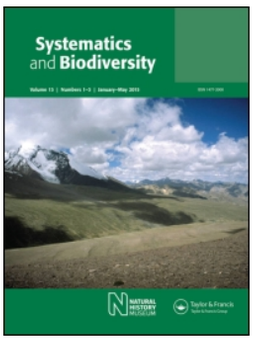


Systematics and Biodiversity

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tsab20

Phylogeny of terraranan frogs based on 2,665 loci and impacts of missing data on phylogenomic analyses

Lucas S. Barrientos, Jeffrey W. Streicher, Elizabeth C. Miller, Marcio R. Pie, John J. Wiens & Andrew J. Crawford

To cite this article: Lucas S. Barrientos, Jeffrey W. Streicher, Elizabeth C. Miller, Marcio R. Pie, John J. Wiens & Andrew J. Crawford (2021): Phylogeny of terraranan frogs based on 2,665 loci and impacts of missing data on phylogenomic analyses, Systematics and Biodiversity, DOI: 10.1080/14772000.2021.1933249

To link to this article: https://doi.org/10.1080/14772000.2021.1933249

Systematics and Biodiversity (2021), **0(0)**: 1–16

Taylor & Francis
Taylor & Francis Group

Research Article

Phylogeny of terraranan frogs based on 2,665 loci and impacts of missing data on phylogenomic analyses

LUCAS S. BARRIENTOS[1], JEFFREY W. STREICHER[2,3], ELIZABETH C. MILLER[2,4], MARCIO R. PIE[5], JOHN J. WIENS[2] & ANDREW J. CRAWFORD[1]

[1]Department of Biological Sciences, Universidad de los Andes, Bogotá, 111711, Colombia
[2]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, 85721-0088, AZ, USA
[3]Department of Life Sciences, The Natural History Museum, South Kensington, London, SW7 5BD, England, UK
[4]School of Aquatic and Fishery Sciences, University of Washington, Seattle, 98195-5020, WA, USA
[5]Departamento de Zoologia, Universidade Federal do Paraná, Curitiba, 81531-980, Paraná, Brazil

Terraranae is a large clade of New World direct-developing frogs that includes 3–5 families and >1,100 described species (~15% of all named frog species). The relationships among major groups of terraranan frogs have been highly contentious, including conflicts among four recent phylogenomic studies utilizing 95, 220, 389, and 2,214 nuclear loci, respectively. In this paper, we re-evaluate relationships within Terraranae using a novel genomic dataset for 16 ingroup species representing most terraranan families and subfamilies. The preferred data matrix consisted of 2,665 nuclear loci from ultraconserved elements (UCEs), with a total of 743,419 aligned base pairs and 57% missing data. Concatenated likelihood analyses and coalescent-based species-tree analyses recovered strong statistical support for the following relationships among terraranan families: (Brachycephalidae, (Eleutherodactylidae, (Craugastoridae + 'Strabomantidae'))). This position for Brachycephalidae agrees with two previous phylogenomic studies but conflicts with two others. Our results infer *Strabomantis* (of the Strabomantidae) to be with (or within) *Craugastor* (Craugastoridae), rather than with other strabomantid genera. This renders Strabomantidae paraphyletic with respect to Craugastoridae. Our results also suggest that Pristimantinae is paraphyletic with respect to Holoadeninae. We also find that using matrices with less missing data (and concomitantly fewer UCE loci) generally resulted in trees with lower mean branch support and problematic phylogenies (e.g. non-monophyly of terraranans). Overall, our results help resolve controversial relationships within one of the largest clades of frogs, with a dataset including ~7 times more loci than those used in previous studies focused on this clade.

Key words: amphibians, Anura, missing data, phylogenomics, Terraranae, ultraconserved elements

Introduction

Terraranae (*sensu* Heinicke et al., 2018) is one of the most species-rich clades of frogs. It contains approximately 1,200 described species, or roughly 15% of described extant anuran species (AmphibiaWeb, 2020). The monophyly of terraranans has been supported in numerous large-scale molecular analyses (Feng et al., 2017; Frost et al., 2006; Hime et al., 2021; Pyron, 2014; Pyron & Wiens, 2011; Streicher et al., 2018). Terraranae is also characterized by several soft anatomical characters (Taboada et al., 2013) and by direct development. Direct development involves the evolutionary loss of the larval stage, such that four-legged hatchlings emerge from fully terrestrial eggs (Duellman & Trueb, 1994).

Terraranans collectively occur from the southern United States to southern Brazil, including the West Indies (Gonzalez-Voyer et al., 2011; Hedges et al., 2008). They are found in a variety of habitats, from deserts to rainforests and from islands to high-elevation páramo and puna (Gonzalez-Voyer et al., 2011; Hedges et al., 2008). In the Neotropics, terraranans have been estimated to make up (on average) >40% of all frog species in local communities, with especially high richness in mesic habitats of Middle America, the Andes of South America, and on Caribbean islands (Pinto-Sánchez et al., 2014).

Several key aspects of the phylogeny and taxonomy of terraranans have been relatively unstable in recent

Correspondence to: Jeffrey W. Streicher. E-mail: j.streicher@nhm.ac.uk

ISSN 1477-2000 print / 1478-0933 online

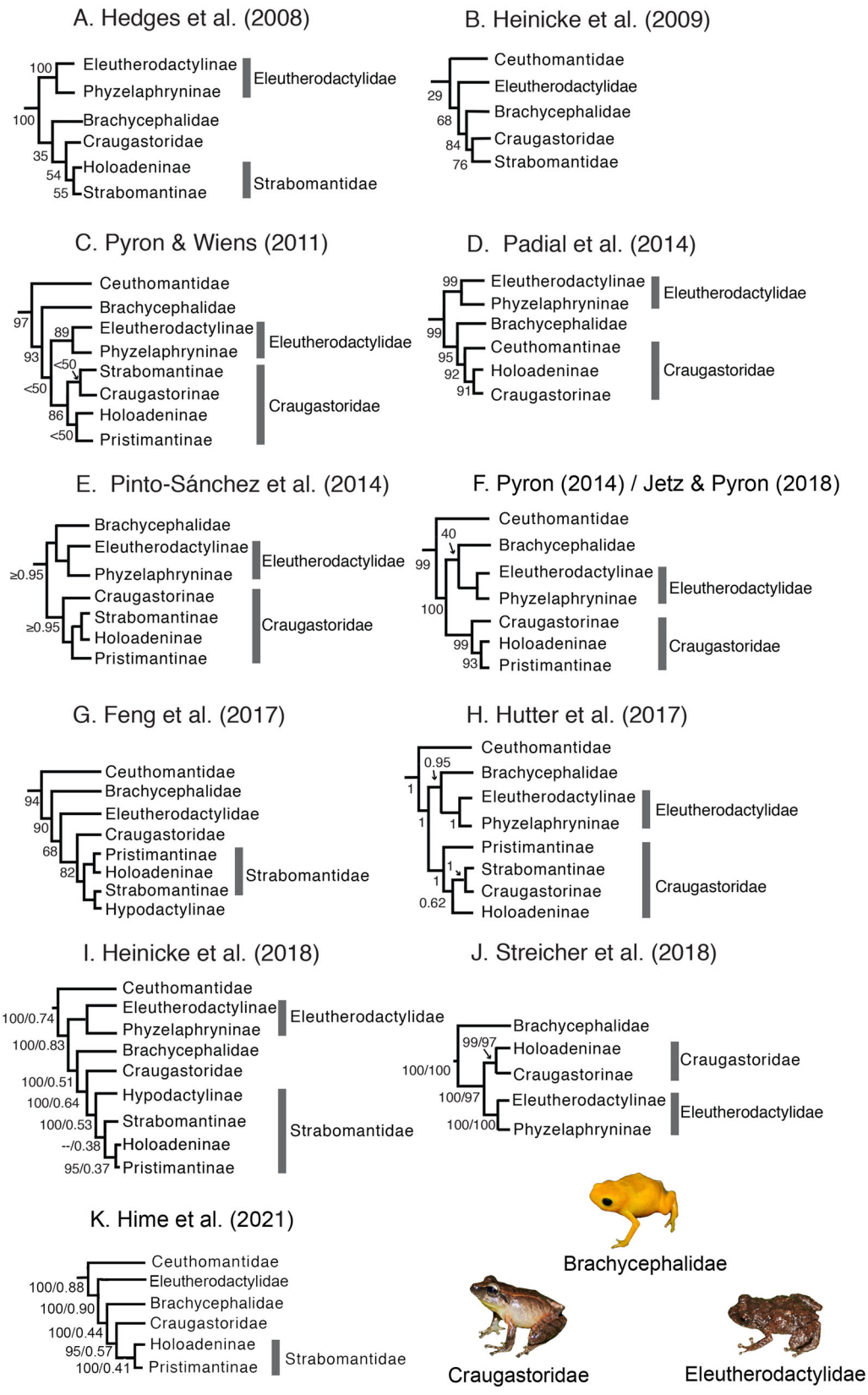
https://dx.doi.org/10.1080/14772000.2021.1933249

