Molecular Phylogenetics and Evolution 119 (2018) 128–143

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Evaluating methods for phylogenomic analyses, and a new phylogeny for a major frog clade (Hyloidea) based on 2214 loci

Jeffrey W. Streicher[a,b,*], Elizabeth C. Miller[a], Pablo C. Guerrero[c,d], Claudio Correa[d], Juan C. Ortiz[d], Andrew J. Crawford[e], Marcio R. Pie[f], John J. Wiens[a]

[a] Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA
[b] Department of Life Sciences, The Natural History Museum, London SW7 5BD, UK
[c] Institute of Ecology and Biodiversity, Faculty of Sciences, University of Chile, 780-0024 Santiago, Chile
[d] Facultad de Ciencias Naturales & Oceanográficas, Universidad de Concepción, Concepción, Chile
[e] Department of Biological Sciences, Universidad de los Andes, A.A. 4976 Bogotá, Colombia
[f] Departamento de Zoologia, Universidade Federal do Paraná, Curitiba, Paraná, Brazil

ARTICLE INFO

Keywords:
Amphibia
Anura
Biogeography
Naive binning
Phylogenomics
Statistical binning

ABSTRACT

Phylogenomic approaches offer a wealth of data, but a bewildering diversity of methodological choices. These choices can strongly affect the resulting topologies. Here, we explore two controversial approaches (binning genes into "supergenes" and inclusion of only rapidly evolving sites), using new data from hyloid frogs. Hyloid frogs encompass ~53% of frog species, including true toads (Bufonidae), glassfrogs (Centrolenidae), poison frogs (Dendrobatidae), and treefrogs (Hylidae). Many hyloid families are well-established, but relationships among these families have remained difficult to resolve. We generated a dataset of ultraconserved elements (UCEs) for 50 ingroup species, including 18 of 19 hyloid families and up to 2214 loci spanning > 800,000 aligned base pairs. We evaluated these two general approaches (binning, rapid sites only) based primarily on their ability to recover and strongly support well-established clades. Data were analyzed using concatenated likelihood and coalescent species-tree methods (NJst, ASTRAL). Binning strongly affected inferred relationships, whereas use of only rapidly evolving sites did not (indicating ~87% of the data contributed little information). The optimal approaches for maximizing recovery and support of well-established clades were concatenated likelihood analysis and the use of a limited number of naive bins (statistical binning gave more problematic results). These two optimal approaches converged on similar relationships among hyloid families, and resolved them with generally strong support. The relationships found were very different from most previous estimates of hyloid phylogeny, and a new classification is proposed. The new phylogeny also suggests an intriguing biogeographical scenario, in which hyloids originated in southern South America before radiating throughout the world.

1. Introduction

Phylogenomic research is now generating massive datasets that can be used to address difficult phylogenetic problems. However, these datasets raise many questions about how the data should be analyzed. For example, should concatenated or coalescent-based (species-tree) analyses be preferred? If coalescent methods are used, which approaches are best? What if the properties of the data do not allow the use of the preferred method (e.g. because of too many genes, too many taxa, or too much missing data)? Should the data primarily determine the choice of methods, or should the choice of methods primarily determine what data are included?

Here, we address three major questions. First, what are the effects of binning on phylogenomic analyses? This approach involves combining sets of genes into bins or "supergenes." These supergenes are intended to provide better estimates of species trees when gene trees are poorly estimated. These supergenes are intended (Bayzid and Warnow, 2013). Simulations suggest that this approach can either improve phylogenetic accuracy (Bayzid and Warnow, 2013) or worsen it (Liu and Edwards, 2015; Liu et al., 2015; but see Springer and Gatesy, 2016), relative to unbinned analyses. There are also many potential approaches to binning, such as naive binning (with different possible numbers of bins) and statistical binning (using compatibility analyses to determine the optimal number of bins; Mirabab et al., 2014b). Second, can accuracy

* Corresponding author at: Department of Life Sciences, The Natural History Museum, London SW7 5BD, UK.
E-mail address: j.streicher@nhm.ac.uk (J.W. Streicher).

https://doi.org/10.1016/j.ympev.2017.10.013
Received 29 May 2017; Received in revised form 21 October 2017; Accepted 22 October 2017
Available online 27 October 2017
1055-7903/

be improved by excluding slower-evolving sites? Recent studies have suggested that accuracy might be improved by including only the fastest evolving sites (e.g. Salichos and Rokas, 2013; Hosner et al., 2016). However, similar to binning, the benefits of this approach have also been disputed (Betancur et al., 2014; Simmons and Gatesy, 2016). Third, what combination of these two approaches optimizes accuracy (i.e. recovery of the true phylogeny)? Previous papers have explored these approaches separately, but it remains unclear what combination of these approaches might yield optimal results.

We evaluate the performance of these approaches using empirical data from frogs. Few empirical systems offer a known phylogeny with which accuracy can be directly evaluated. Nevertheless, those branches that are supported by both molecular and morphological evidence can potentially be used to compare the performance of different sampling and inference methods (e.g. Wiens and Tiu, 2012; Streicher et al., 2016). It is difficult to imagine scenarios by which both molecular and morphological data will be systematically misled to give identical, incorrect relationships, especially if groups are relatively well sampled taxonomically (i.e. no long-branch attraction). Furthermore, empirical data may offer important advantages for evaluating methods relative to simulated data given that empirical data are, by definition, realistic. As one example, phylogenomic datasets often contain some level of missing data, but this is often not incorporated in simulation studies, especially those not specifically focused on this issue. It is not entirely clear what a realistic distribution of missing data would be (e.g. largely random across taxa and genes or more concentrated in certain taxa or genes with particular properties?). This is especially true for phylogenomic data from ultraconserved elements (UCEs; Bejerano et al., 2004; Faircloth et al., 2012), for which many basic properties are still being explored (e.g. Hosner et al., 2016; Meiklejohn et al., 2016; Streicher et al., 2016). Of course, empirical analyses cannot and should not replace simulation studies of method performance. Nevertheless, empirical analyses of method performance offer an important complement to simulation studies, despite remaining relatively underutilized.

In this study, we focus on the phylogeny of hyloid frogs. Hyloid frogs include the majority of frog species (~3600 or ~53%: Pyron and Wiens, 2011; AmphibiaWeb, 2016; Feng et al., 2017). Hyloidea includes many well-known frog families, including the true toads (Bufonidae), treefrogs (Hylidae), glassfrogs (Centrolenidae), and poison frogs (Dendrobatidae). They are distributed globally (especially bufonids and hylids) and include most frog species found in the New World (e.g. Roelants et al., 2007; Pyron and Wiens, 2011; AmphibiaWeb, 2016). Along with Ranoidea, they are one of the two major clades of Neobatrachia, the clade which contains ~95% of all frog species (Ford and Cannatella, 1993; Roelants et al., 2007; Pyron and Wiens, 2011; AmphibiaWeb, 2016).

