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
Research Article

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

LUCAS S. BARRIENTOS1, JEFFREY W. STREICHER2,3, ELIZABETH C. MILLER2,4, MARCIO R. PIE5, JOHN J. WIENS2 & ANDREW J. CRAWFORD1

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

Terraranans 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). Terraranans 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
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 as sister to other terraranans, and Brachycephalidae as sister to Eleutherodactylidae + Craugastoridae. Padial et al. (2014) added two more nuclear and three 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 recovered 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
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, Phyzelaphrynus, and Adelophrynus). 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 Hylactophrynus. From Strabomantidae we sampled four of the 18 genera, including Barycholos from the subfamily...
Phylogenetics of terraranan frogs

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.

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

*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 monogenic 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 Hylidae (Feng et al., 2017; Frost et al., 2006; Hime et al., 2021; Pyron, 2014; Pyron & Wiens, 2011; Streicher et al., 2018). However, relationships among hylid families have been only weakly supported in most previous studies. For outgroups, we included representatives of five hylid 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 Commutabiranata, 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...
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 GTR+GAMMA 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.

---

### 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).

<table>
<thead>
<tr>
<th>Species Family</th>
<th>Museum number</th>
<th>Contigs</th>
<th>UCEs</th>
<th>SRA accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brachycephalus quiririensis</td>
<td>Brachycephalidae</td>
<td>DZUP 522</td>
<td>5415</td>
<td>886</td>
</tr>
<tr>
<td>Craugastor augusti</td>
<td>Craugastoridae</td>
<td>UTA A-60654</td>
<td>3790</td>
<td>738</td>
</tr>
<tr>
<td>Craugastor daryi</td>
<td>Craugastoridae</td>
<td>UTA A-62648</td>
<td>375</td>
<td>71</td>
</tr>
<tr>
<td>Craugastor longirostris</td>
<td>Craugastoridae</td>
<td>MIUA 4809</td>
<td>2569</td>
<td>1301</td>
</tr>
<tr>
<td>Eleutherodactylus johnstonei</td>
<td>Eleutherodactylidae</td>
<td>ANDES-A 1912</td>
<td>4588</td>
<td>1792</td>
</tr>
<tr>
<td>Eleutherodactylus longipes</td>
<td>Eleutherodactylidae</td>
<td>JAC 29834</td>
<td>2710</td>
<td>1243</td>
</tr>
<tr>
<td>Diasporus gularis</td>
<td>Eleutherodactylidae</td>
<td>ANDES-A 3833</td>
<td>19214</td>
<td>427</td>
</tr>
<tr>
<td>Diasporus vocator</td>
<td>Eleutherodactylidae</td>
<td>CH 4786</td>
<td>8247</td>
<td>429</td>
</tr>
<tr>
<td>Adelophryne adiastola</td>
<td>Eleutherodactylidae</td>
<td>ANDES-A 2560</td>
<td>3680</td>
<td>1515</td>
</tr>
<tr>
<td>Phylzephyrne miriame</td>
<td>Eleutherodactylidae</td>
<td>ANDES-A 3834</td>
<td>3875</td>
<td>512</td>
</tr>
<tr>
<td>Strabomantis anomalus</td>
<td>Strabomantidae</td>
<td>ANDES-A 1416</td>
<td>1034</td>
<td>136</td>
</tr>
<tr>
<td>Barycholos pulcher</td>
<td>Strabomantidae</td>
<td>KU 217782</td>
<td>1261</td>
<td>610</td>
</tr>
<tr>
<td>Lynchius nebulanastes</td>
<td>Strabomantidae</td>
<td>ERW 86</td>
<td>11368</td>
<td>2233</td>
</tr>
<tr>
<td>Oreobates quixensis</td>
<td>Strabomantidae</td>
<td>ANDES-A 1954</td>
<td>4975</td>
<td>1198</td>
</tr>
<tr>
<td>Pristimantis simonsi</td>
<td>Strabomantidae</td>
<td>WED 56667</td>
<td>5268</td>
<td>2118</td>
</tr>
<tr>
<td>Pristimantis miyatai</td>
<td>Strabomantidae</td>
<td>ANDES-A 1776</td>
<td>4897</td>
<td>1429</td>
</tr>
<tr>
<td>Espharanus prosoblepon</td>
<td>Centrolenidae</td>
<td>MVZ 149741</td>
<td>6094</td>
<td>1851</td>
</tr>
<tr>
<td>Stefania coxi</td>
<td>Hemiphractidae</td>
<td>ROM 39478</td>
<td>3350</td>
<td>1826</td>
</tr>
<tr>
<td>Dendropsophus leali</td>
<td>Hylidae</td>
<td>KU 215259</td>
<td>6094</td>
<td>1767</td>
</tr>
<tr>
<td>Hylaxalus nixipes</td>
<td>Dendrobatidae</td>
<td>KU 211806</td>
<td>368</td>
<td>1456</td>
</tr>
<tr>
<td>Leptodactylus didymus</td>
<td>Leptodactylidae</td>
<td>MHNSM 14643</td>
<td>5814</td>
<td>1909</td>
</tr>
</tbody>
</table>
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 GTR+gamma 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%
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 (Adelophryn, Phyzelaphryn) 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 Hylactophryn (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 Holoadenninae (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 Holoadenninae (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 (Holoadenninae) was again nested inside of Pristimantinae (Lynchius, Oreobates, and Pristimantis) with strong support for the clade uniting Holoadenninae with Pristimantis (NJst: bs = 97%; ASTRAL-II: bs = 100%).