Published online 21 Jun 2021

molecular studies (Fig. 1). Prior to 2006, traditional taxonomy placed all terraranans in the tribe Eleutherodactylini within the family Leptodactylidae (Lynch, 1971). Based on a parsimony analysis of nuclear and mitochondrial DNA data, Frost et al. (2006) subdivided Leptodactylidae and considered all terraranans to belong to the family Brachycephalidae. Hedges et al. (2008) proposed a new phylogeny for this group of frogs. They gave this group the unranked name of Terrarana (later amended to Terraranae; Dubois, 2009). They divided Terraranae into four families: Brachycephalidae, Craugastoridae, Eleutherodactylidae, and Strabomantidae (Fig. 1A). They found weak support for relationships among families, with Eleutherodactylidae as sister to all other families and Brachycephalidae as sister to Craugastoridae + Strabomantidae. Subsequently, Heinicke et al. (2009) added a fifth family, Ceuthomantidae, sister to all other families (Fig. 1B). Pyron and Wiens (2011) performed a large-scale maximum likelihood analysis of published data for 12 concatenated nuclear and mitochondrial genes for many species, which positioned Strabomantidae within Craugastoridae (Fig. 1C). They also found Ceuthomantidae was sister to other terraranans, and Brachycephalidae was sister to Eleutherodactylidae + Craugastoridae. Padial et al. (2014) added two more nuclear and three more mitochondrial genes to the previous data sets, and analysed relationships using parsimony with dynamic homology (Wheeler, 1999, 2001). This analysis (Fig. 1D) inferred Eleutherodactylidae as the sister taxon to Brachycephalidae + Craugastoridae (including Strabomantidae), and recovered Ceuthomantidae within Craugastoridae (including Strabomantidae). Pinto-Sánchez et al. (2014) combined data from two previous studies (Pinto-Sánchez et al., 2012; Pyron & Wiens, 2011) to analyse relationships among terraranans. They found weak support for Brachycephalidae as the sister taxon to Eleutherodactylidae (Fig. 1E), unlike Pyron and Wiens (2011). Pyron (2014) added published data to the matrix of Pyron and Wiens (2011) and also found weak support for a closer relationship of Eleutherodactylidae to Brachycephalidae than Craugastoridae (including Strabomantidae; Fig. 1F). This same topology was also found in a later study based on further modifications to this dataset (Jetz & Pyron, 2018). Feng et al. (2017) analysed 95 nuclear loci to infer relationships among anurans (Fig. 1G) and found support for the relationships: (Ceuthomantidae (Eleutherodactylidae (Craugastoridae + Strabomantidae))). Hutter et al. (2017) performed a Bayesian analysis of 13 nuclear and seven mitochondrial genes among 158 hyloid genera and found strong support (Fig. 1H) for the relationships: (Ceuthomantidae (Brachycephalidae (Craugastoridae [including Strabomantidae] + Eleutherodactylidae))). Heinicke et al. (2018) analysed 389 nuclear loci and inferred Eleutherodactylidae as the sister taxon to a clade including Brachycephalidae and Craugastoridae + Strabomantidae (Fig. 1I), again with Ceuthomantidae as sister to all other terraranans. Streicher et al. (2018) analysed 2,214 nuclear loci for hyloid frogs (including five terraranan species) and found support for Brachycephalidae as the sister taxon of Craugastoridae (including Strabomantidae) + Eleutherodactylidae (Fig. 1J). Hime et al. (2021) analysed 220 loci, a subset of the loci used by Heinicke et al. (2018), for seven terraranan species. They found support (Fig. 1K) for the same family-level relationships reported in Heinicke et al. (2018): (Ceuthomantidae, (Eleutherodactylidae, (Brachycephalidae, Craugastoridae [including Strabomantidae])).

The results from these studies can be summarized as follows. There are three main competing hypotheses for relationships among terraranan families: (1) Eleutherodactylidae as the sister taxon to Brachycephalidae + Craugastoridae + Strabomantidae (Hedges et al., 2008; Heinicke et al., 2009, 2018; Hime et al., 2021; Padial et al., 2014); (2) Brachycephalidae as the sister taxon to Eleutherodactylidae + (Craugastoridae + Strabomantidae) (Feng et al., 2017; Hutter et al., 2017; Pyron & Wiens, 2011; Streicher et al., 2018); and (3) Brachycephalidae + Eleutherodactylidae as sister to Craugastoridae + Strabomantidae (Jetz & Pyron, 2018; Pinto-Sánchez et al., 2014; Pyron, 2014). Another difference among studies is that the family Ceuthomantidae was revovered as sister to all other terraranan families by Heinicke et al. (2009), and most subsequent studies (Acosta-Galvis et al., 2018; Feng et al., 2017; Hime et al., 2021; Motta et al., 2021; Pyron, 2014; Pyron & Wiens, 2011). However, Padial et al. (2014) inferred Ceuthomantidae (i.e. the genus *Ceuthomantis*) as being nested within Craugastoridae.

What might explain these disagreements? Many previous molecular phylogenetic studies were based on fewer than 10 mitochondrial and 14 nuclear loci, and often had relatively weak support for the key, conflicting relationships. But even the four phylogenomic studies that included multiple terraranan samples, including Feng et al. (2017), Heinicke et al. (2018), Streicher et al. (2018), and Hime et al. (2021), showed conflicting relationships among these families (Fig. 1), despite analysing 97, 389, 2214, and 220, loci, respectively (Table 1). Most studies differed in their sampling of genes and taxa (Table 1), which might have contributed to the conflicting results. Another potential cause of conflict


A. Hedges et al. (2008)
100
Eleutherodactylinae
Phyzelaphryninae
Eleutherodactylidae
100
Brachycephalidae
Craugastoridae
35
54
Holoadeninae
55
Strabomantinae
Strabomantidae
B. Heinicke et al. (2009)
Ceuthomantidae
29
Eleutherodactylidae
68
Brachycephalidae
84
Craugastoridae
76
Strabomantidae
C. Pyron & Wiens (2011)
Ceuthomantidae
Brachycephalidae
97
89
Eleutherodactylinae
Phyzelaphryninae
Eleutherodactylidae
93
<50
Strabomantinae
<50
Craugastorinae
Craugastoridae
86
Holoadeninae
<50
Pristimantinae
D. Padial et al. (2014)
99
Eleutherodactylinae
Phyzelaphryninae
Eleutherodactylidae
99
Brachycephalidae
95
Ceuthomantinae
92
Holoadeninae
Craugastoridae
91
Craugastorinae
E. Pinto-Sánchez et al. (2014)
Brachycephalidae
Eleutherodactylinae
Eleutherodactylidae
≥0.95
Phyzelaphryninae
Craugastorinae
≥0.95
Strabomantinae
Craugastoridae
Holoadeninae
Pristimantinae
F. Pyron (2014) / Jetz & Pyron (2018)
Ceuthomantidae
40
Brachycephalidae
99
Eleutherodactylinae
Eleutherodactylidae
Phyzelaphryninae
100
Craugastorinae
99
Holoadeninae
Craugastoridae
93
Pristimantinae
G. Feng et al. (2017)
Ceuthomantidae
Brachycephalidae
94
Eleutherodactylidae
90
Craugastoridae
68
Pristimantinae
82
Holoadeninae
Strabomantidae
Strabomantinae
Hypodactylinae
H. Hutter et al. (2017)
Ceuthomantidae
0.95
Brachycephalidae
1
Eleutherodactylinae
Eleutherodactylidae
1
1
Phyzelaphryninae
Pristimantinae
1
Strabomantinae
1
Craugastorinae
Craugastoridae
0.62
Holoadeninae
I. Heinicke et al. (2018)
Ceuthomantidae
Eleutherodactylinae
Eleutherodactylidae
100/0.74
Phyzelaphryninae
100/0.83
Brachycephalidae
Craugastoridae
100/0.51
100/0.64
Hypodactylinae
100/0.53
Strabomantinae
Strabomantidae
--/0.38
Holoadeninae
95/0.37
Pristimantinae
J. Streicher et al. (2018)
Brachycephalidae
99/97
Holoadeninae
Craugastoridae
100/100
Craugastorinae
100/97
Eleutherodactylinae
Eleutherodactylidae
100/100
Phyzelaphryninae
K. Hime et al. (2021)
Ceuthomantidae
100/0.88
Eleutherodactylidae
100/0.90
Brachycephalidae
Craugastoridae
100/0.44
95/0.57
Holoadeninae
100/0.41
Pristimantinae
Strabomantidae
Brachycephalidae
Craugastoridae
Eleutherodactylidae

among studies is the different inference methods used (Table 1), including parsimony (with a dynamic optimization criterion), maximum likelihood and Bayesian analysis of concatenated data, and coalescent-based species-tree analyses.

In recent years, a new class of molecular markers for phylogenomic studies has been developed based on ultraconserved genomic elements (UCEs; Bejerano et al., 2004). Because of their conserved nature, researchers are able to enrich and capture DNA sequences from thousands of nuclear loci, even from distantly related taxa (McCormack et al., 2012). UCEs have been used to address relationships within many vertebrate clades, including major groups of reptiles (Crawford et al., 2015) and among families of fishes (Faircloth et al., 2013), frogs (Pie et al., 2019; Streicher et al., 2018), lizards (Portik & Wiens, 2021; Streicher et al., 2016, 2018; Streicher & Wiens, 2017), and snakes (Streicher & Wiens, 2016). UCEs have also been applied to phylogenetic questions at lower taxonomic levels, such as among bird genera (Meiklejohn et al., 2016) and among species of birds and frogs (Alexander et al., 2017; McCormack et al., 2016; Smith et al., 2014).

A potential disadvantage of UCEs is that the data matrices generated may include considerable amounts of missing data. Specifically, including more UCE loci in a given dataset typically requires increasing levels of missing data because data are not available for all taxa for all loci. Using UCE data from iguanian lizards, Streicher et al. (2016) examined the impacts of including different levels of missing data and different numbers of loci on the performance of concatenated and species-tree analyses. Those authors found that the recovery of well-established clades (and overall branch-support) was maximized by including loci with an intermediate level of missing data (i.e., loci that lacked data for up to 50% of the taxa). However, these authors only examined data matrices containing loci with up to $20-60\%$ of taxa lacking data. Here, we use data from thousands of UCE loci to address higher-level relationships within Terraranae using both concatenated and species-tree analyses. We also evaluate the effect of missing data on phylogenetic inference in UCEs across a wider range of sampling strategies (allowing loci with up to 90% of taxa lacking data).

