Untangling relationships among terraranan frogs: a phylogenomic approach based on 2,665 loci

Lucas Santiago Barrientos (lucas.barrientos@gmail.com)
Universidad Del Rosario https://orcid.org/0000-0003-1971-7469
Jeffrey W. Streicher
The Natural History Museum
Elizabeth C. Miller
University of Arizona
Marcio R. Pie
Universidade Federal do Parana
John J. Wiens
University of Arizona
Andrew J. Crawford
Universidad de los Andes Facultad de Ciencias

Research article

Keywords: amphibians, anurans, missing data, phylogenomics, Terraranae

Posted Date: September 5th, 2019

DOI: https://doi.org/10.21203/rs.2.13279/v1

Untangling relationships among terraranan frogs: a phylogenomic approach based on 2,665 loci

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

Author information

Lucas S. Barrientos, Facultad de Ciencias Naturales y Matemáticas, Universidad del Rosario. Bogotá, Colombia. lucas.barrientos@urosario.edu.co

Department of Biological Sciences, Universidad de los Andes, Bogotá, código postal 1117711, Colombia. lucas.barrientos@gmail.com

Andrew J. Crawford, Department of Biological Sciences, Universidad de los Andes, Bogotá, código postal 1117711, Colombia. aj.crawford244@uniandes.edu.co

Jeffrey W. Streicher, Department of Life Sciences, The Natural History Museum, South Kensington, London SW7 5BD, England, UK.; Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721-088, USA.

j.streicher@nhm.ac.uk

John J. Wiens, Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721-088, USA. wiensj@email.arizona.edu

Elizabeth C. Miller, Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721-088, USA. ecmiller@email.arizona.edu

Marcio R. Pie, Departamento de Zoologia, Universidade Federal do Paraná, Curitiba, Paraná, 81531-980, Brazil. marcio.pie@gmail.com

Corresponding author

Correspondence to Lucas Santiago Barrientos Correa

Background

Terraranae is a large clade of New World direct-developing frogs that includes 3–5 families and >1,000 described species, encompassing ~15% of all named frog species. The relationships among major groups of terraranan frogs have been highly contentious, including conflicts among three recent phylogenomic studies utilizing 95, 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.

Results

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 both recovered strong statistical support for the following relationships among terraranan families: (Brachycephalidae, (Eleutherodactylidae, (Craugastoridae + "Strabomantidae"))). Our placement of Brachycephalidae agrees with two previous phylogenomic studies but conflicts with another. Our results place *Strabomantis* (of the Strabomantidae) with (or within) *Craugastor* (Craugastoridae) rather than with other strabomantid genera, rendering Strabomantidae paraphyletic with respect to Craugastoridae.

Conclusions

Our results suggest that Strabomantidae should be placed in the synonymy of the older Craugastoridae. Furthermore, our results suggest that Pristimantinae is paraphyletic with respect to Holoadeninae and should be subsumed into the older Holoadeninae. We also found that using matrices of UCE loci with less missing data (and concomitantly fewer loci) generally decreased support for most nodes on the tree. Overall, our results help resolve controversial relationships

within one of the largest clades of frogs, with a dataset containing ~7 times more loci than previous studies focused on this clade.

Keywords: amphibians, anurans, missing data, phylogenomics, Terraranae

Background

Terraranae (*sensu* Heinicke et al. [1]) is one of the most species-rich clades of frogs. It contains ~1,100 described species, equivalent to ~15% of all known species of anurans [2]. Terraranae is characterized by several soft anatomical characters [3] and by direct development. Direct development involves the evolutionary loss of the larval stage, such that 4-legged hatchlings emerge from fully terrestrial eggs [4]. Terraranae occurs from the southern United States to southern Brazil, including the West Indies [5,6]. Terraranan frogs are found in a variety of habitats, from deserts to rainforests and from islands to high-elevation páramo and puna [5,6]. 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 [7].

Terraranans have had a controversial taxonomy, largely because different studies support contrasting phylogenetic hypotheses (Fig. 1). Prior to 2005, traditional taxonomy placed all terraranans in the tribe Eleutherodactylini within the family Leptodactylidae [8]. Based on a parsimony analysis of nuclear and mitochondrial DNA data, Frost et al. [9] subdivided Leptodactylidae and placed all terraranans in the family Brachycephalidae. Hedges et al. [6] proposed a new phylogeny for this group of frogs, to which they gave the unranked clade name of Terrarana (latter amended to Terraranae [1,10]). They divided this clade into four families: Brachycephalidae, Craugastoridae, Eleutherodactylidae, and Strabomantidae (Fig. 1A). They showed 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. [11] added a fifth family, Ceuthomantidae, which was placed as sister to all other families (Fig. 1B). Pyron and Wiens [12] performed a large-scale maximum likelihood analysis

of 12 concatenated nuclear and mitochondrial genes for many species, which placed Strabomantidae within Craugastoridae (Fig. 1C). They also found Ceuthomantidae as sister to other terraranans, and Brachycephalidae as sister to Eleutherodactylidae + Craugastoridae. Padial et al. [13] added two more nuclear and three more mitochondrial genes to the previous data sets, and analysed relationships using dynamic homology [14,15] with parsimony. This analysis (Fig. 1D) placed Eleutherodactylidae as the sister taxon to Brachycephalidae + Craugastoridae (including Strabomantidae), and placed Ceuthomantidae within Craugastoridae (including Strabomantidae). Pinto-Sánchez et al. [7] combined data from two previous studies [12,16] to analyze relationships in terraranan frogs. They found weak support for placing Brachycephalidae as the sister taxon to Eleutherodactylidae (Fig. 1E), unlike Pyron and Wiens [12]. Pyron [17] added published data to the matrix of Pyron and Wiens [12], and also found weak support for placing Eleutherodactylidae with Brachycephalidae instead of Craugastoridae (including Strabomantidae; Fig. 1F). Feng et al. [18] analysed 95 nuclear loci to infer relationships among anurans and found support for the relationships: (Ceuthomantidae (Eleutherodactylidae (Craugastoridae + Strabomantidae))) (Fig. 1G). Hutter et al. [19] performed a Bayesian analysis of 158 hyloid genera including 13 nuclear and 7 mitochondrial genes that showed strong support (Fig. 1H) for the relationships: (Ceuthomantidae, (Brachycephalidae, (Craugastoridae [including Strabomantidae] + Eleutherodactylidae))). Heinicke et al. [1] analysed 389 nuclear loci and placed Eleutherodactylidae as the sister taxon to a clade including Brachycephalidae and Craugastoridae + Strabomantidae (Fig. 1I), again with Ceuthomantidae as sister to all other terraranans. Finally, Streicher et al. [20] 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).