Many hyloid families are now well established by morphological and molecular data (see below), but relationships among hyloid families have been very difficult to resolve (Fig. 1; Darst and Cannatella, 2004; Frost et al., 2006; Roelants et al., 2007; Pyron and Wiens, 2011; Zhang et al., 2013; Pyron, 2014; Feng et al., 2017; Hutter et al., in press). For example, Pyron and Wiens (2011) conducted a supermatrix analysis of 2871 amphibian species (including 1357 hyloid species), based on likelihood analyses of 12 concatenated genes (Fig. 1d). Among hyloid families, almost no relationships had bootstrap proportions > 70%, except for the clade Terrarana (Brachycephalidae, Ceuthomantidae, Craugastoridae, and Eleutherodactylidae) and the clade uniting Allophrynidae and Centrolenidae. Pyron (2014) analyzed a very similar matrix and obtained very similar results. Other studies have addressed hyloid relationships but with less extensive sampling of taxa (e.g. Darst and Cannatella, 2004; Frost et al., 2006. Roelants et al., 2007; Zhang et al., 2013) and genes (Wiens, 2007, 2011). These studies typically yielded weak support for relationships among hyloid families, and extensive conflicts with other estimates (Fig. 1). In contrast, Feng et al. (2017) found strong support for relationships among a subset of hyloid families using 95 nuclear loci and coalescent analyses (Fig. 1g). However, when they included additional families based on less data, the relationships became weakly supported (Fig. 1h). In summary, relationships among hyloid frogs remain largely unresolved (Fig. 1i). This is unfortunate, especially since numerous studies have now utilized these large-scale estimates of hyloid frog phylogeny, including analyses of life-history evolution (Gomez-Mestre et al., 2012), species richness patterns (Pyron and Wiens, 2013; Hutter et al., in press), diversification (e.g. Roelants et al., 2007; De Lisle and Rowe, 2015; Moen and Wiens, 2017), and ecomorph evolution (Moen et al., 2016).

Here, we analyze relationships among hyloid frogs and empirically evaluate two controversial approaches for phylogenomic data (binning and use of fast sites only). We first generate a novel dataset of ultraconserved (UCE) loci for 50 hyloid species and 5 outgroup taxa (Table 1). We identify 10 clades that are traditionally recognized and are relatively well established by molecular and morphological data. We then evaluate the ability of binning and exclusion of slow-evolving sites to recover and to support these clades (and their support for other clades). We use binning in conjunction with coalescent-based species-tree methods designed for large-scale phylogenomic datasets (NJst: Liu and Yu, 2011; ASTRAL: Mirabab et al., 2014a; Mirabab and Warnow, 2015). Our analyses include naive binning along with weighted and unweighted statistical binning. We also compare these coalescent-based methods to maximum likelihood (ML) analyses of concatenated data, and ML analyses that either include all sites or only fast-evolving sites. We then use the best approach(es) identified by these analyses to infer higher-level phylogenetic relationships among hyloid frogs. Our results offer a strongly supported hypothesis for this important but phylogenetically problematic group.

2. Materials and methods

2.1. Taxon sampling

We sampled 50 species that collectively represent 18 of 19 hyloid families (following the taxonomy of Pyron and Wiens, 2011; AmphibiaWeb, 2016). We were unable to sample the geographically restricted South American family Ceuthomantidae (Heinicke et al., 2009). However, the placement of this family with other terraranan families is well-established (e.g. Heinicke et al., 2009; Pyron and Wiens, 2011; Feng et al., 2017; Hutter et al., in press). We also included representatives of most hyloid subfamilies including: Centroleninae and Hyalinobatrachinae (Centrolenidae), Craugastorinae and Holoadeninae (Craugastoridae), Eleutherodactylinae and Phyzelaphryninae (Eleutherodactylidae), Hylinae, Pelodryadinae, and Phyllomedusinae (Hylidae), Cryptobatrachinae and Hemiphractinae (Hemiphractidae), and Leptodactylinae and Leiuperinae (Leptodactylidae). Note that we differ from AmphibiaWeb (2016) in placing Strabomantidae within Craugastoridae (following Pyron and Wiens, 2011; Padial et al., 2014, and others). We used five non-hyloid taxa as outgroups: *Spea bombifrons* (Scaphiopodidae), *Gastrophryne carolinensis* (Microhylidae), *Rana catesbeiana* (Ranidae), *Calyptocephalella gayi* (Calyptocephalellidae), and *Notaden bennettii* (Myobatrachidae). Many previous analyses have placed Myobatrachidae and Calyptocephalellidae as closely related to Hyloidea, with Ranoidea (including Ranidae and Microhylidae) and Pelobatoidea (Scaphiopodidae) as more distant outgroups (e.g. Roelants et al., 2007; Pyron and Wiens, 2011; Pyron, 2014; Feng et al., 2017). A summary of taxon sampling is presented in Table 1, and voucher information is available in Table S1.

2.2. Targeted sequence capture of ultraconserved elements

We generated UCE data for hyloid anurans using the laboratory protocols described in Faircloth et al. (2012) and the same tetrapod probes as Streicher et al. (2016) and Streicher and Wiens (2016, 2017). These probes are available from http://www.ultraconserved.org (as a FASTA file named “Tetrapods-UCE-5kv1”) and target 5060 UCEs (using