Materials and methods

Taxon sampling

We follow the standard taxonomy of AmphibiaWeb (2020) here and throughout the paper. Tissue samples were from natural history collections (see below). The taxon sampling included 16 ingroup species of Terraranae, including four of the five terraranan families (Table 2). We also included five outgroup species. Data for four of the ingroup species and all five outgroup species were taken from Streicher et al. (2018). From the family Eleutherodactylidae, we included samples from both subfamilies (Eleutherodactylinae and Phyzelaphryninae) and all four genera (*Eleutherodactylus, Diasporus, Phyzelaphryne,* and *Adelophryne*). Brachycephalidae contains only two genera, *Brachycephalus* and *Ischnocnema*, and we included the former. From Craugastoridae, the subfamily Craugastorinae was represented by one of the two genera (*Craugastor*). Within *Craugastor* our samples represented the subgenera *Campbellius, Craugastor,* and *Hylactophryne*. From Strabomantidae we sampled four of the 18 genera, including *Barycholos* from the subfamily

Fig. 1. Summary of hypotheses of higher-level phylogenetic relationships among terraranan frogs. The taxonomy used in each tree follows the taxonomy used in that study. Numbers adjacent to internal branches represent bootstrap support unless otherwise indicated. All branch lengths are arbitrary. Taxon sampling for each study is summarized in Table 1. (A) Hedges et al. (2008) concatenated maximum likelihood (ML) analysis of two mitochondrial genes and two nuclear genes. (B) Heinicke et al. (2009) concatenated ML analysis of six mitochondrial and 11 nuclear genes. (C) Pyron and Wiens (2011) concatenated ML analysis of three mitochondrial and nine nuclear genes. (D) Padial et al. (2014) parsimony analysis of nine mitochondrial and 12 nuclear genes; numbers adjacent to internal branches represent jackknife support. (E) Pinto-Sánchez et al. (2014) concatenated Bayesian analysis of three mitochondrial and nine nuclear genes; internal branches without support represent Bayesian posterior probabilities <0.95. (F) Pyron (2014) concatenated ML analysis of three mitochondrial and nine nuclear genes. This is the same topology reported in Jetz and Pyron (2018). Support values are from Pyron (2014). (G) Feng et al. (2017) concatenated ML analysis and coalescent-based species tree (ASTRAL) of 95 nuclear protein-coding genes. (H) Hutter et al. (2017) concatenated Bayesian analysis of seven mitochondrial and 13 nuclear genes; numbers adjacent to internal branches represent Bayesian posterior probabilities. (I) Heinicke et al. (2018) concatenated ML analysis and coalescent-based species tree (ASTRAL) analysis of 389 nuclear protein-coding genes; numbers adjacent to internal branches represent ML bootstrap support before the slash, followed by local posterior probabilities for ASTRAL. (J) Streicher et al. (2018) concatenated ML analysis of 2,214 UCEs, including loci with up to 60% of taxa lacking data per UCE. Numbers next to each internal branch show support from concatenated analysis before the slash, followed by NJst bootstrap. (K) Hime et al. (2021) concatenated ML analysis and coalescent-based species tree (ASTRAL) analysis of 220 nuclear protein-coding genes; numbers adjacent to internal branches represent ML bootstrap support before the slash, followed by local posterior probabilities for ASTRAL.

Table 1. Summary of previous studies of terraranan relationships using molecular data. The abbreviation for the inference methods are: maximum parsimony (MP), maximum likelihood (ML), Bayesian inference (Bayes), parsimony under direct/dynamic optimization criterion (POY), and coalescent-based species-tree estimation using Accurate Species TRee Algorithm (ASTRAL) and NJst. Phylogenetic result summarizes inter-familial relationships among taxonomic families abbreviated as follows: B is Brachycephalidae, C is Craugastoridae, E is Eleutherodactylidae, and S is Strabomantidae.

Study	Terraranan taxa	Outgroup taxa	Mitochondrial genes	Nuclear genes	Inference method	Phylogenetic result
Hedges et al. (2008)	344	18	2	2	ML, Bayes	(E (B, C, S))
Heinicke et al. (2009)	42	4	6	11	MP, ML, Bayes	(E (B, C, S))
Pyron and Wiens (2011)	340	2533	3	9	ML	(B (E (C, S)))
Padial et al. (2014)	405	25	9	11	POY	(E (B, C, S))
Pyron (2014)	418	2892	3	9	ML	((B, E) (C, S))
Pinto-Sánchez et al. (2014)	363	7	3	9	Bayes	((B, E) (C, S))
Feng et al. (2017)	16	278	–	97	ML, ASTRAL	(B (E (C, S)))
Heinicke et al. (2018)	30	5	–	389	ML, ASTRAL	(E (B, C, S))
Streicher et al. (2018)	5	45	–	2214	ML, NJst, ASTRAL	(B (E (C, S)))
Hutter et al. (2017)	610*	1708*	7	13	Bayes	(B (E (C, S)))
Hime et al. (2021)	7	279	–	220	ML, ASTRAL	(E (B, C, S))
This study	16	5	–	2665	ML, NJst, ASTRAL	(B (E (C, S)))

*Includes well-supported but undescribed species.

Holoadeninae, and representatives of *Lynchius, Pristimantis*, and *Oreobates* from the subfamily Pristimantinae (Table 2). We lacked a sample of *Ceuthomantis*, which is usually considered a distinct family and the sister to all other terraranan families (Feng et al., 2017; Heinicke et al., 2009, 2018; Hime et al., 2021; Pyron, 2014; Pyron & Wiens, 2011). We also lacked *Hypodactylus* (now *Niceforonia,* see Acosta-Galvis et al., 2018), which has been considered a monogeneric subfamily of Strabomantidae (Heinicke et al., 2018). However, our taxon sampling is able to address a major debate among studies of terraranan phylogeny: the relationships among Brachycephalidae, Craugastoridae, Eleutherodactylidae, and Strabomantidae.

Many previous phylogenetic studies have demonstrated that Terraranae is nested within Hyloidea (Feng et al., 2017; Frost et al., 2006; Hime et al., 2021; Pyron, 2014; Pyron & Wiens, 2011; Streicher et al., 2018). However, relationships among hyloid families have been only weakly supported in most previous studies. For outgroups, we included representatives of five hyloid families (Centrolenidae, Dendrobatidae, Hemiphractidae, Hylidae, Leptodactylidae). Based on the well-supported tree of Streicher et al. (2018), the sister group to Terraranae includes Centrolenidae, Dendrobatidae, and Leptodactylidae. These (along with additional families) form the clade Commutabirana, and the sister group to Commutabirana includes Hemiphractidae and Hylidae (members of the clade Amazorana).

Samples were provided by the Círculo Herpetológico de Panamá (CH), Museo de Historia Natural C.J. Marinkelle at the Universidad de los Andes in Bogotá (ANDES), the Museum of Vertebrate Zoology at the University of California, Berkeley (MVZ), Amphibian and Reptile Diversity Research Center at the University of Texas at Arlington (UTA), Museum of Comparative Zoology at Harvard University (MCZ), and the Biodiversity Institute and Natural History Museum at the University of Kansas (KU).

DNA extraction, library preparation, and sequencing

Genomic DNA (gDNA) was extracted using DNeasy® Blood and Tissue kits (Qiagen) or using magnetic beads (Sera-Mag Speedbeads, Fisher Scientific). Samples were digested overnight in 20 µL proteinase K in 180 µL of lysis buffer. Genomic DNA was captured with ca. 360 µL magnetic beads, cleaned with two 700 µL washes of 70% EtOH, and eluted in 70 µL of 10 mM Tris (pH 8). After extraction, we quantified the amount of gDNA via fluorometry using double-stranded DNA high-sensitivity assay kits (Qubit, Life Technologies).

For capture and library preparation we followed the protocol of Faircloth et al. (2012) (available at http://ultraconserved.org), with the modifications used by Streicher et al. (2016). Template gDNA (∼150 ng) was fragmented by either physical shearing with a Bioruptor (Diagenode) using 6 cycles of high-speed agitation (with 30 seconds on and 90 seconds off), or by enzymatic digestion using NEBNext dsDNA Fragmentase (New England Biolabs) at 37 °C for 25 minutes. The post-hybridization PCR was conducted with NEB Phusion DNA polymerase and TruSeq primers, following Streicher et al. (2016). Enriched libraries were visualized for fragment-size distribution and abundance using a Bioanalyzer 7500 (Agilent). We sequenced the three capture libraries on three runs, each with 48

Table 2. Voucher information and amount of DNA data produced for each sample, including number of contigs assembled using Velvet (v1.2.10), and the resulting number of aligned ultraconserved elements (UCEs) obtained. Sequence Read Archive (SRA) accession numbers provide access to all reads obtained for each individual. Family-level taxonomy follows Heinicke et al. (2018). See methods section for definitions of museum collection abbreviations. Three samples currently have only field numbers: Jonathan A. Campbell (JAC; deposited at UTA), Erik R. Wild (ERW; deposited at KU), and William E. Duellman (WED; deposited at KU).

Species	Family	Museum number	Contigs	UCEs	SRA accession
Brachycephalus quiririensis	Brachycephalidae	DZUP 522	5415	886	SAMN05559884
Craugastor augusti	Craugastoridae	UTA A-60654	3790	738	SAMN09873179
Craugastor daryi	Craugastoridae	UTA A-62648	375	71	SAMN09873180
Craugastor longirostris	Craugastoridae	MHUA 4809	2569	1301	SAMN05559889
Eleutherodactylus johnstonei	Eleutherodactylidae	ANDES-A 1912	4588	1792	SAMN09873181
Eleutherodactylus longipes	Eleutherodactylidae	JAC 29834	2710	1243	SAMN09873182
Diasporus gularis	Eleutherodactylidae	ANDES-A 3833	19214	427	SAMN09873183
Diasporus vocator	Eleutherodactylidae	CH 4786	8247	429	SAMN09873184
Adelophryne adiastola	Eleutherodactylidae	ANDES-A 2560	3680	1515	SAMN05559873
Phyzelaphryne miriame	Eleutherodactylidae	ANDES-A 3834	3875	512	SAMN09873185
Strabomantis anomalus	Strabomantidae	ANDES-A 1416	1034	136	SAMN09873186
Barycholos pulcher	Strabomantidae	KU 217782	1261	610	SAMN09873187
Lynchius nebulanastes	Strabomantidae	ERW 86	11368	2233	SAMN05559921
Oreobates quixensis	Strabomantidae	ANDES-A 1954	4975	1198	SAMN09873188
Pristimantis simonsii	Strabomantidae	WED 56667	5268	2118	SAMN09873189
Pristimantis miyatai	Strabomantidae	ANDES-A 1776	4897	1429	SAMN09873190
Espadarana prosoblepon	Centrolenidae	MVZ 149741	6094	1851	SAMN05559886
Stefania coxi	Hemiphractidae	ROM 39478	3350	1826	SAMN05559931
Dendropsophus leali	Hylidae	KU 215259	6094	1767	SAMN05559892
Hyloxalus nexipus	Dendrobatidae	KU 211806	368	1456	SAMN05559914
Leptodactylus didymus	Leptodactylidae	MHNSM 14643	5814	1909	SAMN05559919

individuals (not all individuals were included in the present study). We performed 600-cycle paired-end (300 base pairs) sequencing runs on an Illumina MiSeq at the genomics core facility of the University of Texas at Arlington (Arlington, TX, USA; http://gcf.uta.edu/).