101 The results from these studies can be summarized as follows. There are three main
102 competing hypotheses regarding relationships among terraranan families: (1) Eleutherodactylidae
103 as the sister taxon to Brachycephalidae + Craugastoridae + Strabomantidae [1,6,11,13]; (2)
104 Brachycephalidae as the sister taxon to Eleutherodactylidae + (Craugastoridae + Strabomantidae)
105 [12,18,19,20]; and (3) Brachycephalidae + Eleutherodactylidae as sister to Craugastoridae +
106 Strabomantidae [7,17]. Another difference among studies is that the genus *Ceuthomantis* was
107 assigned to its own family by Heinicke et al. [11], and this was supported in subsequent
108 phylogenies given its placement as sister to all other terraranan families [12,17,18]. However,
109 Padial et al. [13] placed *Ceuthomantis* inside Craugastoridae.
110 What might explain these disagreements? With the exception of Feng et al. [18], Heinicke
111 et al. [1], and Streicher et al. [20], all previous molecular phylogenetic studies were based on
112 fewer than 10 mitochondrial and 14 nuclear loci, and often had relatively weak support for the
113 key, conflicting relationships. Most studies differed in their sampling of genes and taxa (Table 1),
114 which might have contributed to the conflicting results. Another potential cause of conflict
115 among studies is the different inference methods used (Table 1), including parsimony (with a
116 dynamic optimization criterion), maximum likelihood and Bayesian analysis of concatenated
117 data, and coalescent-based species-tree analyses. Even the three recent phylogenomic studies that
118 included multiple terraranan samples showed conflicting relationships among taxonomic families
119 (Fig. 1), despite having 97, 389, and 2214 loci (Table 1).
120 In recent years, a new class of molecular markers for phylogenomic studies has been
121 developed, based on ultraconserved genomic elements (UCEs) [21]. Because of their conserved
122 nature, researchers are able to enrich and capture DNA sequences from thousands of nuclear loci,
123 even from distantly related taxa [22]. UCEs have been used to address relationships within many

vertebrate clades, including major groups of reptiles [23] and among families of fishes [24], frogs [20], lizards [25,26], and snakes [27]. UCEs have also been applied to phylogenetic questions at lower taxonomic levels, such as among bird genera [28] and among species of birds and frogs [29,30,31].

A potential disadvantage of UCEs is that the data matrices generated may include considerable missing data. Specifically, including more UCE loci in a given dataset typically requires increasing levels of missing data. Using UCE data from iguanian lizards, Streicher et al. [25] examined the impacts of including different levels of missing data and different numbers of loci on the performance of concatenated and species-tree analyses. These 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 (up to 50% missing taxa per locus). However, these authors only examined data matrices containing 20−60% missing taxa per locus. Here, we use data from thousands of UCE loci to analyze 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 10−90% missing taxa per locus).

Methods

Taxon sampling

Taxon sampling included 16 ingroup species of Terraranae encompassing representatives of four of the five terraranan families (Table 2). We follow the standard taxonomy of AmphibiaWeb [32] here and throughout the paper. We also included five outgroup species. Data for four of the ingroup species and all five outgroup species were taken from Streicher et al. [20]. From the

147 family Eleutherodactylidae, we included samples from both subfamilies (Eleutherodactylinae and
148 Phyzelaphryninae) and all four genera (*Eleutherodactylus, Diasporus, Phyzelaphryne,* and
149 *Adelophryne*). Brachycephalidae contains only two genera, *Brachycephalus* and *Ischnocnema*,
150 and we included the former. From Craugastoridae, Craugastorinae was represented by one of the
151 two genera (*Craugastor*). Within *Craugastor* our samples represented the subgenera
152 *Campbellius, Craugastor,* and *Hylactophryne*. From Strabomantidae we sampled four of the
153 eighteen genera: *Barycholos* from the subfamily Holoadeninae*,* and representatives of *Lynchius,*
154 *Pristimantis*, and *Oreobates* from the subfamily Pristimantinae (Table 2). We lacked a sample of
155 *Ceuthomantis*, which has been considered a distinct family [1,11,12,17,18]. We also lacked
156 *Hypodactylus,* which has been considered a monogeneric subfamily of Strabomantidae [1].
157 However, our taxon sampling is able to address a major debate among studies of terraranan
158 phylogeny: the relationships among Brachycephalidae, Craugastoridae, Eleutherodactylidae, and
159 Strabomantidae.

160 Many previous phylogenetic studies have demonstrated that Terraranae is nested within
161 Hyloidea [9,12,17,18,20]. However, relationships among hyloid families have generally been
162 only weakly supported in most previous studies. For outgroups, we included representatives of
163 five hyloid families (Centrolenidae, Dendrobatidae, Hemiphractidae, Hylidae, Leptodactylidae).
164 Based on the well-supported tree of Streicher et al. [20], the sister group to Terraranae includes
165 Centrolenidae, Dendrobatidae, and Leptodactylidae. These (along with additional families) form
166 the clade Commutabirana, and the sister group to Commutabirana includes Hemiphractidae and
167 Hylidae (members of the clade Amazorana).

168 All samples used in this study were donated by established natural history museums and
169 herpetological collections. Specifically, samples were provided by the Círculo Herpetológico de

Panamá (CH), Museo de Historia Natural ANDES 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. [33] (available at http://ultraconserved.org), with the modifications used by Streicher et al. [25]. 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. [25]). 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 individuals (not all individuals were included in the present study). We performed 600-cycle paired-end 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. [33] 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 [34,35]. We assembled contigs *de novo* for each sample using Velvet 1.2.10 [36] 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 [37] (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 [38] with default settings to align the resulting UCEs. MAFFT was selected because it has been shown to achieve highly accurate multiple sequence alignments relative to computational costs [39].

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 [40]. 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 [41], 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 [42]. We ran all NJst analyses using the Species Tree Analysis Web (STRAW) Server [43]. As a second species-tree approach, we also used the Accurate Species TRee ALgorithm (ASTRAL-II 5.5.9) [44,45]. 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 [42]. Species trees were rooted using the outgroups.

Missing data

We also evaluated the effect of varying the number of loci included and the concomitant amount of missing data included (by changing the maximum amount 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 missing data per locus (10–90%, in increments of 10). For example, the 10% matrix included only those loci that had data for 90% or more of all taxa. The overall amount of missing data in a matrix could be quite different from the maximum amount of missing taxa allowed per locus. For example, allowing 60% to 90% missing taxa per locus generates matrices with only 49–58% missing data overall (Table 3). We estimated the overall number of parsimony-informative sites and the amount of missing data in each dataset (Table 3) using Geneious pro 8.1 (Biomatters, http://www.geneious.com/) and Perl scripts available from http://phyluce.readthedocs.io/en/latest/tutorial-one.html. We generated alignment statistics using the script get_align_summary_data.py implemented in alignment_assessment_v1 [46].

Following Streicher et al. [20,25], we used the mean bootstrap value of all nodes across the tree to evaluate how missing data and 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 score for any particular clade. However, previous studies [20,25] 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 [47].

Results

Phylogenetic results

Using concatenated maximum likelihood analysis, the tree inferred from the matrix allowing up to 80% missing data per locus (Fig. 2) provided the overall highest mean bootstrap support (bs) values across the tree (Table 4). We describe this tree and then compare it with those obtained based on other matrices and other inference methods. Terrarananans formed a monophyletic group with 100% bootstrap support, confirming previous phylogenetic results [6,9,11–13,18–20,37]. In our results, Brachycephalidae, represented by *Brachycephalus quiririensis*, was recovered as the sister taxon to a strongly supported clade (bs=100%) uniting all other sampled terraranan taxa. Eleutherodactylidae was also strongly supported as monophyletic (bs=100%). Within Eleutherodactylidae, the subfamilies Eleutherodactylinae (*Diasporus, Eleutherodactylus*) and Phyzelaphryninae (*Adelophryne, Phyzelaphryne*) were each strongly supported as monophyletic (bs=100%). Our results show strong support (bs=100%) for a clade uniting Craugastoridae (in our phylogeny represented by *Craugastor augusti*, *C. daryi,* and *C. longirostris*) and Strabomantidae (*sensu* AmphibiaWeb) [1,32] 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 (again following current taxonomy; [32]). There is moderate support (bs=73%) for a 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 *Campbellius* (represented by *C. daryi*) was also well supported (bs=93%). There was a strongly supported clade (bs=100%) uniting *Lynchius, Oreobates*, *Pristimantis,* and *Barycholos*. These genera are currently assigned to the subfamilies Pristimantinae (*Lynchius, Oreobates*, and *Pristimantis*) and Holoadeninae (*Barycholos*) of