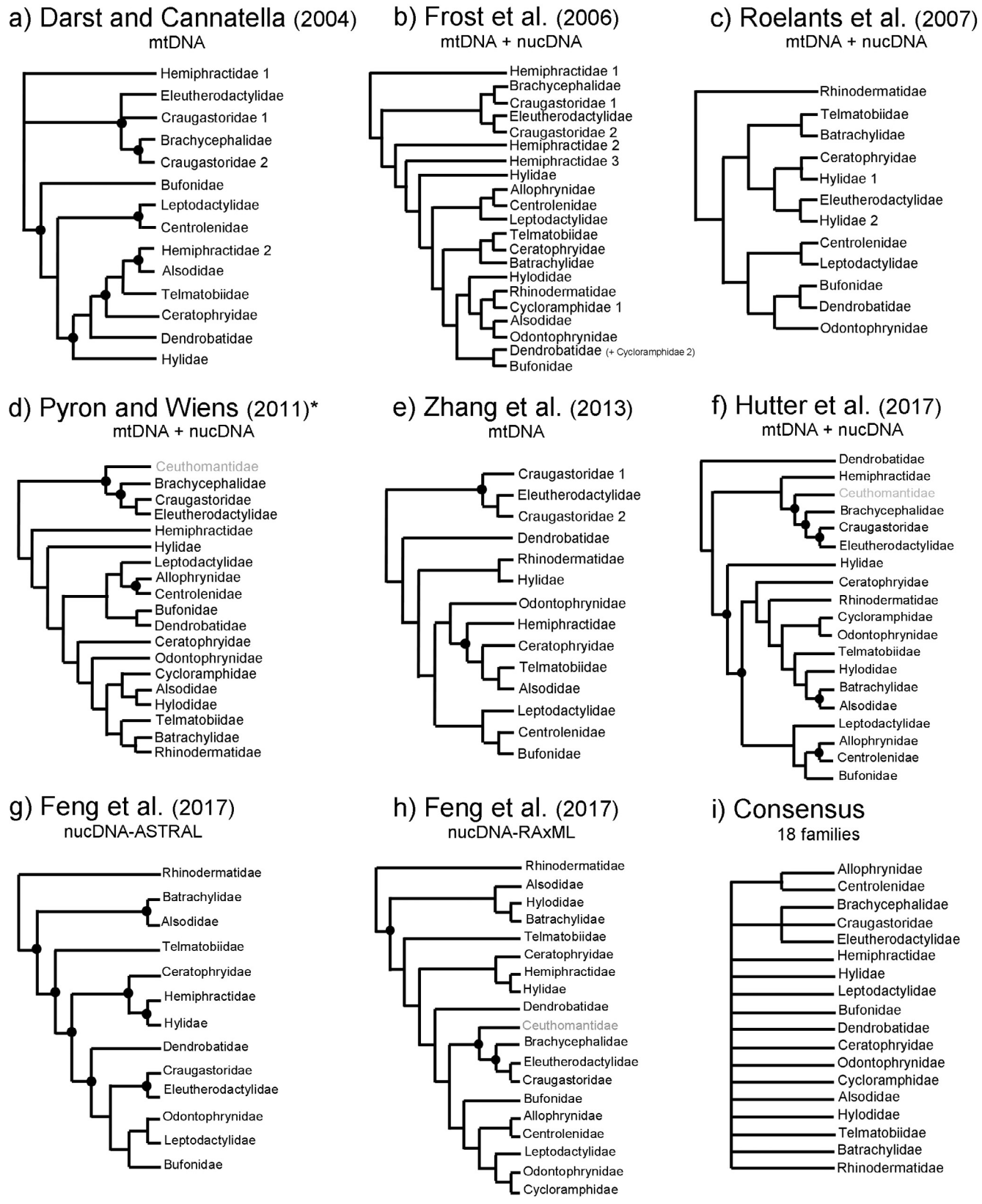

Fig. 1. Phylogenies of hyloid families (sensu Pyron and Wiens, 2011) inferred using molecular data (mitochondrial [mtDNA] and nuclear [nucDNA]). Black circles indicate high support (e.g. > 90% bootstrap or > 0.90 Bayesian posterior probability). (a) Maximum likelihood tree from Darst and Cannatella (2004; their Fig. 2). (b) Consensus of four most parsimonious trees from Frost et al. (2006; their Fig. 50), no support values reported. (c) Bayesian concatenated tree of Roelants et al. (2007; their Fig. 1). (d) Maximum likelihood concatenated tree of Pyron and Wiens (2011; their Fig. 1). The asterisk refers to the maximum likelihood tree from Pyron (2014), which is very similar and therefore not shown separately. (e) Maximum likelihood concatenated tree of Zhang et al. (2013; their Fig. 3). (f) Bayesian concatenated tree of Hutter et al. (in press; their Fig. A1). (g) Coalescent-based species tree (from ASTRAL) of Feng et al. (2017; their Figs. 1, S3). (h) Maximum likelihood concatenated tree of Feng et al. (2017; their Figs. 2, S4), including additional taxa. (i) Consensus of trees in a–e. This graphical consensus was made using only the 18 families we sequenced UCEs from and demonstrates that only two groupings have been consistently supported across studies: (1) Allophrynidae + Centrolenidae and (2) Brachycephalidae + Craugastoridae + Eleutherodactylidae. Not all hypotheses of hyloid relationships are shown, especially those based on fewer genes or taxa (e.g. Gomez-Mestre et al., 2012; Wiens, 2007, 2011). Gray text indicates the one hyloid family (Ceuthomantidae) that was not included in the present study.

Table 1
Species, families, subfamilies, number of velvet contigs, average contig length (n50), number of UCEs sequenced, and NCBI Sequence Read Archive (SRA) accession numbers for anuran amphibians used in this study.