Sequence quality control, assembly, and alignment

UCE data were processed with the pipeline provided by Faircloth et al. (2012) available at http://phyluce.readthedocs.org/en/latest/tutorial-one.html#preparing-data-for-raxml-and-examl. We trimmed sequences to remove adapters and low-quality bases using the Trimmomatic package implemented in Illumiprocessor (Bolger et al., 2014; Faircloth, 2013). We assembled contigs *de novo* for each sample using Velvet 1.2.10 Zerbino, 2010) with a kmer length of 75 and a coverage cutoff of 10. Following contig assembly, we processed the data using programs available from PHYLUCE 1.5.0 (Faircloth, 2016) (http://phyluce.readthedocs.org/en/latest/tutorial-one.html#preparing-data-for-raxml-and-examl). We identified the UCE contigs from *de novo* assemblies on a sample-by-sample basis. We used MAFFT 7.130 (Katoh et al., 2002) with default settings to align the resulting UCEs, because this program has been shown to achieve highly accurate multiple sequence alignments relative to computational costs (Pais et al., 2014). Recent analyses (Portik & Wiens, 2021) suggest that different alignment and trimming options can have relatively little impact on phylogenomic analyses of UCE data, and we generally followed the recommendations of that study (i.e., MAFFT without aggressive trimming).

Concatenated phylogenetic analyses

We inferred phylogenetic relationships from each concatenated data matrix using maximum likelihood (ML) analysis as implemented in RAxML version 8.0.19 (Stamatakis, 2014). We used the standard GTRGAMMA substitution model. We did not search the data for partitions, given the very large number of loci. Furthermore, there are few obvious *a priori* partitions for UCE data (e.g., many loci are not protein coding, so most sites cannot be assigned to codon positions). We ran two RAxML analyses for each dataset. First, we ran 20 replicate searches to find the optimal ML tree. Second, we performed bootstrapping using the autoMRE option, which automatically determines a sufficient number of bootstrap replicates. The bootstrap support values are shown on the inferred best ML tree, and all trees were rooted using the outgroup species (see above).

Species-tree analyses

We used two coalescent-based species-tree approaches designed to work on large phylogenomic datasets.

First, we used a species-tree approach (NJst) based on a matrix of internode distances across gene trees (Liu & Yu, 2011), which approximates the species tree under the multi-species coalescent. To build our species tree, we generated 100 bootstrap samples per locus using RAxML version 8.0.19 and the GTRGAMMA model. To obtain bootstrap support values for the inferred species tree, we used a two-stage bootstrap procedure in which genes were randomly resampled followed by random resampling of base pairs within the resampled genes (Seo, 2008). We ran all NJst analyses using the Species Tree Analysis Web (STRAW) Server (Shaw et al., 2013). As a second species-tree approach, we also used the Accurate Species TRee ALgorithm (ASTRAL-II 5.5.9; Mirarab & Warnow, 2015; Mirarab et al., 2014). This method estimates an unrooted species tree given a set of unrooted gene trees, under a multi-species coalescent model. Branch support for both NJst and ASTRAL-II analyses was estimated using the same bootstrap method proposed by (Seo, 2008). Species trees were rooted using the outgroups.

Missing data

Matrices for UCE data contain some loci that lack DNA sequence data for at least some samples (i.e. taxa). This occurs because the UCE probes do not bind sufficiently to all of the >2000 targeted loci in all samples. The lack of binding may reflect sequence variation among taxa and/or differences in tissue quality among samples (among other things). In deciding which loci to include in phylogenomic analyses, researchers typically choose a threshold for the maximum allowable percentage of taxa with no data, such that all loci representing less than this percentage of taxa will be excluded (e.g. loci that lack data for 50% of the taxa are excluded). To evaluate how the choice of threshold affects phylogenetic inference for our study, we varied the number of loci included in each dataset by changing the maximum percentage of missing taxa allowed for a locus to be included. We used PHYLUCE to filter the alignments to create nine data matrices that differed based on the number of loci included, where the decision to include a locus was based on the maximum percentage of taxa (including outgroups) that were missing data for that locus. We created nine matrices, each allowing different maximum amounts of taxa that were missing data per locus (10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%). For example, the 10% matrix included only those loci that had data for 90% or more of all the sampled taxa. We quantified the overall amount of missing data in a data matrix as the percentage of all cells containing '?', where the columns were nucleotide sites (not loci). The overall amount of missing data could be quite different from the maximum percentage of missing taxa allowed per locus. For example, allowing 60% to 90% missing taxa per locus generated matrices with only 49–58% missing data overall (Table 3).

We estimated the overall number of parsimony-informative sites and the overall amount of missing data in each data matrix (Table 3) using Geneious pro 8.1 (Biomatters, http://www.geneious.com/) and Python scripts available from https://phyluce.readthedocs.io/en/latest/tutorials/tutorial-1.html. We generated alignment statistics using the script get_align_summary_data.py implemented in alignment_assessment_v1 (Portik et al., 2016).

Following Streicher et al. (2016, 2018), we used the mean bootstrap value of all nodes across the tree to evaluate how the percentage of taxa that are missing data per locus and concomitant number of loci impacted the phylogenetic results. We did not consider any clades within terraranans to be well-established based on non-molecular data. Therefore, we did not focus on the bootstrap values for any particular clade. However, previous studies (Streicher et al., 2016, 2018) found that bootstrap support for well-established clades was strongly correlated with mean bootstrap support for other nodes across the tree. Therefore, we used mean support across nodes as a provisional index of performance for each approach. Nevertheless, we acknowledge that there are conditions under which mean bootstrap support could be misleading as a measure of method performance (Hillis & Bull, 1993).

Results

Phylogenetic results

Using concatenated ML, the tree inferred from the matrix allowing up to 80% of taxa lacking data per locus (Fig. 2) provided the overall highest mean bootstrap support (bs) values (Table 4). This matrix also provided the highest mean bs in the analyses with ASTRAL-II, but not NJst. For NJst, all matrices allowing up to 60 to 90% missing taxa per locus had very similar mean bs values (90.9–91.3). We considered the matrix allowing up to 80% missing taxa to be our preferred data matrix overall. This matrix contained 2,665 loci with 743,419 sites (24,079 informative) and 57% missing data cells overall.

We describe the concatenated likelihood tree from this matrix and then compare it with those obtained based on other matrices and other inference methods. Terraranans formed a monophyletic group with 100%

Table 3. Summary of data matrices based on ultraconserved elements (UCEs), organized by the maximum percentage of taxa that lack data per locus in each matrix. The summary includes the total number of UCE loci per matrix, total number of characters in aligned base pairs (bp), number of parsimony-informative sites summed across all UCEs, total number of missing data cells in each matrix, percent of sites across all loci that contain parsimony-informative variation (excluding sites that contain gap characters) in relation to the total length of the alignment, and the overall percentage of missing data cells in each matrix.

Maximum % of taxa lacking data per locus	Number of UCEs	Total DNA sequence length in bp	Parsimony-informative sites	Missing data cells	% informative sites	% missing data cells
90%	2,745	754,266	24,079	9,264,413	3.10%	58%
80%	2,665	743,419	24,079	8,844,779	3.20%	57%
70%	2,368	639,195	22,083	7,126,259	3.50%	53%
60%	1,906	503,502	17,390	5,166,952	3.50%	49%
50%	1,262	337,741	11,217	3,097,704	3.30%	44%
40%	632	172,052	5,513	1,359,560	3.20%	38%
30%	202	53,240	1,243	342,194	2.30%	31%
20%	22	5,565	83	23,879	1.50%	20%
10%	4	1,265	14	944	1.10%	4%

bs. Brachycephalidae, represented by *Brachycephalus quiririensis*, was recovered as the sister taxon to a maximally-supported clade (bs = 100%) uniting all other sampled terraranan taxa. Eleutherodactylidae was also maximally supported as monophyletic. Within Eleutherodactylidae, the subfamilies Eleutherodactylinae (*Diasporus, Eleutherodactylus*) and Phyzelaphryninae (*Adelophryne, Phyzelaphryne*) were each maximally supported as monophyletic (bs = 100%). We found maximal support also for a clade uniting Craugastoridae (in our phylogeny represented by *Craugastor augusti*, *C. daryi,* and *C. longirostris*) and Strabomantidae (*sensu* AmphibiaWeb, 2020) represented by *Barycholos pulcher, Lynchius nebulanastes*, *Oreobates quixensis, Pristimantis miyatai, P. simonsii,* and *Strabomantis anomalus*. However, Strabomantidae is non-monophyletic in this tree, because Craugastoridae is nested within Strabomantidae (or, put another way, because *Strabomantis* groups with *Craugastor* instead of with other genera of Strabomantidae). There is moderate support (bs = 73%) for the clade uniting *Craugastor* (Craugastoridae) and *Strabomantis* (Strabomantidae). The monophyly of *Craugastor* was strongly supported (bs = 98%). The sister relationship between the subgenera *Hylactophryne* (represented by *C. augusti*) and *Craugastor* (represented by *C. longirostris*) relative to the subgenus *Campbellius* (represented by *C. daryi*) was also well supported (bs = 93%). A maximally supported clade united *Lynchius, Oreobates*, *Pristimantis,* and *Barycholos*. These genera are currently assigned to the subfamilies Pristimantinae (*Lynchius, Oreobates*, and *Pristimantis*) and Holoadeninae (*Barycholos*) of Strabomantidae. The genera *Lynchius* and *Oreobates* formed the sister clade (bs = 100%) to a clade including *Barycholos* and *Pristimantis* (bs =100%; Fig. 2). Thus, Pristimantinae was paraphyletic with respect to Holoadeninae (because *Barycholos* was nested inside of Pristimantinae), with strong support for the relevant relationships.