285 Strabomantidae. The genera *Lynchius* and *Oreobates* together formed the sister clade (bs=100%)
286 to a clade including *Barycholos* and *Pristimantis* (bs =100%; Fig. 2). Thus, Pristimantinae was
287 paraphyletic with respect to Holoadeninae, with strong support for the relevant relationships.
288 Coalescent-based species-tree analyses using NJst and ASTRAL-II applied to the 80%
289 missing-data matrix yielded phylogenetic trees that generally agreed with each other and with the
290 concatenated ML results (see above) in terms of relationships among families and among
291 subfamilies (Figs. 2 and Additional file 1: Figure S1). We compare the statistical support
292 provided by these two methods, and then highlight the one topological difference with respect to
293 the concatenated ML results. NJst and ASTRAL-II provided strong support for the monophyly of
294 Terraranae (bs=100% for both methods), and moderately strong support for the clade uniting
295 Eleutherodactylidae, Craugastoridae, and Strabomantidae to the exclusion of Brachycephalidae
296 (NJst: bs=92%: ASTRAL-II: bs=85%). Support was strong for monophyly of
297 Eleutherodactylidae (NJst: bs=96%; ASTRAL-II: bs=100%), Eleutherodactylinae (NJst:
298 bs=99%; ASTRAL-II: bs=100%), and Phyzelaphryninae (NJst: bs=98%; ASTRAL-II:
299 bs=100%). Support for the monophyly of the clade uniting Craugastoridae and Strabomantidae
300 was moderately strong for NJst and very strong for ASTRAL-II (NJst: bs=85%; ASTRAL-II:
301 bs=96%). Craugastoridae was again placed inside Strabomantidae, rendering Strabomantidae
302 paraphyletic. However, support was moderate for the clade including *Strabomantis* and
303 Craugastoridae (NJst: bs=89%; ASTRAL-II: bs=70%). Relationships among *Barycholos,*
304 *Lynchius, Oreobates*, and *Pristimantis* were the same with NJst and ASTRAL-II as with
305 concatenated ML, and remained strongly supported (bs=94–100%). Importantly, *Barycholos*
306 (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 major topological difference we observed among methods involved relationships among craugastorids and *Strabomantis* (see Additional file 1: Figure S1). The concatenated ML tree placed *Strabomantis* as sister to *Craugastor*. In contrast, the NJst and ASTRAL-II trees placed *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 analyses, whereas only Strabomantidae was paraphyletic in the ML analyses. The different placements 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 missing data were allowed per locus (from 10% up to 90%), the number of UCE loci included and total sequence length increased (Table 3). However, the range of variation in the overall amount of missing data was relatively limited (no more than 58% missing data overall). The number of parsimony-informative sites included also increased with the number of missing bases (Table 3; Spearman's rank correlation, $r_s = 0.98$; $P < 0.0001$). This may explain why bootstrap support values increased when higher levels of missing data were allowed. There were also strong correlations between the number of loci included and mean 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 $r_s = 0.88$, $P = 0.0017$).

329 The phylogenetic results described above are based on the data matrix with up to 80%
330 missing data per locus. With matrices with 50% to 90% missing data per locus (44–58% missing
331 data overall), all analyses (ML, NJst, and ASTRAL-II; Additional files 1–4: Figures S1–S4)
332 supported the monophyly of Terraranae, Eleutherodactylidae, and Strabomantidae +
333 Craugastoridae. All of these analyses also placed Brachycephalidae as the sister taxon to
334 Eleutherodactylidae + Craugastoridae + Strabomantidae. Importantly, these trees also placed
335 *Strabomantis* with or within Craugastoridae (rendering Strabomantidae paraphyletic) and
336 *Barycholos* with *Pristimantis* (rendering Pristimantinae paraphyletic). The number of loci
337 included varied from 1,262 when allowing 50% missing taxa per locus to 2,745 UCEs when
338 allowing 90% missing taxa per locus (Table 3).
339 The phylogenetic results were more variable when allowing ≤40% missing taxa per locus.
340 When loci were included with a maximum of 40% missing taxa, the data matrix included only
341 632 loci, and the ASTRAL-II analysis placed *Brachycephalus quiririensis* inside the clade of
342 Craugastoridae + Strabomantidae (bs=7%; see Supplementary Fig. S4E online), whereas ML and
343 NJst topologies were unaffected. The concatenated ML and NJst analyses based on matrices with
344 40% or more missing data placed Brachycephalidae as sister to Eleutherodactylidae +
345 Craugastoridae + Strabomantidae. Thus, the ASTRAL-II analyses were more sensitive than NJst
346 and ML to the limited number of loci. The matrix allowing up to 30% missing taxa included only
347 202 loci. Mean bootstrap values dropped precipitously across the three inference methods (Table
348 4). In the data matrices allowing only 10 or 20% missing taxa per locus, the total number of loci
349 was very low (4 and 22 UCEs, respectively), and the ML, NJst, and ASTRAL-II analyses did not
350 recover the monophyly of Eleutherodactylidae or Craugastoridae (Additional file 2–4: Figures

S2–S4). The topologies were incongruent with many previously hypothesized relationships, including monophyly of Terraranae (Additional file 2–4: Figures S2–S4).

Discussion

Previous studies showed considerable uncertainty regarding the higher-level relationships among terraranan frogs, with many conflicting trees and weak support values (Fig. 1). Based on concatenated ML and species-tree analyses, our phylogenomic dataset of 2,665 loci provides a well-supported hypothesis that may resolve much of this uncertainty (Fig. 2). The 80% matrix (including loci with up to 80% missing taxa per locus) provided the overall best-supported phylogeny for two of the three methods used (ML, ASTRAL-II; Table 4), and the inferred relationships were mostly consistent regardless of the inference method employed (Fig. 2, Additional file 1–4: Figures S1–S4 online).

The family-level phylogenetic relationships recovered here were most similar to those of Pyron and Wiens [12], Feng et al. [18], and Streicher et al. [20]. Specifically, we placed the family 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 Pyron and Wiens [12], based on concatenated ML analysis using a dataset of 3 nuclear and 9 mitochondrial genes from 340 terraranan species. Using a dataset of 95 nuclear genes from 16 terraranan species, Feng et al. [18] also placed Brachycephalidae as the 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 [26]. Our results contradict Pyron [17], who used 3 nuclear and 9 mitochondrial

loci with 418 species, and placed Brachycephalidae + Eleutherodactylidae as sister to Craugastoridae + Strabomantidae. The same topology was also supported by Pinto-Sánchez et al. [7] and Hutter et al. [19]. Our findings also contradict Heinicke et al. [1] who placed Eleutherodactylidae as sister to Brachycephalidae + Craugastoridae + Strabomantidae (see also Padial et al. [13]). Heinicke et al. [1] used a dataset of 389 genes for 30 terraranan species. However, they found only weak support from a species-tree method for most relationships, including the clade excluding Eleutherodactylidae (Fig. 1I). Among all these studies, our results are based on the largest number of loci. Our results for these family-level relationships are generally well-supported by both concatenated analyses and coalescent-based species-tree analyses.

We do not think that the systematic relationships recovered here are spurious due to limited taxon sampling. In fact, our family-level results are consistent with analyses based on 340 species [12]. Furthermore, analyses based on >300 species show considerable conflict, with some placing Brachycephalidae with Eleutherodactylidae, others placing Eleutherodactylidae as sister to Brachycephalidae, Craugastoridae, and Strabomantidae, and others placing Eleutherodactylidae with Craugastoridae and Strabomantidae (Fig. 1; Table 1). There are similar conflicts among studies with more limited taxon sampling (e.g. [1] vs. [18]). Furthermore, limited taxon sampling would also be grounds to dismiss the results of Heinicke et al. [1], which are based on a similar level of taxon sampling but much weaker sampling of genes. Thus, we do not believe that conflicts among previous studies are due to limited taxon sampling, or that there is a consistent bias associated with trees with fewer taxa. Instead, a recurring theme is that the conflicting relationships often have relatively weak branch support. Using UCE loci, we recover strong support for nodes that were contentious or poorly supported in previous studies.