Taxon	Family	Subfamily	Reads	Velvet contigs	n50	UCEs	SRA accession
Adelophryne adiastola	Eleutherodactylidae	Phyzelaphryninae	1,582,691	6948	384	1648	SAMN05559873
Agalychnis callidryas	Hylidae	Phyllomedusinae	515,689	3105	424	1373	SAMN05559871
Allophryne ruthveni	Allophrynidae	N/A	1,271,714	8771	386	2127	SAMN05559872
Alsodes nodosus	Alsodidae	N/A	899,243	5513	448	2157	SAMN05559874
Alsodes pehuenche	Alsodidae	N/A	655,724	4221	462	2150	SAMN05559876
Amazophrynella minuta	Bufonidae	N/A	463,060	1750	326	498	SAMN05559878
Ameerega trivittata	Dendrobatidae	N/A	432,522	1881	324	595	SAMN05559879
Atelognathus salai	Batrachylidae	N/A	419,143	4947	402	1034	SAMN05559880
Atelopus peruensis	Bufonidae	N/A	319,804	2137	388	601	SAMN05559881
Batrachyla leptopus	Batrachylidae	N/A	503,007	4119	469	2037	SAMN05559882
Batrachyla taeniata	Batrachylidae	N/A	839,108	5910	446	2282	SAMN05559883
Brachycephalus quirirensis	Brachycephalidae	N/A	3,252,412	22,624	255	1015	SAMN05559884
Ceratophrys cornuta	Ceratophryidae	N/A	5,301,907	40,359	284	1967	SAMN05559887
Chaltenobatrachus grandisonae	Batrachylidae	N/A	493,707	3936	433	1687	SAMN05559888
Craugastor longirostris	Craugastoridae	Craugastorinae	670,754	2569	382	1367	SAMN05559889
Cycloramphus boraceiensis	Cycloramphidae	N/A	3,522,778	14,118	415	2173	SAMN05559891
Dendropsophus leali	Hylidae	Hylinae	2,070,023	8555	352	1996	SAMN05559892
Eleutherodactylus monensis	Eleutherodactylidae	Eleutherodactylinae	3,149,425	19270	342	2080	SAMN05559901
Espadarana prosoblepon	Centrolenidae	Centroleninae	3,957,199	25,601	272	1857	SAMN05559886
Eupsophus emiliopugini	Alsodidae	N/A	562,563	4299	402	1313	SAMN05559902
Eupsophus roseus	Alsodidae	N/A	450,385	3634	382	1341	SAMN05559905
Flectonotus pygmaeus	Hemiphractidae	Cryptobatrachinae	780,816	3237	330	911	SAMN05559906
Gastrotheca nicefori	Hemiphractidae	Hemiphractinae	276,257	1199	366	453	SAMN05559907
Hemiphractus proboscideus	Hemiphractidae	Hemiphractinae	755,673	5250	405	1654	SAMN05559908
Hyalinobatrachium colymbiphyllum	Centrolenidae	Hyalinobatrachinae	549,489	2756	335	1326	SAMN05559909
Hyla cinerea	Hylidae	Hylinae	164,007	764	296	126	SAMN05559910
Hylodes heyeri	Hylodidae	N/A	420,657	1992	323	1061	SAMN05559911
Hylodes phyllodes	Hylodidae	N/A	2,086,810	7415	387	1952	SAMN05559912
Hylorina sylvatica	Batrachylidae	N/A	707,405	4663	451	2111	SAMN05559913
Hyloxalus nexipus	Dendrobatidae	N/A	725,085	3716	368	1537	SAMN05559914
Hypsiboas lanciformis	Hylidae	Hylinae	846,581	4949	377	1057	SAMN05559916
Incilius alvarius	Bufonidae	N/A	117,981	330	283	133	SAMN05559915
Insuetophrynus acarpicus	Rhinodermatidae	N/A	735,779	5852	434	1764	SAMN05559918
Lepidobatrachus laevis	Ceratophryidae	N/A	2,186,159	5814	347	2226	SAMN05559917
Leptodactylus didymus	Leptodactylidae	Leptodactylinae	994,805	5814	369	2053	SAMN05559919
Litoria caerulea	Hylidae	Pelodryadinae	5,297,219	30,118	201	1898	SAMN05559920
Lynchius nebulanastes	Craugastoridae	Holoadeninae	1,511,042	11,368	479	2346	SAMN05559921
Melanophryniscus stelzneri	Bufonidae	N/A	317,175	1977	366	705	SAMN05559922
Megaelosia apuana	Hylodidae	N/A	602,495	2744	335	1298	SAMN05559923
Odontophrynus americanus	Odontophrynidae	N/A	1,843,432	14,355	346	1994	SAMN05559924
Osornophryne guacamayo	Bufonidae	N/A	184,610	343	290	81	SAMN05559925
Phyllomedusa tomopterna	Hylidae	Phyllomedusinae	147,660	564	260	134	SAMN05559926
Physalaemus cuvieri	Leptodactylidae	Leiuperinae	244,604	1022	335	708	SAMN05559927
Rhinella margaritifera	Bufonidae	N/A	619,572	2885	350	1605	SAMN05559928
Rhinoderma darwinii	Rhinodermatidae	N/A	953,562	6505	426	1981	SAMN05559929
Scinax catharinae	Hylidae	Hylinae	697,312	3413	377	1273	SAMN05559930
Stefania coxi	Hemiphractidae	Hemiphractinae	520,138	3350	446	1930	SAMN05559931
Telmatobius carrillae	Telmatobiidae	N/A	715,750	5032	383	1863	SAMN05559932
Telmatobius truebae	Telmatobiidae	N/A	839,489	5013	395	2233	SAMN05559933
Thoropa miliaris	Cycloramphidae	N/A	2,226,820	9472	436	2129	SAMN05559934
Outgroup taxa							
Calyptocephalella gayi	Calyptocephalellidae	N/A	1,816,831	14,750	415	2372	SAMN05559935
Gastrophryne carolinensis	Microhylidae	Gastrophryninae	2,186,159	12,090	347	2226	SAMN05559936
Rana catesbeiana	Ranidae	N/A	471,302	1059	327	432	SAMN05559937
Notaden bennettii	Myobatrachidae	N/A	347,794	1101	398	547	SAMN05559938
Spea bombifrons	Scaphiopodidae	N/A	3,771,407	22125	285	1821	SAMN05559939

5472 unique probes; note that some probes target the same locus). A particularly appealing aspect of this UCE tetrapod probe set is its ability to capture orthologous loci from highly divergent taxa. Using a similar set of probes, Faircloth et al. (2012) demonstrated that UCEs can reliably be obtained across tetrapods despite > 300 million years of evolution separating some lineages. However, they only included a single species of amphibian, *Xenopus tropicalis*. Interestingly, they identified ~ 1000 UCEs from *X. tropicalis*, thousands fewer loci than for other tetrapod clades (i.e. birds, mammals, squamates). This was expected given the levels of divergence between amphibians (the earliest branching tetrapod clade) and the taxa that the probes were primarily designed from (birds and lizards; Faircloth et al., 2012). Therefore, an ancillary goal of our study was to ascertain the rate of UCE capture success for this probe set for additional amphibians, to assess its utility for anuran phylogenomics.

In the laboratory, we modified several steps of the protocol of Faircloth et al. (2012), as previously described (Streicher et al., 2016). Briefly, these modifications included: (1) initial DNA shearing using NEBNext® dsDNA Fragmentase® (New England Biolabs) or a sonicator (Biorupter®, Diagenode), (2) pooling up to 48 individually-barcoded samples prior to each capture and not diluting the Sure Select XT target enrichment kit (Agilent), and (3) size-selection prior to shotgun library amplification using a Pippin Prep electrophoresis system (Sage Science) on 12 ligated samples that each had a concentration of 10–30 ng/μL

(total of 100–300 ng of DNA/pool). In total we performed three rounds of sequencing on an Illumina® MiSeq at the University of Texas at Arlington Genomics Core Facility (Arlington, Texas, USA; http://gcf.uta.edu). Each sequencing run contained UCE data from 48 individuals (but not all individuals were for this study).

We processed demultiplexed raw Illumina® data as previously described in Streicher et al. (2016). We used *illumiprocessor* 2.0.2 (Faircloth, 2013; Bolger et al., 2014) for quality control, Velvet 1.2.10 (Zerbino and Birney, 2008) for *de novo* assembly of contigs for each species, and PHYLUCE 2.0.0 (Faircloth et al., 2012; also see Faircloth, 2016) to identify and align UCEs, process sequence data, and generate alignments for phylogenetic analyses. We aligned contigs in PHYLUCE 2.0.0 using the MUSCLE alignment algorithm (Edgar, 2004) with settings described in Streicher and Wiens (2016).