Coalescent-based species-tree analyses using NJst and ASTRAL-II were then applied to the preferred matrix (up to 80% missing taxa per locus). These analyses yielded trees that generally agreed with each other and with the concatenated ML results (Fig. 2 and Additional file 1: Fig. S1). We describe the branch support from these two methods, and then explain the one topological difference with the concatenated results. NJst and ASTRAL-II provided maximal support for the monophyly of Terraranae, and moderately strong support for the clade uniting Eleutherodactylidae, Craugastoridae, and Strabomantidae to the exclusion of Brachycephalidae (NJst: bs = 92%: ASTRAL-II: bs = 85%). Support was strong for monophyly of Eleutherodactylidae (NJst: bs = 96%; ASTRAL-II: bs = 100%), Eleutherodactylinae (NJst: bs = 99%; ASTRAL-II: bs = 100%), and Phyzelaphryninae (NJst: bs = 98%; ASTRAL-II: bs = 100%). Support for the monophyly of the clade uniting Craugastoridae and Strabomantidae was moderately strong for NJst and very strong for ASTRAL-II (NJst: bs = 85%; ASTRAL-II: bs = 96%). Craugastoridae was again inside Strabomantidae, rendering Strabomantidae paraphyletic (again, given the sister relationship of *Strabomantis* with *Craugastor*). However, support was moderate for the clade uniting *Strabomantis* and Craugastoridae (NJst: bs = 89%; ASTRAL-II: bs = 70%). Relationships among *Barycholos, Lynchius, Oreobates*, and *Pristimantis* were the same with NJst and ASTRAL-II as with concatenated ML, and remained strongly supported (bs= 94 – 100%). Importantly, *Barycholos* (Holoadeninae) was again nested inside of Pristimantinae (*Lynchius, Oreobates*, and *Pristimantis*) with strong support for the clade uniting Holoadeninae with *Pristimantis* (NJst: bs = 97%; ASTRAL-II: bs = 100%).

The only topological difference among trees inferred by different methods involved relationships among

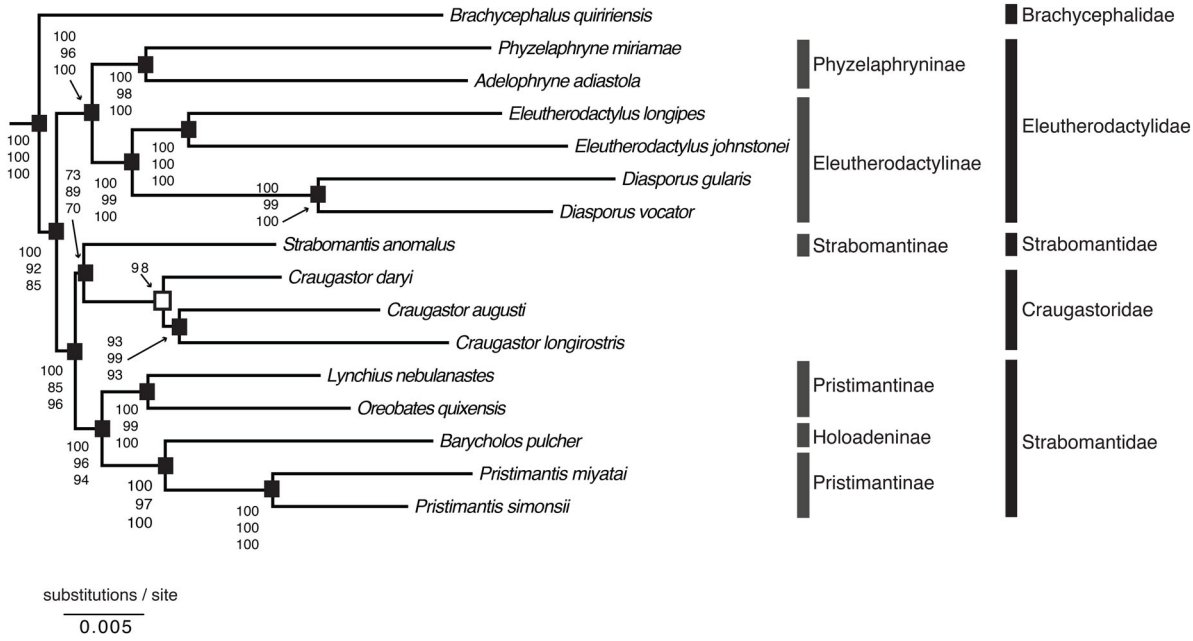


Fig. 2. Relationships among terraranan frogs based on a concatenated maximum likelihood (ML) phylogenetic analysis. The data matrix included 2,665 UCE loci for a total of 743,419 aligned base pairs and included loci with up to 80% of taxa lacking data for those loci. Numbers next to each internal branch indicate bootstrap support values from concatenated ML analysis (top), coalescent-based species-tree analyses from NJst (middle), and ASTRAL-II (bottom). The black squares indicate the clades that were recovered in all three analyses of this data matrix. The white square indicates an internal branch (monophyly of Craugastoridae) that was not recovered in the NJst or ASTRAL-II analyses, and thus has only the ML bootstrap score. In the NJst and ASTRAL-II analyses of this dataset, *Strabomantis* is inside *Craugastor* (see Supplementary Fig. S1 online). Note that Strabomantidae and Pristimantinae are paraphyletic in all of these trees. The full trees from each analysis (including branch lengths and outgroups) are provided in Additional file: Fig. S1 online. Family-level taxonomy follows AmphibiaWeb (2020).

craugastorids and *Strabomantis* (see Additional file 1: Fig. S1). The concatenated ML tree inferred *Strabomantis* as sister to a monophyletic *Craugastor*, but the NJst and ASTRAL-II trees inferred *S. anomalus* and *C. daryi* as sister taxa (NJst: bs = 92%; ASTRAL-II: bs = 73%), with this clade as the sister taxon of the other two sampled species of *Craugastor*. Thus, both Craugastoridae and Strabomantidae were paraphyletic in the NJst and ASTRAL trees, whereas only Strabomantidae was paraphyletic in the concatenated ML tree. The different positions of *Strabomantis* and *C. daryi* may reflect the low number of UCE loci recovered for these taxa (136 and 71 loci for *S. anomalus* and *C. daryi*, respectively; Table 2).

Impact of missing data

As more taxa without data were allowed per locus (from 10% up to 90%), the number of UCE loci included and the total alignment length increased (Table 3). However, the range of variation in the overall percentage of missing data was limited. Specifically, the percentage of cells in the data matrix containing '?' (missing bases) ranged from 4–58%, but mostly from 30–58%). The number of parsimony-informative sites included also increased with the number of missing base pairs included (Table 3; Spearman's rank correlation, $r_s = 0.976$; $P < 0.0001$). This almost certainly explains why mean bootstrap support values were greater when loci with higher percentages of missing taxa were included. There were strong correlations between the number of loci included and mean bootstrap support values for all three methods (Spearman's rank correlation, ML $r_s = 0.95$, $P < 0.0001$; NJst $r_s = 0.92$, $P = 0.0004$; ASTRAL-II $r_s = 0.88$, $P = 0.0017$).

The phylogenetic results described above were based on the matrix allowing up to 80% missing taxa per locus (57% missing data cells overall). For matrices with 50% to 90% missing taxa per locus (44–58% missing data cells overall), all analyses (ML, NJst, and ASTRAL-II; Additional files 1–4: Figs. S1–S4) supported the monophyly of Terraranae, Eleutherodactylidae, and Strabomantidae + Craugastoridae. All of these analyses also placed Brachycephalidae as the sister taxon to

Table 4. Mean bootstrap support across all internal branches in trees inferred from nine data matrices. Each matrix included different numbers of loci (Table 3), and each locus included different maximum percentages of taxa that lack data per locus (10% to 90%). Results are given for three phylogenetic methods (concatenated ML, and the coalescent-based species-tree methods NJst and ASTRAL-II).

Maximum % of taxa lacking data per locus	Mean bootstrap support across all internal branches		
	ML	NJst	ASTRAL-II
90%	94.9	91.6	93.0
80%	95.7	90.9	94.1
70%	95.4	91.6	89.6
60%	93.8	91.3	93.0
50%	91.8	88.9	93.0
40%	91.6	80.3	83.4
30%	51.7	56.3	56.0
20%	41.3	15.0	13.0
10%	18.0	16.4	12.5

Eleutherodactylidae + Craugastoridae + Strabomantidae. Importantly, these trees also inferred *Strabomantis* to be with or within Craugastoridae (rendering Strabomantidae paraphyletic) and *Barycholos* with *Pristimantis* (rendering Pristimantinae paraphyletic). The number of loci included varied from 1,262 when allowing 50% missing taxa per locus to 2,745 UCEs when allowing 90% missing taxa per locus (Table 3).