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

---

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.

<table>
<thead>
<tr>
<th>Maximum % of taxa lacking data per locus</th>
<th>Number of UCEs</th>
<th>Total DNA sequence length in bp</th>
<th>Parsimony-informative sites</th>
<th>Missing data cells</th>
<th>% informative sites</th>
<th>% missing data cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>2,745</td>
<td>754,266</td>
<td>24,079</td>
<td>9,264,413</td>
<td>3.10%</td>
<td>58%</td>
</tr>
<tr>
<td>80%</td>
<td>2,665</td>
<td>743,419</td>
<td>24,079</td>
<td>8,844,779</td>
<td>3.20%</td>
<td>57%</td>
</tr>
<tr>
<td>70%</td>
<td>2,368</td>
<td>639,195</td>
<td>22,083</td>
<td>7,126,259</td>
<td>3.50%</td>
<td>53%</td>
</tr>
<tr>
<td>60%</td>
<td>1,906</td>
<td>503,502</td>
<td>17,390</td>
<td>5,166,952</td>
<td>3.50%</td>
<td>49%</td>
</tr>
<tr>
<td>50%</td>
<td>1,262</td>
<td>377,741</td>
<td>11,217</td>
<td>3,097,704</td>
<td>3.30%</td>
<td>44%</td>
</tr>
<tr>
<td>40%</td>
<td>632</td>
<td>172,052</td>
<td>5,513</td>
<td>1,359,560</td>
<td>3.20%</td>
<td>38%</td>
</tr>
<tr>
<td>30%</td>
<td>202</td>
<td>53,240</td>
<td>1,243</td>
<td>342,194</td>
<td>2.30%</td>
<td>31%</td>
</tr>
<tr>
<td>20%</td>
<td>22</td>
<td>5,565</td>
<td>83</td>
<td>23,879</td>
<td>1.50%</td>
<td>20%</td>
</tr>
<tr>
<td>10%</td>
<td>4</td>
<td>1,265</td>
<td>14</td>
<td>944</td>
<td>1.10%</td>
<td>4%</td>
</tr>
</tbody>
</table>
craugastoridae 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
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 quiririensis 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).

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).

<table>
<thead>
<tr>
<th>Maximum % of taxa lacking data per locus</th>
<th>Mean bootstrap support across all internal branches</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ML</td>
</tr>
<tr>
<td>90%</td>
<td>94.9</td>
</tr>
<tr>
<td>80%</td>
<td>95.7</td>
</tr>
<tr>
<td>70%</td>
<td>95.4</td>
</tr>
<tr>
<td>60%</td>
<td>93.8</td>
</tr>
<tr>
<td>50%</td>
<td>91.8</td>
</tr>
<tr>
<td>40%</td>
<td>91.6</td>
</tr>
<tr>
<td>30%</td>
<td>51.7</td>
</tr>
<tr>
<td>20%</td>
<td>41.3</td>
</tr>
<tr>
<td>10%</td>
<td>18.0</td>
</tr>
</tbody>
</table>

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 quiririensis 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 Pristimantinae. 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 monogenic 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, Plistimantis, 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, Plistimantis). Both species-tree methods found Strabomantis inside Craugastoridae, 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 Plistimantinae also differ between our results and previous studies. Here, we found Plistimantinae (represented by Lynchius, Oreobates, and Plistimantis) 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, (Plistimantis, (Lynchius, Oreobates))). On the other hand, Padial et al. (2014) recovered a third topology: (Plistimantis, (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 Holoadeninæa (Barycholos) inside Pristimantinæa (Lynchius, Pristimantis, and Öreobates) 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 Holoadeninæa being nested inside Pristimantinæa.

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 Terraranæa. 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 (Holoadeninæa) within Pristimantinæa. 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 Pristimantinæa). 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

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASTRAL-II</td>
<td>Accurate Species TRee Algorithm</td>
</tr>
<tr>
<td>bs</td>
<td>bootstrap support</td>
</tr>
<tr>
<td>gDNA</td>
<td>Genomic DNA</td>
</tr>
<tr>
<td>ICZN</td>
<td>International Commission for Zoological Nomenclature</td>
</tr>
<tr>
<td>MAFFT</td>
<td>multiple sequence alignment program</td>
</tr>
<tr>
<td>ML</td>
<td>maximum likelihood</td>
</tr>
<tr>
<td>STRAW</td>
<td>Species TRee Analysis Web server</td>
</tr>
<tr>
<td>UCEs</td>
<td>Ultraconserved Elements</td>
</tr>
</tbody>
</table>

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 Ramirez-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, JW, 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.

Disclosure statement

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

Supplemental material

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

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.


**Associate Editor: David Gower**