Another conflict among recent molecular phylogenetic studies of Terraranae is the placement of *Strabomantis* and the recognition of the family Strabomantidae. We found that Strabomantidae is paraphyletic with respect to Craugastoridae, as also found by Pyron and Wiens [12] and Hutter et al. [19]. Again, our results are not simply an artefact of limited taxon sampling, given that the study with the most extensive taxon sampling [19] found a congruent result. Following Article 23.1 of the International Code of Zoological Nomenclature (ICZN, 1999), we suggest synonymizing Strabomantidae with Craugastoridae, given that Craugastoridae is the (slightly) older name (described on page 3 vs. page 5 of Hedges et al. [6]). Within Craugastoridae, Pyron and Wiens [12] recognized the monogeneric subfamily Strabomantinae, which is consistent with our ML results, but conflicts with our species-tree results, which place *Strabomantis* within *Craugastor* (Additional files 1, 3, 4: Figures S1, S3, S4). While its exact placement is uncertain, we did not support placing *Strabomantis* with other members of Strabomantidae. Therefore, we recommend that Craugastoridae should be expanded to include *Strabomantis* and all other genera of Strabomantidae. In contrast, Heinicke et al. [1] recognized Strabomantidae as the sister taxon of Craugastoridae. Within Strabomantidae, they placed *Strabomantis* as the sister taxon of a clade including *Barycholos*, *Oreobates*, and *Pristimantis*, among other genera. Feng et al. [18] also recognized Strabomantidae, placing *Strabomantis*, *Pristimantis*, *Hypodactylus*, and *Barycholos* in a clade that was the sister taxon of *Craugastor*. In our concatenated ML tree, *Strabomantis* is placed as the sister taxon of *Craugastor* (bs=73%), rather than with the other sampled genera of Strabomantidae (*Barycholos, Lynchius, Oreobates, Pristimantis*). Both species-tree methods placed *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, we support Pyron and Wiens [12],

Gomez-Mestre et al. [49], and others in placing Strabomantidae within Craugastoridae. We note that this taxonomy would still be consistent with the phylogeny even if we were wrong about the placement of *Strabomantis* and some previous hypotheses were correct instead (as long as Craugastoridae and "Strabomantidae" are sister taxa). Importantly, we recognize that our data were relatively limited for *Strabomantis* (136 loci), despite the large number of loci overall. Finally, we note that there is little basis for saying that Strabomantidae must be recognized on the grounds of taxonomic stability. Both Craugastoridae and Strabomantidae have only been recognized very recently, and many previous authors considered Strabomantidae to be part of Craugastoridae [7,12,17,19,20,49].

The relationships among genera of Holoadeninae and Pristimantinae also differ between our study and previous studies. Here, were 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). In contrast, among previous studies that sampled these four genera (along with many other genera), Hedges et al. [6], Heinicke et al. [1], Hutter et al. [19], Pyron [17], and Pyron and Wiens [12] all recovered the following relationships (*Barycholos*, (*Pristimantis*, (*Lynchius*, *Oreobates*))). On the other hand, Padial et al. [13] 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 more extensive in most previous studies, despite their more limited sampling of genes [1,12,17,19]. Our results placing Holoadeninae (*Barycholos*) inside Pristimantinae (*Lynchius*, *Pristimantis*, and *Oreobates*) suggest that one of these two subfamilies is not valid. We note that the name Holoadeninae proposed by Hedges et al. [6] is available and

443 has priority over Pristimantinae [12], according to Article 23.1 of ICZN [50]. Therefore, we
444 recommend expanding Holoadeninae to encompass Pristimantinae. Again, we note that this
445 taxonomy would still be consistent with the phylogeny even if previous hypotheses were correct
446 about the placement of *Barycholos*. Thus, we propose that the family Craugastoridae should
447 contain the subfamilies Craugastorinae (containing the genera *Craugastor* and *Haddadus*),
448 Holoadeninae (containing *Barycholos, Bryophryne, Dischidodactylus, Euparkerella, Holoaden,*
449 *Lynchius, Microkayla, Niceforonia, Noblella, Oreobates, Phrynopus, Pristimantis,*
450 *Psychophrynella,* and *Yunganastes*), Hypodactylinae (containing *Hypodactylus*), and
451 Strabomantinae (containing *Strabomantis*).
452 In summary, our analyses using concatenated ML analysis and coalescent-based species-
453 tree methods (NJst and ASTRAL-II) resulted in high support for Brachycephalidae as the sister
454 taxon to the clade Eleutherodactylidae + Craugastoridae + Strabomantidae, as also found by
455 Pyron and Wiens [12], Feng et al. [18], and Streicher et al. [20], but *contra* Heinicke et al.
456 [1,11,36], Hedges et al. [52], Padial et al. [13], Pinto-Sánchez et al. [7], and Pyron [17]. Our
457 results also show that Craugastoridae is nested inside of Strabomantidae (suggesting that
458 Strabomantidae should be placed in the older Craugastoridae). We also find that Holoadeninae is
459 nested inside Pristimantinae, and that Pristimantinae should be synonymized with Holoadeninae.
460 Our results may help resolve some controversial issues in terraranan phylogenetics, but
461 much additional work is needed. We suggest that the highest priority for future studies will be to
462 include more taxa. For example, because we had no samples of *Ceuthomantis* in this study, we
463 could not evaluate the hypothesis that this genus is sister to all other Terrananae. Nevertheless,
464 this hypothesis has been consistently supported by all other recent studies [1,11,12,17,18], except
465 Padial et al. [13]. The possible placement of *Strabomantis* within *Craugastor* should also be

addressed with better taxon sampling from both genera, and additional taxonomic changes may be needed if the results from our species-tree analyses are supported. It will also be important to include several genera that have not yet been included in molecular phylogenetic analyses (i.e., *Atopophrynus, Dischidodactylus*, *Geobatrachus*, and *Niceforonia*).

Missing data

A major concern in phylogenomic analyses is the possible impact of missing data on phylogenetic inference [25,53,54,55,56,57]. Here, we found that branch support values from concatenated, ASTRAL-II, and NJst analyses increased in datasets containing more, not fewer, missing data (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 may occur because including more missing data per locus allows for the inclusion of more sites, contributing to concatenated analyses, and more loci, contributing to coalescent analyses (Table 3). Many simulation and empirical studies have now shown that increasing the number of genes (for species-tree methods) and characters (for concatenated analyses) may increase the accuracy of phylogenetic analyses, despite the increase in missing data [25,38,41,42].

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 most terraranan families and subfamilies based on concatenated and species-tree methods, including the largest number of genetic loci considered so far. At the family level, our results show that Brachycephalidae is the

sister taxon of Eleutherodactylidae + Craugastoridae + Strabomantidae. This result conflicts with some previous studies [1,6,7,13,17,19] but is congruent with others [12,18,20]. Within these families, we find that Strabomantidae is paraphyletic with respect to Craugastoridae, and Pristimantinae is paraphyletic with respect to Holoadeninae. Given these results, we recommend synonymizing Strabomantidae with the older Craugastoridae and Pristimantinae with the older Holoadeninae. Note that these changes would make the taxonomy consistent with the phylogeny even if most previous hypotheses were correct instead. Finally, we explored the effect of including loci with progressively more missing data on three phylogenetic methods (concatenated likelihood, and two species-tree methods, NJst and ASTRAL-II). We found that including loci with more missing data generally increased mean bootstrap values for all three methods, given that including more missing data allowed more loci to be included. Our results add more support to the idea that the benefits of including more loci can potentially overcome the negative consequences of including more missing data in phylogenomic analyses.