The phylogenetic results were more variable when allowing ≤40% missing taxa per locus. When loci were included with a maximum of 40% missing taxa per locus, the data matrix included only 632 loci, and the ASTRAL-II analysis inferred that *Brachycephalus quiriensis* was inside the clade of Craugastoridae + Strabomantidae (bs = 7%; see Supplementary Fig. S4E online), whereas the concatenated ML and NJst topologies were unaffected. The concatenated and NJst analyses based on more loci inferred Brachycephalidae as sister to Eleutherodactylidae + Craugastoridae + Strabomantidae. Thus, the ASTRAL-II analyses were more sensitive than NJst and ML to the limited number of loci. The matrix allowing only up to 30% missing taxa per locus included only 202 loci. Mean bootstrap values dropped precipitously across the three inference methods (Table 4). In the data matrices allowing 10 or 20% missing taxa per locus, the total number of loci was very low (4 and 22 UCEs, respectively), and the concatenated ML, NJst, and ASTRAL-II analyses did not recover the monophyly of Eleutherodactylidae or Craugastoridae (Additional file 2–4: Figs. S2–S4). The topologies were incongruent with many previously hypothesized relationships, including monophyly of Terraranae (Additional file 2–4: Figs. S2–S4).

Discussion

Previous studies of higher-level relationships among terraranan frogs have inferred many conflicting trees with often weak support values (Fig. 1). Here, we addressed terraranan phylogeny using the largest number of loci so far, including >2,600 loci for the 16 ingroup taxa. We also explored the impacts of missing data on phylogenomic analyses, and found that the best-supported phylogeny for all three methods used (concatenated ML, ASTRAL-II, NJst) was obtained by including many loci for which some taxa lacked data (Table 4). Based on concatenated likelihood and species-tree analyses of the optimal dataset (2,665 loci, and 57% missing data in the matrix overall), our results provided a generally well-supported hypothesis (Fig. 2). This phylogeny should help resolve the controversial relationships among brachycephalids, eleutherodactylids, and other families, and reveals the non-monophyly of the large family Strabomantidae and the subfamily Pristimatinae. Below, we describe the implications of our results for terraranan phylogeny and taxonomy, and then address the topic of missing data.

Terraranan phylogeny

The family-level phylogenetic relationships recovered here are most similar to those of Pyron and Wiens (2011), Feng et al. (2017), Hutter et al. (2017), and Streicher et al. (2018). Specifically, we recovered Brachycephalidae as the sister to Eleutherodactylidae + Craugastoridae + Strabomantidae, with relatively strong support for the latter clade (concatenated ML: bs = 100%; NJst: bs = 92%; ASTRAL-II: bs = 85%). This is consistent with the analysis of Pyron and Wiens (2011), based on concatenated ML analysis of three nuclear and nine mitochondrial genes from 340 terraranan species. Using a dataset of 95 nuclear genes from 16 terraranan species, Feng et al. (2017) also inferred Brachycephalidae to be sister to Eleutherodactylidae + Craugastoridae + Strabomantidae, but with only moderate support for the latter clade (ML bs = 68%). This

same topology was also recovered with strong support (bs = 100%) in a study that used 2,214 UCEs but included only five terraranan taxa (Streicher et al., 2018). Our results contradict Pyron's (2014) analysis of three nuclear and nine mitochondrial genes with 418 species, which inferred Brachycephalidae + Eleutherodactylidae as sister to Craugastoridae + Strabomantidae. The same topology was also supported by Pinto-Sánchez et al. (2014) and Hutter et al. (2017). Our findings also contradict the analysis of Heinicke et al. (2018) based on 389 nuclear genes for 30 terraranan species, which inferred Eleutherodactylidae to be sister to Brachycephalidae + Craugastoridae + Strabomantidae (see also Padial et al., 2014). However, they found only weak support from a species-tree method for most relationships, including the clade excluding Eleutherodactylidae (Fig. 1I). Hime et al. (2021), using a subset of 220 anchored hybrid enrichment loci from Heinicke et al. (2018), recovered the same topology as Heinicke et al. (2018) with strong support from RAxML analysis but very weak support for most terraranan relationships using a species-tree method (ASTRAL). Among all these studies, our results are based on the largest number of loci, although from fewer terraranan taxa than sampled in some studies. Unlike most previous studies, our results for these family-level relationships are generally well supported by both concatenated analyses and coalescent-based species-tree analyses.

Another conflict among recent studies is the position of *Strabomantis* and the recognition of the family Strabomantidae. We found Strabomantidae paraphyletic with respect to Craugastoridae. This result was also found by Pyron and Wiens (2011) and Hutter et al. (2017), using fewer loci but much more extensive taxon sampling. Within Craugastoridae, Pyron and Wiens (2011) recognized the monogeneric subfamily Strabomantinae, which is consistent with our concatenated ML results, but conflicts with our species-tree results, which found *Strabomantis* within *Craugastor* (Additional files 1, 3, 4: Figs. S1, S3, S4). Although its exact position is uncertain, we did not find support for a close relationship between *Strabomantis* and other members of Strabomantidae. In contrast, Heinicke et al. (2018) recognized Strabomantidae as the sister taxon of Craugastoridae, and within the former they found *Strabomantis* to be the sister taxon of a clade including *Barycholos*, *Oreobates*, and *Pristimantis* (among other genera). Feng et al. (2017) also recognized Strabomantidae, with *Strabomantis*, *Pristimantis*, *Niceforonia* (as *Hypodactylus*), and *Barycholos* in a clade that was the sister taxon of *Craugastor*. In our concatenated ML tree, *Strabomantis* is the sister taxon of *Craugastor* (bs = 73%), rather than more closely related to other sampled strabomantid genera (*Barycholos, Lynchius, Oreobates, Pristimantis*). Both species-tree methods found *Strabomantis* inside *Craugastor*, as sister to *C. daryi* (NJst: bs = 89%; ASTRAL-II: bs = 73%). Either topology would render Strabomantidae non-monophyletic with respect to Craugastoridae. Therefore, our results support those of Pyron and Wiens (2011), Gomez-Mestre et al. (2012), and others in suggesting that Strabomantidae is within Craugastoridae. We note that *Strabomantis* cannot be simply removed from Strabomantidae because it is the type genus of the family. We also note that synonymizing Strabomantidae and placing strabomantid taxa within an expanded Craugastoridae would still be consistent with the phylogeny even if our results were wrong about the position of *Strabomantis* and some previous hypotheses were correct instead, as long as Craugastoridae and Strabomantidae are sister taxa. All previous analyses have been consistent with this result (Fig. 1). Thus, this move would avoid the non-monophyly of Strabomantidae found in this and previous studies (e.g. Hutter et al., 2017; Pyron & Wiens, 2011) and would be very conservative in its assumptions about the phylogeny. Importantly, we recognize that our data were relatively limited for *Strabomantis* (136 loci), despite the large number of loci overall.

Another possible explanation for our results is that there is a problem in the taxonomic placement of the species *Strabomantis anomalus* (the only sampled *Strabomantis* species) rather than a problem in the phylogenetic position of *Strabomantis* and Strabomantinae overall. However, a large-scale phylogeny of anurans (Pyron & Wiens, 2011) showed that *S. anomalus* is clearly within *Strabomantis*, as sister to *S. bufoniformis*. This phylogeny also agreed with ours in placing *Strabomantis* as sister to Craugastorinae (see also Hutter et al., 2017).

The relationships among genera of Holoadeninae and Pristimantinae also differ between our results and previous studies. Here, we found Pristimantinae (represented by *Lynchius, Oreobates,* and *Pristimantis*) to be paraphyletic with respect to Holoadeninae (represented by *Barycholos*), with strong support from all three methods (Fig. 2). Several previous studies sampled these four genera (along with many other genera), including Hedges et al. (2008), Heinicke et al. (2018), Hutter et al. (2017), Pyron (2014), and Pyron and Wiens (2011). All inferred the following relationships (*Barycholos*, (*Pristimantis*, (*Lynchius*, *Oreobates*))). On the other hand, Padial et al. (2014) recovered a third topology: (*Pristimantis*, (*Barycholos*, (*Lynchius*, *Oreobates*))). Our support here for the relationships

among these four genera is higher than that obtained in most previous studies and is based on many more loci (but with only 610 loci in *Barycholos*). We also note that taxon sampling was much more extensive in most previous studies, despite their more limited sampling of genes (Heinicke et al., 2018; Hutter et al., 2017; Pyron, 2014; Pyron & Wiens, 2011). Our results with Holoadeninae (*Barycholos*) inside Pristimantinae (*Lynchius*, *Pristimantis*, and *Oreobates*) suggest that one of these two subfamilies may not be valid, but this hypothesis would benefit from further testing with more taxa and more loci in all sampled taxa.

In summary, our analyses using concatenated ML analysis and coalescent-based species-tree methods (NJst and ASTRAL-II) found high support for Brachycephalidae as the sister taxon to the clade Eleutherodactylidae + Craugastoridae + Strabomantidae, as also found by Pyron and Wiens (2011), Feng et al. (2017), Hutter et al. (2017), and Streicher et al. (2018), but *contra* Hedges et al. (2008), Heinicke et al. (2009, 2018), Padial et al. (2014), Pinto-Sánchez et al. (2014), Pyron (2014), and Hime et al. (2021). Our results also provide support for Craugastoridae being nested inside a paraphyletic Strabomantidae, as also found by Pyron and Wiens (2011) and Hutter et al. (2017). We also found strong support for Holoadeninae being nested inside Pristimantinae.

We suggest that the highest priority for future studies will be to include more taxa using datasets that also sample very large numbers of nuclear loci. For example, because we had no samples of *Ceuthomantis* in this study, we could not evaluate the hypothesis that this genus is sister to the rest of Terraranae. Nevertheless, this hypothesis has been consistently supported by all other recent studies (Feng et al., 2017; Heinicke et al., 2009, 2018; Hime et al., 2021; Pyron, 2014; Pyron & Wiens, 2011), except that of Padial et al. (2014). The possible position of *Strabomantis* within *Craugastor* should also be addressed with greater taxon sampling from both genera, and more work is needed on the placement of *Barycholos* (Holoadeninae) within Pristimantinae. Additional taxonomic changes may be needed if these results are supported in future studies. It will also be important to include several genera that have not yet been included in molecular phylogenetic analyses (i.e., *Atopophrynus, Dischidodactylus*, and *Geobatrachus*).

