth state (SAP5th), analysis.

Data S5. Total-evidence data matrix analyzed in the similarity-alignment + parsimony, with gaps as binary characters (SAPg), analysis.

Data S6. Total-evidence data matrix analyzed in the similarity-alignment + parsimony, with gaps as missing data (SAPm), analysis.

Data S7. Total-evidence data matrix analyzed in the similarity-alignment + maximum Likelihood, with gaps as binary characters (SALg), analysis.

Data S8. Total-evidence data matrix analyzed in the similarity-alignment + maximum likelihood, with gaps as missing data (SALm), analysis.

Appendix 1

Genus *Gastrotheca* Fitzinger, 1843

Gastrotheca fissipes species group

Equivalent to: Subgenus *Eotheca* Duellman, 2015

Content: *Gastrotheca fissipes*, *G. flamma*, *G. megacephala*, *G. prasina*, *G. pulchra*, and *G. recava*.

Remarks: This clade includes the undescribed species *Gastrotheca* sp. L.

Gastrotheca longipes species group

Equivalent to: Subgenus *Amphignathodon* Boulenger, 1882 *sensu* Duellman (2015).

Content: *Gastrotheca andaquiensis*, *G. angustifrons*, *G. antomia*, *G. bufona*, *G. cornuta*, *G. dendronastes*, *G. guentheri*, *G. helenae*, and *G. longipes*.

Remarks: Equivalent in content to the *Gastrotheca longipes* species group of Castroviejo-Fisher et al. (2015) except for the exclusion in this work of *G. walkeri* and *G. williamsoni*. Our analyses did not include *G. andaquiensis*, *G. angustifrons*, *G. antomia* and *G. bufona*. They remain tentatively assigned to this clade, as in Castroviejo-Fisher et al. (2015), based on similarity with the other species of the group.

Gastrotheca marsupiata species group

Equivalent to: Subgenus *Gastrotheca* Fitzinger, 1843 *sensu* this work.

Synonyms of Subgenus *Gastrotheca*: Subgenus *Duellmania* Dubois, 1987, **new synonym**; Subgenus *Edaphoteca* Duellman, 2015, **new synonym**.

Content: *Gastrotheca abdita*, *G. aguaruna*, *G. antoniiochoai*, *G. aratia*, *G. argenteovirens*, *G. atympana*, *G. aureomaculata*, *G. caeruleomaculata*, *G. carinaceps*, *G. christiani*, *G. chrysosticta*, *G. cuencana*, *G. dissimilis*, *G. dunni*, *G. dysprosita*, *G. elicioi*, *G. espeletia*, *G. excubitor*, *G. galeata*, *G. gracilis*, *G. griswoldi*, *G. lateonota*, *G. litonedis*, *G. lojana*, *G. marsupiata*, *G. monticola*, *G. nebulanastes*, *G. nicefori*, *G. ochoai*, *G. oresbios*, *G. orophylax*, *G. ossilaginis*, *G. pacchamama*, *G. pachachacae*, *G. peruana*, *G. phalarosa*, *G. phelloderma*, *G. piperata*, *G. plumbea*, *G. pseustes*, *G. psychrophila*, *G. rebeccae*, *G. riobambae*, *G. ruizi*, *G. spectabilis*, *G. splendens*, *G. stictopleura*, *G. testudinea*, *G. trachyceps*, *G. turnerorum*, *G. yacuri*, and *G. zeugocystis*.

Remarks: The only difference in content with the *G. marsupiata* species group of Castroviejo-Fisher et al. (2015) is the exclusion in this work of *G. ovifera*. The subgenera *Gastrotheca* and *Duellmania* are non-monophyletic. Although *Edaphoteca* is monotypic, the preservation of this subgenus would require further nomenclatural changes to *Gastrotheca* and *Duellmania* in all analyses. The simplest solution (i.e., fewer nomenclatural changes) is to consider *Duellmania* and *Edaphoteca* as synonyms of the subgenus *Gastrotheca*. This also renders the *G. marsupiata* species group equivalent to the subgenus *Gastrotheca*.

The following species were not included in our analyses: *G. abdita*, *G. caeruleomaculata*, *G. carinaceps*, *G. cuencana*, *G. dysprosita*, *G. lateonota*, *G. ossilaginis*, *G. pacchamama*, *G. peruana*, *G. piperata*, and *G. splendens*. These species remain assigned to this clade because of its Andean distribution and morphological similarity to other species of the group.

Gastrotheca microdiscus species group

Equivalent to: Subgenus *Australotheca* Duellman, 2015

Content: *G. albolineata*, *G. ernestoi*, *G. fulvorufa*, and *G. microdiscus*.

Gastrotheca ovifera (Lichtenstein and Weinland, 1854)

Equivalent to: Subgenus *Opisthodelphys* Günther, 1859 *sensu* Duellman (2015).

Remarks: Castroviejo-Fisher et al. (2015) and Duellman (2015) recovered *G. ovifera* as sister of the large Andean radiation of *Gastrotheca*. Our preferred hypothesis places *G. ovifera* as sister of a monophyletic *G. fissipes* group. We leave *G. ovifera* unassigned to a species group within *Gastrotheca*.

Gastrotheca walkeri species group

Equivalent to: Subgenus *Cryptotheca* Duellman, 2015

Content: *G. walkeri* and *G. williamsoni*.

Remarks: In the *G. longipes* species group of Castroviejo-Fisher et al. (2015). Although this clade is the sister group of the *G. longipes* species group and hence its monophyly as in Castroviejo-Fisher et al. (2015) is not compromised, we prefer to recognize this clade as its own species group to highlight the presence of a brood pouch that laterally penetrates the body wall into the coelomic cavity (Duellman, 1980, 2015). *Gastrotheca williamsoni* was not included in our analyses, but based on its overall similarity with *G. walkeri* (Duellman, 2015), we assume a close evolutionary relationship between both species and extend any supraspecific taxonomic rearrange on *G. walkeri* to *G. williamsoni*.