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

Declarations

Acknowledgements

We are grateful for tissue loans from Círculo Herpetológico de Panamá (CH), Museo de Historia Natural ANDES 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), and the Museo de Herpetología de la Universidad de Antioquia (MHUA). We thank Andres M. Cuervo, Luisa Dueñas, Carlos E. Guarnizo, Santiago Herrera, Andrea Paz, and Camila Plata 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, to Jorge Avendaño for his advice with the figures.

Funding

Funding for this project came from Colciencias grant number 567 for PhD studies and *Crédito Condonable* of the Universidad de los Andes to LSB, the FAPA fund of the Universidad de los Andes to AJC, and the University of Arizona and U.S. National Science Foundation grant DEB 1655690 to JJW.

Availability of data and materials

The genetic data alignments datasets supporting the results of this article are available in the Dryad repository (XXXX)

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 re-writing 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.

Consent for publication

All authors give consent for their work to be published in this paper.

Competing interests

The authors declare no competing interests.

References

1. Heinicke, M. P., Lemmon, A. R., Moriarty, E., Mcgrath, K. & Hedges, S. B. Phylogenomic support for evolutionary relationships of New World direct-developing frogs (Anura: Terraranae). Mol. Phylogenet. Evol. 2018;118:145–155.

2. Frost, D. R. Amphibian species of the world: An online reference. Version 6.0. Am. Mus. Nat. Hist., New York, *USA.* 2017; Available from: http://research.amnh.org/herpetology/amphibia/index.html

3. Taboada, C., Grant, T., Lynch, J. D. & Faivovich, J. New morphological synapomorphies for the New World direct-developing frogs (Amphibia: Anura: Terrarana). Herpetologica 2013;69:342–357.

4. Duellman, W. E. & Trueb, L. *Biology of amphibians*. The Johns Hopkins University Press: Baltimore; 1994.

5. Gonzalez-Voyer, A., Padial, J. M., Castroviejo-Fisher, S., de la Riva, I. & Vilà, C. Correlates of species richness in the largest Neotropical amphibian radiation. *J. Evol. Biol.* 2011;24:931–942.

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

7. Pinto-Sánchez, N. R., Crawford, A. J. & Wiens, J. J. Using historical biogeography to test for community saturation. Ecol. Lett. 2014;17:1077–1085.

8. Lynch, J. D. Evolutionary relationships, osteology, and zoogeography of leptodacyloid frogs. Univ. Kansas Mus. Nat. Hist. Misc. Publ. 1971;53:1–238.

9. Frost, D. R. *et al.* The amphibian tree of life. Bull. Am. Mus. Nat. Hist. 2006; 297:1–377.

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

11. Heinicke, M. P. *et al.* A new frog family (Anura : Terrarana) from South America and an

expanded direct-developing clade revealed by molecular phylogeny. Zootaxa. 2009;35:1–35.

12. Pyron, R. A. & Wiens, J. J. A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. Mol. Phylogenet. Evol. 2011;61:543–83.

13. Padial, J. M., Grant, T. & Frost, D. R. Molecular systematics of terraranas (Anura: Brachycephaloidea) with an assessment of the effects of alignment and optimality criteria. Zootaxa. 2014;3825:1–132.

14. Wheeler, W. Fixed character states and the optimization of molecular sequence data. Cladistics. 1999;15:379–385.

15. Wheeler, W. Homology and the optimization of DNA sequence data. Cladistics. 2001;17: S3–S11.

16. Pinto-Sánchez, N. R. *et al.* The Great American Biotic Interchange in frogs: Multiple and early colonization of Central America by the South American genus *Pristimantis* (Anura : Craugastoridae). Mol. Phylogenet. Evol. 2012;62:954–972.

17. Pyron, R. A. Biogeographic analysis reveals ancient continental vicariance and recent oceanic dispersal in amphibians. Syst. Biol. 2014;63:779–797.

18. Feng, Y.-J. *et al.* Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous–Paleogene boundary. Proc. Natl. Acad. Sci. USA. 2017;114:E5864–E5870.

19. Hutter, C. R., Lambert, S. M. & Wiens, J. J. Rapid diversification and time explain amphibian richness at different scales in the tropical Andes, Earth's most biodiverse hotspot. Am. Nat. 2017;190:828–843.

20. Streicher, J. W. *et al.* Evaluating methods for phylogenomic analyses, and a new phylogeny for a major frog clade (Hyloidea) based on 2,214 loci. Mol. Phylogenet. Evol. 2018;119:128–143.
21. Bejerano, G. *et al.* Ultraconserved elements in the human genome. Science. 2004;304: 1321–1325.
22. McCormack, J. E. *et al.* Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. Mol. Phylogenet. Evol. 2012;62:397–406.
23. Crawford, N. G. *et al.* A phylogenomic analysis of turtles. Mol. Phylogenet. Evol. 2015;83:250–257.
24. Faircloth, B. C., Sorenson, L., Santini, F. & Alfaro, M. E. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). PLoS ONE. 2013;8:e65923.
25. Streicher, J. W., Schulte, J. A. & Wiens, J. J. How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in iguanian lizards. Syst. Biol. 2016;65:128–145.
26. Streicher, J. W. & Wiens, J. J. Phylogenomic analyses of more than 4000 nuclear loci resolve the origin of snakes among lizard families. Biol. Lett. 2017;13:1–4.
27. Streicher, J. W. & Wiens, J. J. Phylogenomic analyses reveal novel relationships among snake families. Mol. Phylogenet. Evol.2016*;*100:160–169.
28. Meiklejohn, K. A., Faircloth, B. C., Glenn, T. C., Kimball, R. T. & Braun, E. L. Analysis of a rapid evolutionary radiation using ultraconserved elements: Evidence for a bias in some multispecies coalescent methods. Syst. Biol. 2016;65:612–627.
29. Alexander, A. M. *et al.* Genomic data reveals potential for hybridization, introgression,

and incomplete lineage sorting to confound phylogenetic relationships in an adaptive radiation of narrow-mouth frogs. Evolution 2016;71:475–488.

30. McCormack, J. E., Tsai, W. L. E. & Faircloth, B. C. Sequence capture of ultraconserved elements from bird museum specimens. Mol. Ecol. Res. 2016;16:1189–1203.

31. Smith, B. T., Harvey, M. G., Faircloth, B. C., Glenn, T. C. & Brumfield, R. T. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Syst. Biol. 2014:63:83–95.

32. University of California, Berkeley, CA. *AmphibiaWeb*. 2018; Available from https://amphibiaweb.org.

33. Faircloth, B. C. *et al.* Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 2012:61:717–726.

34. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30;2114–2120.

35. Faircloth, B. C. illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. 2013. Available from https://doi.org/10.6079/J9ILL.

36. Zerbino, D. R. Using the Velvet de novo assembler for short-read sequencing technologies. Curr. Prot. Bioinformatics. 2010;31:11.5.1-11.5.12.

37. Faircloth, B. C. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics. 2015;32;786–788.

38. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30: 3059–3066.

39. Pais, F. S.-M., Ruy, P. de C., Oliveira, G. & Coimbra, R. S. Assessing the efficiency of

multiple sequence alignment programs. Algorithms Mol. Biol. 2014;9:4.

40. Stamatakis, A. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–1315.

41. Liu, L. & Yu, L. Estimating species trees from unrooted gene trees. Syst. Biol. 2011;60: 661–667.

42. Seo, T. K. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. Algorithms Mol. Biol. 2008;**25**;960–971.

43. Shaw, T. I., Ruan, Z., Glenn, T. C. & Liu, L. STRAW: Species TRee Analysis Web server. Nucleic Acids Res. 2013;41:W238-41.