Taxonomic implications

An obvious solution to the problem of Craugastoridae being nested inside of Strabomantidae is to revert to the earlier taxonomy, in which the subfamilies of Strabomantidae were considered subfamilies of Craugastoridae. Importantly, such a taxonomy is consistent both with our phylogeny, and with all recent alternative phylogenies for terraranans (Fig. 1). We do not think that there is any logical argument for preferring a taxonomy that recognizes a possibly paraphyletic family (Strabomantidae) over one that is instead consistent with all proposed phylogenies for the group.

We note that our analyses also imply the potential need for changes in the taxonomy of craugastorid subfamilies (i.e. non-monophyly of Pristimantinae). However, in this case, our results here do not have such clear precedents in previous phylogenetic studies. Moreover, we would prefer to see more complete sampling of the genera in the relevant subfamilies before changing the assignment of these genera to subfamilies.

Impacts of excluding missing data and loci

A major concern in phylogenomic analyses is the possible impact of missing data on phylogenetic inference (Crotti et al., 2019; Jiang et al., 2014; Philippe et al., 2004; Roure et al., 2013; Streicher et al., 2016; Wiens & Morrill, 2011; Xi et al., 2016). Our results suggest that a larger problem may instead be excluding loci because of concerns about a lack of data for some of the sampled taxa. Here, we found that mean branch support values from concatenated ML, ASTRAL-II, and NJst analyses were greater for datasets allowing more, not fewer, missing taxa per locus (Tables 3 and 4; Spearman's rank correlation: concatenated ML: $r_s = 0.95$, $P < 0.0001$; NJst: $r_s = 0.92$, $P = 0.0004$; ASTRAL-II: $r_s = 0.88$, $P = 0.0017$). This almost certainly occurs because permitting more missing taxa per locus allowed for the inclusion of a greater number of loci and sites, which increased mean branch support in both species-tree and concatenated ML analyses (Table 3). Similarly, recent studies of genome-wide single nucleotide polymorphism data also found that mean branch support values in concatenated ML analyses increased in datasets with more sites but with a larger proportions of missing taxa (Crotti et al., 2019). Although increased branch support may not always reflect phylogenetic accuracy, other simulation and empirical studies have now shown that increasing the number of genes included can increase the accuracy of phylogenetic analyses, despite the potential increase in the number of missing data cells included (Jiang et al., 2014; Liu & Yu, 2011; Roure et al., 2013; Streicher et al., 2016). These results strongly suggest that the consequences of drastically reducing the number of loci sampled appear to be worse than those of including loci for which some taxa lack data. Finally, although some

might argue that this issue has already been resolved, empirical phylogenomic studies continue to exclude substantial numbers of loci because of concerns about missing data (e.g. Hime et al., 2021).

Conclusions

Many previous studies of relationships among major groups of terraranan frogs were in disagreement and had weak support for key branches (Fig. 1, Table 1). Our results provide a generally well-supported estimate of relationships among the sampled terraranan families and subfamilies based on concatenated ML and species-tree methods, including the largest number of loci considered so far, albeit for a relatively small taxon set. The preferred phylogeny inferred Eleutherodactylidae to be the sister to Strabomantidae + Craugastoridae, in contrast to many (but not all) previous studies with fewer loci but more taxa. The results also suggest that Strabomantidae is not monophyletic (as also found in previous studies with much greater taxon sampling), a problem that can easily be rectified by re-expanding Craugastoridae to include Strabomantidae. We also found that the strabomantid subfamily Pristimantiae appears to be paraphyletic with respect to the subfamily Holoaedeninae. Finally, we found that the highest mean branch support for these data is achieved by including large numbers of loci, even when many loci are missing data for some of the sampled taxa. These results argue against the continuing practice of excluding loci a priori because of the fear of increasing the absolute amount of missing data.

List of abbreviations

ASTRAL-II	Accurate Species TRee Algorithm
bs	bootstrap support
gDNA	Genomic DNA
ICZN	International Commission for Zoological Nomenclature
MAFFT	multiple sequence alignment program
ML	maximum likelihood
STRAW	Species TRee Analysis Web server
UCEs	Ultraconserved Elements

Acknowledgements

We are grateful for tissue loans from the museum collections cited in the methods. Thanks to IT Management and *ExaCore* of the Office of the Vice President for Research & Creation at the Universidad de Los Andes for high-performance computing services and technical assistance. We thank Andres M. Cuervo, Luisa Dueñas, Carlos E. Guarnizo, Santiago Herrera, Andrea Paz, Camila Plata, Evan Twomey and two anonymous reviewers for providing valuable comments on earlier versions of this manuscript. Special thanks to David O. Botero, Catalina Ramírez-Portilla, and Fanny Gonzáles for help with script writing and bioinformatic advice. Thanks to Carolina Amorocho and Luisa Castellanos for help in the lab and with paperwork. We thank Jorge Avendaño for his advice with the figures.

Availability of data and materials

The genetic data alignments datasets supporting the results of this article are available via the NHM Data Portal (https://data.nhm.ac.uk/).

Author contributions

LSB, AJC, JJW, and JWS conceived and designed the study; LSB, JWS, and ECM performed laboratory work; LSB, JWS, and MRP conducted analyses; LSB and AJC wrote an initial draft, with extensive reviewing and rewriting by all authors; LSB, AJC, and JJW acquired funding.

Ethics approval and consent to participate

All samples used in this study were donated by natural history museums and herpetological collections. No live animals were used in this study.

Disclosure statement

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

Supplemental material

Supplemental material for this article can be accessed here: https://doi.org/doi/10.1080/14772000.2021.1933249.

Funding

This work was supported by Colciencias under grant number 567 for PhD studies; Universidad de los Andes through a *Crédito Condonable*; Universidad de los Andes FAPA fund; and U.S. National Science Foundation grant DEB 1655690.

References

Acosta-Galvis, A. R., Streicher, J. W., Manuelli, L., Cuddy, T., & de Sá, R. O. (2018). Molecular insights into the phylogenetic placement of the poorly known genus *Niceforonia* Goin & Cochran 1963 (Anura: Brachycephaloidea). *Zootaxa*, *4514*(4), 487–500. https://doi.org/10.11646/zootaxa.4514.4.3

Alexander, A. M., Su, Y.-C., Oliveros, C. H., Olson, K. V., Travers, S. L., & Brown, R. M. (2017). Genomic data reveals potential for hybridization, introgression, and incomplete lineage sorting to confound phylogenetic relationships in an adaptive radiation of narrow-mouth frogs. *Evolution*, *71*(2), 475–488. https://doi.org/10.1111/evo.13133

AmphibiaWeb. 2020. http://amphibiaweb.org. University of California, Berkeley, CA. Accessed 15 Nov 2020.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004). Ultraconserved elements in the human genome. *Science*, *304*(5675), 1321–1325. https://doi.org/10.1126/science.1098119

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Crawford, N. G., Parham, J. F., Sellas, A. B., Faircloth, B. C., Glenn, T. C., Papenfuss, T. J., Henderson, J. B., Hansen, M. H., & Simison, W. B. (2015). A phylogenomic analysis of turtles. *Molecular Phylogenetics and Evolution*, *83*, 250–257. https://doi.org/10.1016/j.ympev.2014.10.021

Crotti, M., Barratt, C. D., Loader, S. P., Gower, D. J., & Streicher, J. W. (2019). Causes and analytical impacts of missing data in RADseq phylogenetics: insights from an African frog (*Afrixalus*). *Zoologica Scripta*, *48*(2), 157–167. https://doi.org/10.1111/zsc.12335

Dubois, A. (2009). Miscellanea nomenclatorica batrachologica 20. Class-series nomina are nouns in the nominative plural: Terrarana Hedges, Duellman & Heinicke, 2008 must be emended. *Alytes*, *26*, 165–175.

Duellman, W. E., & Trueb, L. (1994). *Biology of Amphibians*. The Johns Hopkins University Press.

Faircloth, B. C. (2013). Illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. https://doi.org/10.6079/J9ILL

Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, *32*(5), 786–788. https://doi.org/10.1093/bioinformatics/btv646

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, *61*(5), 717–726. https://doi.org/10.1093/sysbio/sys004

Faircloth, B. C., Sorenson, L., Santini, F., & Alfaro, M. E. (2013). A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). *PLoS One*, *8*(6), e65923. https://doi.org/10.1371/journal.pone.0065923

Feng, Y.-J., Blackburn, D. C., Liang, D., Hillis, D. M., Wake, D. B., Cannatella, D. C., & Zhang, P. (2017). Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous–Paleogene boundary. *Proceedings of the National Academy of Sciences*, *114*(29), E5864–E5870. https://doi.org/10.1073/pnas.1704632114

Frost, D. R., Grant, T., Faivovich, J., Bain, R. H., 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. M., & Wheeler, W. C. (2006). The amphibian tree of life. *Bulletin of the American Museum Natural History*, *297*, 1–377. https://doi.org/10.1206/0003-0090(2006)297[0001:TATOL2.0.CO;2]

Gomez-Mestre, I., Pyron, R. A., & Wiens, J. J. (2012). Phylogenetic analyses reveal unexpected patterns in the evolution of reproductive modes in frogs. *Evolution*, *66*(12), 3687–3700. https://doi.org/10.1111/j.1558-5646.2012.01715.x

Gonzalez-Voyer, A., Padial, J. M., Castroviejo-Fisher, S., de la Riva, I., & Vilà, C. (2011). Correlates of species richness in the largest Neotropical amphibian radiation. *Journal of Evolutionary Biology*, *24*(5), 931–942. https://doi.org/10.1111/j.1420-9101.2011.02243.x

Hedges, S. B., Duellman, W. E., & Heinicke, M. P. (2008). New World direct-developing frogs (Anura: Terrarana): Molecular phylogeny, classification, biogeography, and conservation. *Zootaxa*, *1737*(1), 1–182. https://doi.org/10.11646/zootaxa.1737.1.1

Heinicke, M. P., Duellman, W. E., Trueb, L., Means, D. B., MacCulloch, R. D., & Hedges, S. B. (2009). A new frog family (Anura: Terrarana) from South America and an expanded direct-developing clade revealed by molecular phylogeny. *Zootaxa*, *2211*(1), 1–35. https://doi.org/10.11646/zootaxa.2211.1.1