We used output from Velvet 1.2.10 to determine (for the sample for each species) the number of (Table 1): (1) quality-trimmed reads, (2) assembled contigs, and (3) average contig length (n50 statistics). To determine the total number of UCEs captured for each species we used output from PHYLUCE 2.0.0, which reports the number of contigs matching the probe set FASTA file (Table 1). We deposited our raw FASTQ files for each species in the NCBI sequence read archive (http://www.ncbi.nlm.nih.gov/sra; Table 1). A summary of commands used to execute *illumiprocessor*, PHYLUCE, and Velvet is available in our supporting documents (Command Summary 1).

2.3. Phylogenetic analyses

Our primary analyses involved a comparison of the performance of a coalescent method (NJst; Liu and Yu, 2011) under different levels of naive binning and with inclusion and exclusion of the putative slowest sites. We also compared these results to those from concatenated likelihood analyses, including and excluding the slowest sites, and to an alternative coalescent method (ASTRAL; Mirabab et al., 2014a), with naive, unweighted and weighted statistical binning (Mirabab et al., 2014b; Bayzid et al., 2015), and to NJst with statistical binning. Most analyses were conducted on the same baseline dataset, using the same level of missing data (up to 60% per included locus, see below). We also performed a more limited series of secondary analyses that addressed other issues, including (a) different levels of missing data, (b) use of the putative slowest sites only (the slowest 60–90%), and (c) exclusion of highly incomplete taxa. We note that examining the questions in the primary analyses across all the conditions in the secondary analyses would have led to a very large parameter space, beyond the scope of the present study. Importantly, our secondary analyses strongly suggested that most of the results would be redundant with those of our primary analyses (or the secondary analyses already described).

We describe the primary analyses first. For these analyses, we generated a baseline dataset that excluded any UCE locus with > 60% missing taxa. This level of missing data allowed inclusion of many loci (2214 UCEs; Fig. 2a). Decreasing the amount of missing data dramatically decreased the number of loci included (to 807), and had little affect on topology and support (see Section 3). Concatenated datasets and UCE trees for coalescent analyses were assembled as described by Streicher et al. (2016). Briefly, we performed concatenated maximum likelihood (ML) analyses using RAxML 8.2 (Stamatakis, 2014) with a single GTRCAT model across the alignment (partitions are unclear for UCE data because they are generally not protein coding, and therefore partitions were not used). We used the Cyberinfrastructure for Phylogenetic Research (CIPRES) science gateway (Miller et al., 2010) for all concatenated analyses. We generated gene trees for all NJst analyses using RAxML 8.0 and assessed branch support via the bootstrapping method of Seo (2008). We ran all NJst analyses using the Species Tree Analysis Web Server (STRAW; Shaw et al., 2013). In theory, other coalescent methods could have been used. However, the set of possible methods is limited by the large size of our dataset. Further, among these candidate methods, NJst has accuracy similar to ASTRAL II in simulations (Mirabab and Warnow, 2015).

We performed some analyses using an alternative coalescent method (ASTRAL). We assessed branch support for ASTRAL using the method of Seo (2008), and ran ASTRAL 4.7.6 analyses on the Natural History Museum computing clusters. ASTRAL 4.7.6 differs from NJst in requiring the species tree be estimated before bootstrapping is performed. We followed the recommendation of the developers and inferred a species tree from an ASTRAL 4.7.6 analysis of the best maximum likelihood (ML) tree for each locus estimated by RAxML 8.0. ASTRAL was applied to two datasets, one including loci with up to 40% missing data (807 loci) and one allowing loci with up to 60% missing data (2214 loci). For each dataset, bootstrap support was estimated by generating 100 bootstrap pseudoreplicate gene trees for each locus. Bootstrap support for each node was then drawn on the ASTRAL tree inferred from the best ML tree for each locus. A summary of commands used to execute RAxML, NJst, and ASTRAL 4.7.6 analyses is available in our supporting documents (Command Summary 2).

2.4. Primary analyses: binning strategies

We analyzed four different levels of naive binning, either (a) no bins (standard analysis, all genes treated separately) or else datasets consisting of (b) 4 supergenes, (c) 40 supergenes, and (d) 400 supergenes. To create these supergenes, we used the Geneious 8.0 (BioMatters Ltd.) alignment editor (4 and 40 supergenes) or a customized perlscript (400 supergenes) to arbitrarily split the concatenated alignment generated by PHYLUCE for the 60% missing taxa per locus dataset into evenly-sized bins (Table 5). We calculated the size of naive supergenes by dividing the length of the concatenated alignment by 4, 40, or 400, respectively. We then divided the alignment up into supergenes based on the original order of genes in the concatenated alignment. For example, for the 4 supergene “all sites” analysis, each supergene comprised ca. 200,542 base pairs (i.e. 802,167/4). We then made the supergenes by separating the first 200,542 base pairs (naive supergene 1; bases 1–200,542 of the original alignment), second 200,542 base pairs (naive supergene 2; bases 200,543–401,084 of the original alignment), third 200,542 base pairs (naive supergene 3; bases 401,085–601,626 of the original alignment), and remaining 200,541 base pairs (naive supergene 4; bases 601,627–802,167 of the original alignment). We considered the potential for non-random concatenation to influence tree estimation when using naive supergenes. The concatenated alignment was made using individual UCE alignments in PHYLUCE, so there was no risk of concatenating loci on the basis of gene-tree topology. We also tested for (and did not find) biases in evolutionary rate and data completeness along the 802,167 bp concatenated alignment by examining the 4 supergenes “all sites” alignments. Specifically, for each supergene we used RAxML to identify proxy measures for the relative amount of missing data (proportion of gaps and completely undetermined characters) and evolutionary rate (number of distinct alignment patterns) with the expectation that if PHYLUCE concatenates UCEs using one of these characteristics, we would see a gradual increase or decrease in their values across the supergenes. We observed similar proportions of alignment patterns and missing data across the supergenes, with no consistent increase or decrease in their values (naive supergene 1: 76,841 distinct patterns, 55.49% missing; naive supergene 2: 78,080 distinct patterns, 55.52% missing; naive supergene 3: 75,603 distinct patterns, 55.59% missing; naive supergene 4: 77,078 distinct pattern, 55.56% missing). We acknowledge that using 4, 40, and 400 supergenes is somewhat arbitrary. However, our goal here was to explore how our phylogenetic results are affected by binning across three orders of magnitude, and to reveal trends in how changing the number of bins affects our criteria for method performance. Following standard practice, we used a single selection of genes for the supergenes for each level of binning (but we acknowledge that there are numerous ways to assemble genes into supergenes).