44. Mirarab, S. *et al.* ASTRAL: Genome-scale coalescent-based species tree estimation. Bioinformatics. 2014;30:541–548.

45. Mirarab, S. & Warnow, T. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics. 2015;31:i44–i52.

46. Portik, D. M., Smith, L. L. & Bi, K. An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). Mol. Ecol. Res. 2016;**16**:1069–1083.

47. Hillis, D. M. & Bull, J. An empirical test of bootstraping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 1993;42:182–192.

48. Crawford, A. J. & Smith, E. N. Cenozoic biogeography and evolution in direct-developing frogs of Central America (Leptodactylidae: *Eleutherodactylus*) as inferred from a phylogenetic analysis of nuclear and mitochondrial genes. Mol. Phylogenet. Evol. 2005;35:536–555 (2005).

49. Gomez-Mestre, I., Pyron, R. A. & Wiens, J. J. Phylogenetic analyses reveal unexpected

patterns in the evolution of reproductive modes in frogs. Evolution. 2012;66:3687–3700.

50. International Commission for Zoological Nomenclature. International code of zoological nomenclature, 4th ed. (ICZN), London. 1999.

51. Heinicke, M. P., Duellman, W. E. & Hedges, S. B. Major Caribbean and Central American frog faunas originated by ancient oceanic dispersal. Proc. Natl. Acad. Sci. *USA* 2007;104:9913–10294.

52. Hedges, S. B., Duellman, W. E. & Heinicke, M. P. A replacement name for *Isodactylus* Hedges, Duellman, and Heinicke, 2008. Zootaxa. 2008;68:67–68.

53. Jiang, W., Chen, S. Y., Wang, H., Li, D. Z. & Wiens, J. J. Should genes with missing data be excluded from phylogenetic analyses? Mol. Phylogenet. Evol. 2014;80:308–318 (2014).

54. Philippe, H. *et al.* Phylogenomics of eukaryotes: Impact of missing data on large alignments. Mol. Phylogenet. Evol. 2004;21:1740–1752.

55. Roure, B., Baurain, D. & Philippe, H. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. Mol. Phylogenet. Evol. 2013;30:197–214.

56. Wiens, J. J. & Morrill, M. C. Missing data in phylogenetic analysis: Reconciling results from simulations and empirical data. Syst. Biol. 2011;60:719–731.

57. Xi, Z., Liu, L. & Davis, C. C. The impact of missing data on species tree estimation. *Mol.* Phylogenet. Evol. 2016;33:838–860.

695 **Table 1.** Summary of previous studies of the phylogenetic relationships within Terraranae using
696 molecular data. The abbreviation for the inference methods are: maximum parsimony (MP),
697 maximum likelihood (ML), Bayesian inference (Bayes), parsimony under direct /dynamic
698 optimization criterion (POY), and coalescent-based species-tree estimation using Accurate
699 Species TRee Algorithm (ASTRAL) and NJst.
700

Study	Terraranan taxa	Outgroup taxa	Mitochondrial genes	Nuclear genes	Inference method
Hedges et al. [6]	344	18	2	2	ML, Bayes
Heinicke et al. [11]	42	4	6	11	MP, ML, Bayes
Pyron and Wiens [12]	340	2533	3	9	ML
Padial et al. [13]	405	25	9	11	POY
Pyron [17]	418	2892	3	9	ML
Pinto-Sánchez et al. [7]	363	7	3	9	Bayes
Feng et al. [18]	16	278	--	97	ML, ASTRAL
Hutter et al. [19]	610*	1708*	7	13	Bayes
Heinicke et al. [1]	30	5	--	389	ML, ASTRAL
Streicher et al. [20]	5	45	--	2214	ML, NJst, ASTRAL

701 * Includes well-supported but undescribed species.
702

Table 2. Voucher information and amount of DNA data produced for each sample, including number of contigs assembled usi the resulting number of aligned ultraconserved elements (UCEs) obtained. Sequence Read Archive (SRA) accession numbers p obtained for each individual. Family-level taxonomy follows Heinicke et al. [1]. Accession numbers starting with SAMN were Streicher et al. [20]. Museum numbers are from the following natural history collections: Museo Herpetológico de la Universid (MHUA), Royal Ontario Museum (ROM), Museo de Historia Natural, Universidad Nacional Mayor de San Marcos (MHNSM de Panamá (CH), Museo de Historia Natural ANDES at the Universidad de los Andes in Bogotá (ANDES), the Museum of Ve University of California, Berkeley (MVZ), Amphibian and Reptile Diversity Research Center at the University of Texas at Arl Comparative Zoology at Harvard University (MCZ), and the Biodiversity Institute & Natural History Museum at Kansas Univ Herpetológica do Departamento de Zoologia, Universidade Federal do Paraná, Curitiba, state of Paraná (DZUP). Three sample numbers: Jonathan A. Campbell (JAC), Erik R. Wild (ERW), and William E. Duellman (WED).

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

716 **Table 3.** Summary of data matrices based on ultraconserved elements (UCEs), organized by the
717 maximum percentage of missing data allowed per locus in each matrix. The summary includes
718 the total number of UCE loci per matrix, total number of characters in aligned base pairs (bp),
719 number of parsimony-informative sites summed across all UCEs, total number of missing data
720 cells in each matrix, percent of sites across all loci that contain parsimony-informative variation
721 (excluding sites with gap characters) in relation to the total length of the alignment, and the
722 overall percentage of missing data cells in each matrix.
723

Maximum missing data per locus	Number of UCEs	Total DNA sequence length in bp	Parsimony-informative sites	Missing data cells	% informative sites	% missing data in matrix
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%

724
725

726 **Table 4.** Mean bootstrap support across all nodes in trees inferred from nine data matrices, each
727 allowing different amounts of missing taxa per locus (10% to 90%). Results are given for three
728 phylogenetic methods (concatenated likelihood using RAxML, and the coalescent-based species
729 tree methods NJst and ASTRAL-II).
730

Missing taxa/locus	Average bootstrap support values across all nodes		
	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

731

FIGURE LEGENDS

Figure 1. Summary of hypotheses of higher-level phylogenetic relationships among terraranan frogs. The taxonomy used in each tree here follows the taxonomy used in that study. Numbers adjacent to nodes 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. [6] concatenated maximum likelihood (ML) analysis of two mitochondrial genes and two nuclear genes. (B) Heinicke et al. [11] concatenated ML analysis of 6 mitochondrial and 11 nuclear genes. (C) Pyron and Wiens [12] concatenated ML analysis of 3 mitochondrial and 9 nuclear genes. (D) Padial et al. [13] parsimony analysis of 9 mitochondrial and 12 nuclear genes; numbers adjacent to nodes represent jackknife support. (E) Pinto-Sánchez et al. [7] concatenated Bayesian analysis of 3 mitochondrial and 9 nuclear genes; nodes without support represent Bayesian posterior probabilities <0.95. (F) Pyron [17] concatenated ML analysis of 3 mitochondrial and 9 nuclear genes. (G) Feng et al. [18] concatenated ML analysis and coalescent-based species tree (ASTRAL) of 95 nuclear protein-coding genes. (H) Hutter et al. [19] concatenated Bayesian analysis of 7 mitochondrial and 13 nuclear genes; numbers adjacent to nodes represent Bayesian posterior probabilities. (I) Heinicke et al. [1] concatenated ML analysis and coalescent-based species tree (ASTRAL) based on 389 nuclear protein-coding genes; numbers adjacent to nodes represent ML bootstrap support before the slash, followed by local posterior probabilities for ASTRAL. (J) Streicher et al. [20] concatenated ML analysis based on 2,214 UCEs, including loci with up to 60% missing taxa per UCE. Numbers next to each branch show support from concatenated analysis before the slash, followed by NJst bootstrap.