Heinicke, M. P., Lemmon, A. R., Moriarty, E., Mcgrath, K., & Hedges, S. B. (2018). Phylogenomic support for evolutionary relationships of New World direct-developing frogs (Anura: Terraranae). *Molecular Phylogenetics and Evolution*, *118*, 145–155. https://doi.org/10.1016/j.ympev.2017.09.021

Hillis, D. M., & Bull, J. (1993). An empirical test of bootstraping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology*, *42*(2), 182–192. https://doi.org/10.1093/sysbio/42.2.182

Hime, P. M., Lemmon, A. R., Lemmon, E. C. M., Prendini, E., Brown, J. M., Thomson, R. C., Kratovil, J. D., Noonan, B. P., Pyron, R. A., Peloso, P. L. V., Kortyna, M. L., Keogh, J. S., Donnellan, S. C., Mueller, R. L., Raxworthy, C. J., Kunte, K., Ron, S. R., Das, S., Gaitonde, N., … Weisrock, D. W. (2021). Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. *Systematic Biology*, *70*(1), 49–66. https://doi.org/10.1093/sysbio/syaa034

Hutter, C. R., Lambert, S. M., & Wiens, J. J. (2017). Rapid diversification and time explain amphibian richness at different scales in the tropical Andes, Earth's most biodiverse hotspot. *The American Naturalist*, *190*(6), 828–843. https://doi.org/10.1086/694319

Jetz, W., & Pyron, R. A. (2018). The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. *Nature Ecology & Evolution*, *2*(5), 850–858. https://doi.org/10.1038/s41559-018-0515-5

Jiang, W., Chen, S. Y., Wang, H., Li, D. Z., & Wiens, J. J. (2014). Should genes with missing data be excluded from phylogenetic analyses? *Molecular Phylogenetics and Evolution*, *80*, 308–318. https://doi.org/10.1016/j.ympev.2014.08.006

Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Liu, L., & Yu, L. (2011). Estimating species trees from unrooted gene trees. *Systematic Biology*, *60*(5), 661–667. https://doi.org/10.1093/sysbio/syr027

Lynch, J. D. (1971). Evolutionary relationships, osteology, and zoogeography of leptodactyloid frogs. *University of Kansas Museum of Natural History Miscellaneous Publications*, *53*, 1–238.

McCormack, J. E., Maley, J. M., Hird, S. M., Derryberry, E. P., Graves, G. R., & Brumfield, R. T. (2012). Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. *Molecular Phylogenetics and Evolution*, *62*(1), 397–406. https://doi.org/10.1016/j.ympev.2011.10.012

McCormack, J. E., Tsai, W. L. E., & Faircloth, B. C. (2016). Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*, *16*(5), 1189–1203. https://doi.org/10.1111/1755-0998.12466

Meiklejohn, K. A., Faircloth, B. C., Glenn, T. C., Kimball, R. T., & Braun, E. L. (2016). Analysis of a rapid evolutionary radiation using ultraconserved elements: Evidence for a bias in some multispecies coalescent methods. *Systematic Biology*, *65*(4), 612–627. https://doi.org/10.1093/sysbio/syw014

Mirarab, S., Reaz, R., Bayzid, M. S., Zimmermann, T., Swenson, M. S., & Warnow, T. (2014). ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics*, *30*(17), i541–548. https://doi.org/10.1093/bioinformatics/btu462

Mirarab, S., & Warnow, T. (2015). ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, *31*(12), i44–i52. https://doi.org/10.1093/bioinformatics/btv234

Motta, A. P., Taucce, P. P. G., Haddad, C. F. B., & Canedo, C. (2021). A new terraranan genus from the Brazilian Atlantic Forest with comments on the systematics of Brachycephaloidea (Amphibia: Anura). *Journal of Zoological Systematics and Evolutionary Research*, *59*(3), 663–679. https://doi.org/10.1111/jzs.12452

Padial, J. M., Grant, T., & Frost, D. R. (2014). Molecular systematics of terraranas (Anura: Brachycephaloidea) with an assessment of the effects of alignment and optimality criteria. *Zootaxa*, *3825*, 1–132. https://doi.org/10.11646/zootaxa.3825.1.1

Pais, F. S.-M., Ruy, P., de, C., Oliveira, G., & Coimbra, R. S. (2014). Assessing the efficiency of multiple sequence alignment programs. *Algorithms for Molecular Biology : AMB*, *9*(1), 4. https://doi.org/10.1186/1748-7188-9-4

Philippe, H., Snell, E. A., Bapteste, E., Lopez, P., Holland, P. W. H., & Casane, D. (2004). Phylogenomics of eukaryotes: Impact of missing data on large alignments. *Molecular Biology and Evolution*, *21*(9), 1740–1752. https://doi.org/10.1093/molbev/msh182

Pie, M. R., Bornschein, M. R., Ribeiro, L. F., Faircloth, B. C., & McCormack, J. E. (2019). Phylogenomic species delimitation in microendemic frogs of the Brazilian Atlantic Forest. *Molecular Phylogenetics and Evolution*, *141*, 106627. https://doi.org/10.1016/j.ympev.2019.106627

Pinto-Sánchez, N. R., Crawford, A. J., & Wiens, J. J. (2014). Using historical biogeography to test for community saturation. *Ecology Letters*, *17*(9), 1077–1085. https://doi.org/10.1111/ele.12310

Pinto-Sánchez, N. R., Ibáñez, R., Madriñán, S., Sanjur, O. I., Bermingham, E., & Crawford, A. J. (2012). The Great American Biotic Interchange in frogs: Multiple and early colonization of Central America by the South American genus *Pristimantis* (Anura: Craugastoridae). *Molecular Phylogenetics and Evolution*, *62*(3), 954–972. https://doi.org/10.1016/j.ympev.2011.11.022

Portik, D. M., Smith, L. L., & Bi, K. (2016). An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). *Molecular Ecology Resources*, *16*(5), 1069–1083. https://doi.org/10.1111/1755-0998.12541

Portik, D. M., & Wiens, J. J. (2021). Do alignment and trimming methods matter for phylogenomic (UCE) analyses? *Systematic Biology*, *70*(3), 440–462. https://doi.org/10.1093/sysbio/syaa064

Pyron, R. A. (2014). Biogeographic analysis reveals ancient continental vicariance and recent oceanic dispersal in amphibians. *Systematic Biology*, *63*(5), 779–797. https://doi.org/10.1093/sysbio/syu042

Pyron, R. A., & Wiens, J. J. (2011). A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution*, *61*(2), 543–583. https://doi.org/10.1016/j.ympev.2011.06.012

Roure, B., Baurain, D., & Philippe, H. (2013). Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular Biology and Evolution*, *30*(1), 197–214. https://doi.org/10.1093/molbev/mss208

Seo, T. K. (2008). Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Algorithms in Molecular Biology*, *25*(5), 960–971. https://doi.org/10.1093/molbev/msn043

Shaw, T. I., Ruan, Z., Glenn, T. C., & Liu, L. (2013). STRAW: Species TRee Analysis Web server. *Nucleic Acids Research*, *41*(Web Server issue), W238–241. https://doi.org/10.1093/nar/gkt377

Smith, B. T., Harvey, M. G., Faircloth, B. C., Glenn, T. C., & Brumfield, R. T. (2014). Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology*, *63*(1), 83–95. https://doi.org/10.1093/sysbio/syt061

Stamatakis, A. (2014). RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1315. https://doi.org/10.1093/bioinformatics/btu033

Streicher, J. W., Miller, E. C., Guerrero, P. C., Correa, C., Ortiz, J. C., Crawford, A. J., Pie, M. R., & Wiens, J. J. (2018). Evaluating methods for phylogenomic analyses, and a new phylogeny for a major frog clade (Hyloidea) based on 2,214 loci. *Molecular Phylogenetics and Evolution*, *119*, 128–143. https://doi.org/10.1016/j.ympev.2017.10.013

Streicher, J. W., Schulte, J. A., & Wiens, J. J. (2016). How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in iguanian lizards. *Systematic Biology*, *65*(1), 128–145. https://doi.org/10.1093/sysbio/syv058

Streicher, J. W., & Wiens, J. J. (2016). Phylogenomic analyses reveal novel relationships among snake families. *Molecular Phylogenetics and Evolution*, *100*, 160–169. https://doi.org/10.1016/j.ympev.2016.04.015

Streicher, J. W., & Wiens, J. J. (2017). Phylogenomic analyses of more than 4000 nuclear loci resolve the origin of snakes among lizard families. *Biology Letters*, *13*(9), 20170393. https://doi.org/10.1098/rsbl.2017.0393

Taboada, C., Grant, T., Lynch, J. D., & Faivovich, J. (2013). New morphological synapomorphies for the New World direct-developing frogs (Amphibia: Anura: Terrarana). *Herpetologica*, *69*(3), 342–357. https://doi.org/10.1655/HERPETOLOGICA-D-13-00019

Wheeler, W. (1999). Fixed character states and the optimization of molecular sequence data. *Cladistics*, *15*(4), 379–385. https://doi.org/10.1111/j.1096-0031.1999.tb00274.x

Wheeler, W. (2001). Homology and the optimization of DNA sequence data. *Cladistics*, *17*(1), S3–S11. https://doi.org/10.1111/j.1096-0031.2001.tb00100.x

Wiens, J. J., & Morrill, M. C. (2011). Missing data in phylogenetic analysis: Reconciling results from simulations and empirical data. *Systematic Biology*, *60*(5), 719–731. https://doi.org/10.1093/sysbio/syr025

Xi, Z., Liu, L., & Davis, C. C. (2016). The impact of missing data on species tree estimation. *Molecular Biology and Evolution*, *33*(3), 838–860. https://doi.org/10.1093/molbev/msv266

Zerbino, D. R. (2010). Using the Velvet de novo assembler for short-read sequencing technologies. *Current Protocols in Bioinformatics*, *31*(1), 11.5.1–11.5.12. https://doi.org/10.1002/0471250953.bi1105s31

Associate Editor: David Gower