We also used ASTRAL to investigate two non-naive binning strategies:

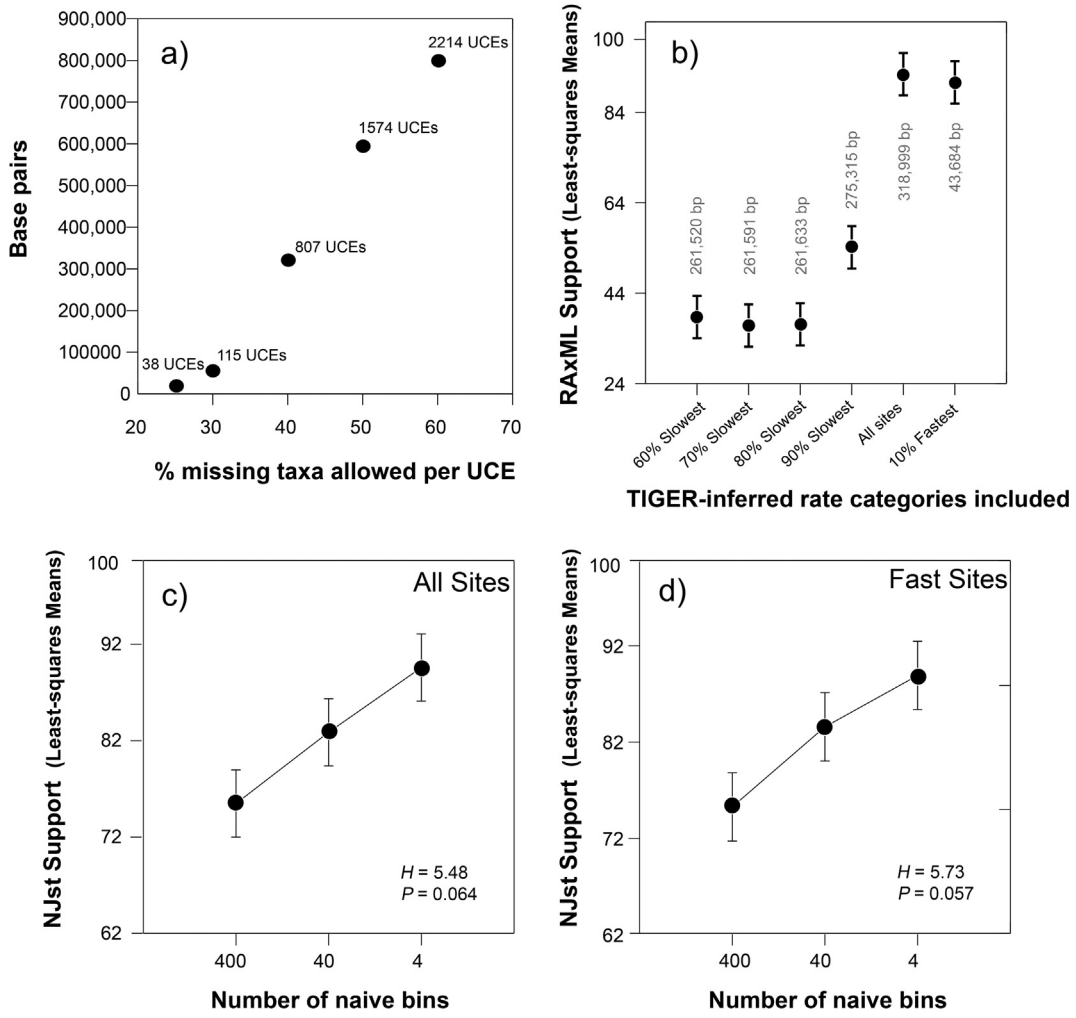

Fig. 2. Basic properties of the UCE data analyzed in this study for the 55 sampled frog species. (a) Relationship between overall dataset size and the maximum percentage of missing taxa allowed per UCE locus (i.e. 60% indicates that loci with missing data in > 60% of the taxa are excluded). The y-axis shows the number of base pairs in the concatenated alignment, and the number above each datapoint is the number of loci included. Results for b, c, and d are based on including genes with up to 60% missing taxa per locus. For b–d, bars around data points indicate standard error on the mean. Kruskal-Wallis statistics (testing for differences in the mean support values across the different number of bin categories) are also shown in c and d. (b) The relationship between mean bootstrap support (from concatenated likelihood analyses in RAxML, across all branches in each estimated tree) and the sites included in each analysis (total number of sites included is in gray above data point), where sites are evolving at different rates (based on TIGER-inferred rate categories). (c) Effect of different numbers of supergenes (used in naive binning) on mean branch support for all branches for NJst, when including all sites. (d) Effect of different numbers of supergenes (used in naive binning) on mean branch support for all branches for NJst, when including only the fastest 10% of sites.

unweighted and weighted statistical binning. These binning methods differ from naive binning in that before supergenes were constructed, UCEs were grouped into bins based on a statistical test for compatibility (Mirabab et al., 2014b). Unweighted binning uses the statistically sorted bins directly to infer the species tree. In contrast, weighted binning weights each supergene tree by the number of UCEs used to construct that supergene when estimating the species tree (Bayzid et al., 2015). We conducted these analyses using the 60% missing (2214 UCEs) dataset. To perform statistical binning we used the pipeline available at https://github.com/smirarab/binning. This pipeline uses some elements of the program DendroPy 3.12 (Sukumaran and Holder, 2010). We ran the pipeline using a support threshold of 50 (see Bayzid et al., 2015) and inferred the resulting supergene trees using the same RAxML criteria as our naive binning analyses. To compare unbinned, unweighted, and weighted analyses, we used LPP support measures generated in the newer version of ASTRAL, 4.10.12 (Sayyari and Mirabab, 2016). For comparison, we also performed an analysis of 4 supergenes using ASTRAL 4.7.6 (using the same naive bins described above), with branch support estimated by bootstrapping. A summary of commands used to execute the statistical binning pipeline and ASTRAL 4.10.12 is available in our supporting documents (Command Summary 3).

The phylogenomic methods we used estimate branch lengths in different ways. RAxML branch lengths are in average substitutions per site (Stamatakis, 2014). NJst branch lengths are the average number of internodes (Liu and Yu, 2011). ASTRAL 4.10.12 branch lengths are equivalenet to coalescent units (Sayyari and Mirabab, 2016). In ASTRAL 4.7.6, branch lengths are not estimated, and trees are presented as cladograms.