Figure 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% missing taxa. Numbers next to each node indicate bootstrap support values from concatenated ML analysis (top) and coalescent-based species-tree analyses from NJst (middle) and ASTRAL-II (bottom number). The black squares indicate the clades that were recovered in all three analyses of this data matrix. The white square indicates a node (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 placed inside *Craugastor* (see Supplementary Fig. S1 online). Note that Strabomantidae and Pristimantinae are paraphyletic in all of these trees. The full trees from each method (including branch lengths and outgroups) are provided in Additional file: Fig. S1 online. Family-level taxonomy follows AmphibiaWeb [32].

Additional file: **Figure S1.** Phylogeny of terraranan frogs based on three inference methods, using the data matrix including loci with up to 80% missing data per locus and including 2,665 ultraconserved element (UCE) loci for a total of 743,419 concatenated base pairs. The family-level taxonomy follows AmphibiaWeb [32]. (A) Maximum likelihood tree inferred from concatenated data using RAxML. (B) Coalescent-based species tree inferred using NJst. (C) Coalescent-based species tree inferred using ASTRAL-II. The numbers adjacent to the nodes represent bootstrap support (see Methods). All trees were rooted on *Stefania coxi* + *Dendropsophus leali*. Note that in Additional file: Fig. S1C online the resulting tree is inconsistent with previous results and was rooted only on *S. coxi*.

Additional file: **Figure S2.** Concatenated maximum likelihood trees inferred using RAxML, with bootstrap support values indicated next to each branch. The family-level taxonomy is indicated for trees inferred from matrices with ≥ 40% maximum missing data. Taxonomy follows AmphibiaWeb [32]. The eight trees are based on increasingly complete data matrices that have fewer UCE loci. (A) Tree inferred from a data matrix allowing up to 90% missing taxa per UCE locus, for a total of 2,745 loci; (B) allowing up to 70% missing taxa per UCE locus, including 2,368 loci; (C) allowing up to 60% missing taxa per UCE locus, including 1,906 loci; (D) allowing up to 50% missing taxa per UCE locus, including 1,262 loci; E) allowing a maximum of 40% missing taxa per UCE locus, including 632 loci; (F) allowing up to 30% missing taxa per UCE locus, including 202 loci; (G) allowing up to 20% missing taxa per UCE locus, thus including 22 loci; (F) allowing up to 10% missing taxa per UCE locus, including just 4 loci. The numbers next to each node represent bootstrap support (see Methods). Note that the tree based on the 80% missing-data matrix is shown in Additional file: Fig. S1A online. Outgroups not shown (see Additional file: Fig. S1 online), unless ingroup is non-monophyletic.

Additional file: **Figure S3.** Coalescent-based species trees inferred using NJst, with bootstrap support values indicated next to each branch. The eight trees are based on increasingly complete data matrices that have fewer UCE loci. The family-level taxonomy is indicated for trees inferred from matrices with ≥ 40% maximum missing data. Taxonomy follows AmphibiaWeb [32]. (A) Tree inferred from a data matrix allowing up to 90% missing taxa per UCE locus, including 2,745 loci; (B) allowing up to 70% missing taxa per UCE locus, and including 2,368 loci; (C) allowing up to 60% missing taxa per UCE locus, including 1,906 loci; (D) allowing up to 50%

missing taxa per UCE locus, including 1,262 loci; (E) allowing up to 40% missing taxa per UCE locus, including 632 loci; (F) allowing up to 30% missing taxa per UCE locus, including 202 loci; (G) allowing up to 20% missing taxa per UCE locus, thus including 22 loci; (F) allowing a maximum of 10% missing taxa per UCE locus, thus including only 4 loci. The numbers next to each node represent the bootstrap support (see Methods). Note that the tree based on the 80% missing-data matrix is shown in Additional file: Fig. S1B. Outgroups not shown (see Additional file: Fig. S1 online), unless ingroup is non-monophyletic.

Additional file: **Figure S4.** Coalescent-based species trees inferred using ASTRAL-II, with bootstrap support values indicated next to each branch. The eight trees are based on increasingly complete data matrices that have fewer UCE loci. The family-level taxonomy is indicated for trees inferred from matrices with ≥ 40% maximum missing data. Taxonomy follows AmphibiaWeb [32]. (A) Tree inferred from a data matrix allowing up to 90% missing taxa per UCE locus, including 2,745 loci; (B) allowing up to 70% missing taxa per UCE locus, including 2,368 loci; (C) allowing up to 60% missing taxa per UCE locus, including 1,906 loci; (D) allowing up to 50% missing taxa per UCE locus, including 1,262 loci; (E) allowing up to 40% missing taxa per UCE locus, including 632 loci; (F) allowing up to 30% missing taxa per UCE locus, including 202 loci; (G) allowing up to 20% missing taxa per UCE locus, thus including 22 loci; (F) allowing a maximum of 10% missing taxa per UCE locus, including 4 loci. The numbers next to each node represent the bootstrap support (see Methods). Note that the tree based on the 80% missing-data matrix is shown in Additional file: Fig. S1C. Outgroups not shown (see Additional file: Fig. S1 online), unless ingroup is non-monophyletic.