2.5. Primary analyses: site-exclusion strategies

We also analyzed two different patterns of site exclusion. We either included only the fastest 10% of sites across all included loci, or else included all sites. We identified putatively rapidly evolving sites using Tree Independent Generation of Evolutionary Rates (TIGER; Cummins and McInerney, 2011). We used the default 10 rate categories. Sites in the tenth rate category represented the most rapidly evolving sites (hereafter called the fastest 10% sites). Note that the fastest 10% of sites inferred by TIGER do not perfectly correspond to 10% of the dataset. Instead, TIGER places sites in rate categories in part on the basis of levels of character conflict (Cummins and McInerney, 2011). For example, in a UCE dataset, most sites will be placed in the first (i.e.

slowest) rate category because they are “ultraconserved” and thus invariant across taxa. Note that this analysis is dataset dependent so “fast sites” identified from hyloid UCEs may have slower or faster evolutionary rates than “fast sites” in other studies. For coalescent analyses that incorporated only the fastest sites and different levels of naive binning (NJst), slow sites were first excluded and the resulting fast-only alignment was subsequently binned into supergenes (Table 5).

A recent study has suggested that the rate categories generated by TIGER may be biased under certain circumstances (Simmons and Gatesy, 2016). To test if the subset of our data identified as fastest by TIGER was reasonable, we assessed the distribution of characters across TIGER-inferred rate categories with the expectation that the most variable sites (those with 3 and 4 observed states) should be most common in the fastest rate categories (all other things being equal; Hansmann and Martin, 2000). We counted the number of character states observed across all taxa for each site using PAUP 4.0 (Swofford, 2002) and Microsoft Excel and found that 99.4% of 3-state sites and 99.7% of 4-character state sites were restricted to the upper 5 rate categories (upper 50%). Furthermore, the category with the presumed fastest 10% of sites contained 75% of the 3-state sites and 79% of the 4-state sites (Table 6). Thus, we assumed that TIGER was adequate at partitioning site-specific rates overall and that the tenth rate category was a reasonable proxy for the set of fastest sites.

A potential issue with using only rapidly evolving sites is that these sites may be more prone to homoplasy and thus mislead phylogenetic inference (e.g. Cummins and McInerney, 2011). We used saturation plots to test for putative homoplasy in the form of site saturation. We used a method proposed by Philippe et al. (1994) to generate saturation plots for the 60% missing 2214 UCE datasets with: (1) all sites and (2) fastest 10% sites. Comparisons of raw genetic distances to TrN model-corrected distances revealed that the fast sites dataset had a concave slope indicative of site saturation (Fig. S1). However, the resulting trees from these datasets do not suggest that homoplasy is a problem given that they are highly similar to each other in topology (Figs. S2 and S3) and support values (Fig. 2a and c).

We note that an obvious alternative approach would be to include only the slowest sites. We included this analysis also, and found that it gave very poor support values, and was clearly problematic (Fig. 2b; Figs. S4–S7). We also note that focusing on the fastest 10% is entirely arbitrary (vs. the fastest 15% or 25%, for example). However, our results show that there is relatively little difference between including only the fastest 10% and including all sites (Fig. 2b). Therefore, using a different criterion seems unlikely to give very different results.

2.6. Primary analyses: criteria for accuracy

We evaluated the performances of different approaches based on three criteria. An “approach” was a combination of phylogenetic method (NJst, ASTRAL, concatenated ML), binning strategy (no bins; 4, 40, and 400 supergenes; weighted and unweighted statistical binning), and site-inclusion strategy (e.g. fastest 10% vs. all sites included; see Fig. 2c and d). Our first criterion was the proportion of well-established clades that were recovered by a given approach. We considered 10 clades to be well-established (see below). Our second criterion was the mean support for these well-established clades from a given approach (e.g. mean bootstrap values across the 10 clades, with a bootstrap value of 0 assigned if the clade was not recovered). Our third criterion was the mean support across all other clades (i.e. unknown clades). We acknowledge that this third criterion is less obvious as a measure of accuracy. However, our previous analyses (Streicher et al., 2016) support the idea that this criterion can be strongly related to the other two (see also Section 3). We tested whether these criteria were correlated with each other using non-parametric Spearman rank correlation in R version 3.1.3 (R Core Team, 2013). In theory, these different criteria might prefer very different approaches. However, we found that they were generally correlated and agreed broadly upon the optimal approach (es).

We considered 10 clades to be well established, based on congruent support from (a) traditional taxonomy, (b) Bayesian analyses of morphology (Wiens et al., 2005), and (c) analyses of multiple molecular datasets. These 10 clades were as follows: (1) Ranoidea, (2) Bufonidae, (3) Centrolenidae, (4) Ceratophryidae, (5) Dendrobatidae, (6) Hemiphractidae, (7) Hylidae, (8) Phyllomedusinae (subfamily of Hylidae), (9) Leptodactylidae, and (10) Terrarana (i.e. Brachycephalidae, Craugastoridae, Eleutherodactylidae). These clades were strongly supported in Bayesian morphological analyses (Wiens et al., 2005; albeit with limited taxon sampling in some cases), except Leptodactylidae (with moderate support, posterior probability = 0.79), Dendrobatidae (only one species included), and Hylidae and Ranoidea (not monophyletic). Nevertheless these groups are supported by derived phenotypic traits (reviewed in Ford and Cannatella, 1993), including Ranoidea (e.g. fused epicoracoid cartilages) and Dendrobatidae (e.g. retroarticular process on mandible, divided dermal scutes on dorsal surfaces of fingers). Leptodactylidae (formerly Leptodactylinae) is supported by two striking derived traits: an ossified sternal style and use of foam nests (Lynch, 1971). For Hylidae, many of the phenotypic traits that potentially unite them (most associated with arboreality) are also shared with the largely arboreal allophrynids, centrolenids, and hemiphractids (in morphological analyses, hylids form a non-monophyletic group in a clade uniting these families; Wiens et al., 2005). However, monophyly of Hylidae has been strongly supported in many molecular analyses (e.g. Wiens et al., 2005; Pyron and Wiens, 2011; Pyron, 2014) and the derived traits uniting hylids are presumably synapomorphies if previous molecular analyses are correct in showing that hylids are not the sister taxon to allophrynids, centrolenids, or hemiphractids (i.e. as is the case in most trees in Fig. 1).

We acknowledge that, in theory, one or more of these 10 clades still might not be correct (a weakness of almost all empirical studies of method performance). However, there is little support for this idea from our phylogenomic analyses (see Section 3). Thus, we find that the method that recovers all 10 of these clades does so with relatively strong mean branch support, and shows strong mean support for most other relationships across the tree. In contrast, when methods fail to recover one or more of these 10 clades, the support for the conflicting relationships tends to be weak, and there is weaker support for other branches across the tree (i.e. our performance criteria tend to be correlated across methods). Furthermore, we found that in many cases, the failure to support a well-established clade was caused by the apparent misplacement of a single, relatively incomplete taxon (i.e. those with few UCEs). Thus, phylogenomic results that contradict any of these 10 clades generally seem to reflect weak support rather than hidden evidence that one or more of these 10 clades are actually incorrect.