Figures

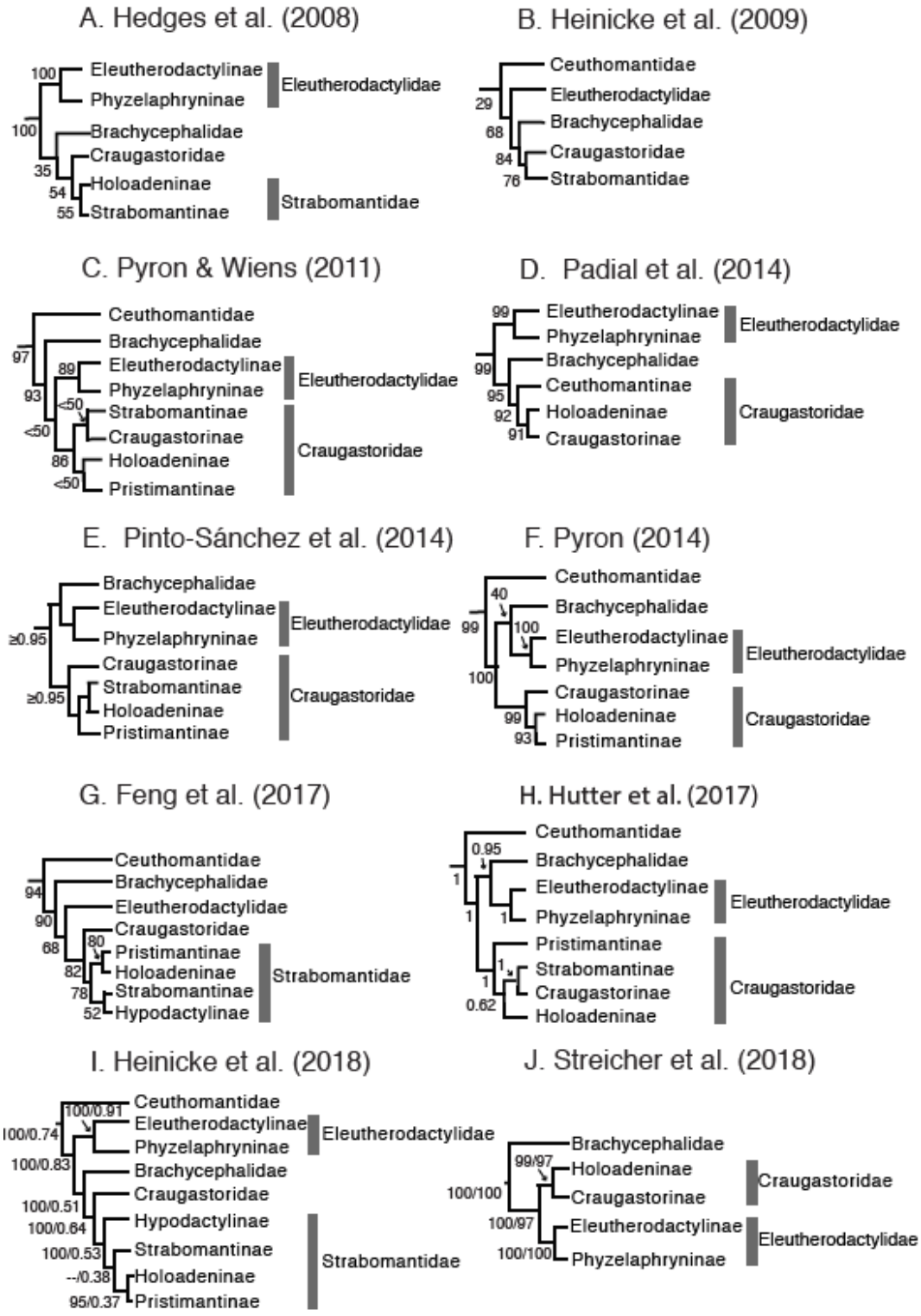

Figure 1

Summary of hypotheses of higher-level phylogenetic relationships among terraranan frogs. The taxonomy used in each tree here follows the taxonomy used in that study. Numbers adjacent to nodes 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. [6] concatenated maximum likelihood (ML) analysis of two mitochondrial genes and two nuclear genes. (B) Heinicke et al. [11] concatenated ML analysis of 6 mitochondrial and 11 nuclear genes. (C) Pyron and Wiens [12] concatenated ML analysis of 3 mitochondrial and 9 nuclear genes. (D) Padial et al. [13] parsimony analysis of 9 mitochondrial and 12 nuclear genes; numbers adjacent to nodes represent jackknife support. (E) Pinto-Sánchez et al. [7] concatenated Bayesian analysis of 3 mitochondrial and 9 nuclear genes; nodes without support represent Bayesian posterior probabilities <0.95. (F) Pyron [17] concatenated ML analysis of 3 mitochondrial and 9 nuclear genes. (G) Feng et al. [18] concatenated ML analysis and coalescentbased species tree (ASTRAL) of 95 nuclear protein-coding genes. (H) Hutter et al. [19] concatenated Bayesian analysis of 7 mitochondrial and 13 nuclear genes; numbers adjacent to nodes represent Bayesian posterior probabilities. (I) Heinicke et al. [1] concatenated ML analysis and coalescent-based species tree (ASTRAL) based on 389 nuclear protein-coding genes; numbers adjacent to nodes represent ML bootstrap support before the slash, followed by local posterior probabilities for ASTRAL. (J) Streicher et al. [20] concatenated ML analysis based on 2,214 UCEs, including loci with up to 60% missing taxa per UCE. Numbers next to each branch show support from concatenated analysis before the slash, followed by NJst bootstrap.

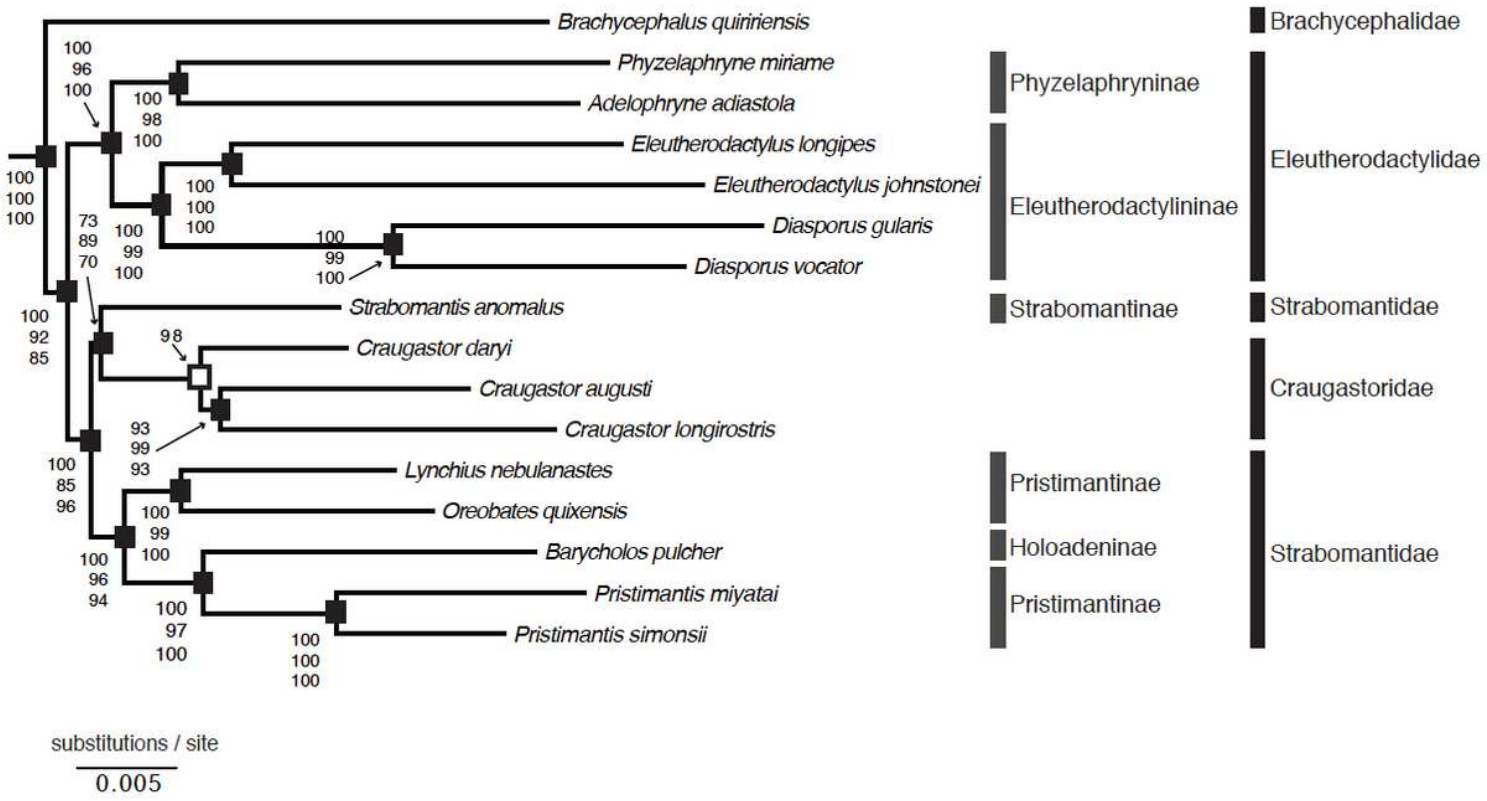

Figure 2

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% missing taxa. Numbers next to each node indicate bootstrap support values from concatenated ML analysis (top) and coalescent-based species-tree analyses from NJst (middle) and

ASTRAL-II (bottom number). The black squares indicate the clades that were recovered in all three analyses of this data matrix. The white square indicates a node (monophyly of Craugastoridae) that was not recovered in the NJst or ASTRALII analyses, and thus has only the ML bootstrap score. In the NJst and ASTRAL-II analyses of this dataset, Strabomantis is placed inside Craugastor (see Supplementary Fig. S1 online). Note that Strabomantidae and Pristimantinae are paraphyletic in all of these trees. The full trees from each method (including branch lengths and outgroups) are provided in Additional file: Fig. S1 online. Family-level taxonomy follows AmphibiaWeb [32].

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigS2.pdf
- FigS3.pdf
- FigS4.pdf
- FigS1.pdf