2.7. Secondary analyses: missing data and gene-tree resolution

We also performed a series of secondary analyses that addressed (a) exclusion of loci with different levels of missing data, (b) exclusion of highly incomplete taxa, (c) use of the slowest sites alone, and (d) potential for poor gene-tree resolution. First, using PHYLUCE we allowed different levels of missing data, alternatively excluding UCEs with either > 25%, > 30%, > 40%, > 50% and > 60% missing taxa (Fig. 2a). We then identified two levels of missing data (including loci with up to 40% and 60% missing taxa/UCE) that produced relatively large datasets while still representing different levels of data completeness (49.1% and 55.5% missing data; 318,999 bp and 802,167 bp in total length, respectively). For each of these two datasets, we ran concatenated ML, NJst, and ASTRAL analyses.

Second, to explore the influence of highly incomplete taxa, we constructed a dataset that excluded any taxon with < 500 UCEs enriched (starting from the 60% missing dataset) and then ran our standard analyses (concatenated ML and NJst). This taxonomically reduced dataset included 2455 UCEs from 46 taxa and was 892,640 bp in length.

Note that this dataset had more loci than the other 60% dataset because (after excluding highly incomplete taxa) each locus was more complete on average and fewer taxa were needed to meet the criterion for locus inclusion (i.e. non-missing data were needed for only 28 taxa to retain a UCE in the 46 taxa dataset vs. 33 taxa needed to keep a UCE in the 55 taxa dataset). Given that the major differences between RAxML and NJst persisted despite the removal of highly incomplete taxa (e.g. placement of Rhinodermatidae), we did not perform these analyses with the 40% missing dataset.

Third, to examine the relationship between inclusion of putative fast evolving sites and overall support, we used TIGER to remove the fastest 10%, 20%, 30%, and 40% of sites for one of the levels of missing data (40% missing; 318,999 bp total). We then conducted concatenated ML analyses on each dataset and compared them to concatenated ML results obtained using the entire dataset and the 10% fastest sites alone. A summary of commands used to execute secondary analyses and TIGER is available in our supporting documents (Command Summary 4).

Fourth, we explored a potential issue for our coalescent analyses related to an intrinsic property of UCEs: that they are largely conserved, with most variability occurring in the regions that flank the core of the UCE-anchor (Faircloth et al., 2012). Shorter UCE contigs are therefore prone to having smaller flanking regions and less variability. Thus, the shorter the UCE contig, the more unlikely it is to yield a well-resolved gene tree. Across thousands of gene trees, this property of UCEs could ostensibly lead to a sampling error that confuses low phylogenetic signal with incomplete lineage sorting. As described by Betancur et al. (2014), phylogenetic analyses of short sequences (in general) are more prone to this type of sampling error than longer sequences (Rasmussen and Kellis, 2007). To test our main dataset for this sampling bias (60% missing 2214 UCE loci), we conducted a similar analysis to that of Betancur et al. (2014). To establish levels of incongruence between gene trees and species trees, we measured the topological distance between the tree from each UCE contig and two alternative tree topologies including all the data (concatenated all sites 60% missing [Fig. S2]; ASTRAL no bins, all sites 60% missing [Fig. S8]). We then tested for a relationship between incongruence and UCE contig length, with the expectation that, if sampling bias were present, we would see a negative relationship between incongruence and UCE contig length (as observed in Betancur et al. (2014)). We calculated topological distances between the best ML tree for each UCE locus (same trees used in the ASTRAL LPP analyses) and alternative trees using the R package APE (Paradis et al., 2004; R Core Team, 2013). We used PH85 distances, which are calculated as twice the number of internal branches defining different bipartitions of the tips (Penny and Hendy, 1985).

3. Results

The results are organized as follows. First, we present our results comparing the performance of different phylogenomic approaches. Second, from this comparison, we identify the approach that should provide the best estimate of hyloid frog relationships. Third, we present our secondary analyses, which address the robustness of this estimate to alternative approaches. Fourth, we present our preferred estimate of hyloid frog phylogeny and compare this to previous estimates.

3.1. Evaluating phylogenomic approaches

We evaluated a total of 14 approaches, where an approach is defined here as the combination of phylogenetic method (e.g. ML, NJst, ASTRAL), binning procedure, and site-inclusion strategy (Table 2). The 14 approaches were: standard concatenated ML with all sites (1); ML with the fastest 10% of sites only (2); standard NJst with all sites and no bins (3); and NJst with 4, 40, and 400 supergenes, comprising either all sites (4–6) or only the fastest 10% of sites (7–9); NJst with all sites and unweighted statistical binning (10); standard ASTRAL, with all sites and no bins (11); ASTRAL with unweighted statistical binning (12); ASTRAL with weighted statistical binning (13); and ASTRAL with 4 supergenes (14). All 14 analyses were applied to the dataset that included loci with up to 60% missing taxa per locus, which maximized the number of loci included while reducing missing data overall. We acknowledge that we did not examine every possible combination of site inclusion approach (all vs. fastest 10%), coalescent method (NJst and ASTRAL), and binning strategy (no bins; naive binning with 4, 40, or 400 supergenes; unweighted statistical binning; weighted statistical binning). This is because our results (see below) showed that: (a) the use of all sites versus fast sites alone had little affect on the results, and (b) the use of many supergenes (40 and 400), and statistical binning did not strongly improve NJst or ASTRAL analyses relative to analyses without bins. Also, we did not perform ASTRAL or NJst analyses of the unbinned data with only the fastest 10% of sites because the reduced number of sites would be divided among thousands of loci, and the limited number of included sites per locus would then be so small as to be impractical for estimating gene trees.

Our three criteria for evaluating phylogenomic approaches were generally correlated across the 14 different approaches compared. Specifically, there was a significant, positive correlation between the proportion of well-established clades recovered and the mean support for those clades across the 14 approaches ($r = 0.81$, $P = .001$; Fig. 3a). There was also a significant, positive correlation between mean support for well-established clades and mean support for all other clades ($r = 0.74$, $P = .002$; Fig. 3b). However, there was not a significant correlation between the proportion of well-established clades recovered by a given approach and the mean support for all other clades across the tree ($r = 0.34$, $P = .2350$). It should be noted that most ASTRAL analyses used local posterior probabilities (LPP; Sayyari and Mirabab, 2016) as a measure of branch support, whereas bootstrapping was used in the concatenated, NJst, and 4 supergene ASTRAL analyses. We multiplied LPP support by 100 for comparisons. We also compared LPP and bootstrap support from ASTRAL analyses based on the same underlying data (60% missing, no bins) and found them to be strongly correlated (Fig. S9; $r = 0.91$, $P < .001$).

These three criteria generally agreed as to the best approach (Table 2). Specifically, the concatenated ML analysis (with all sites or with only the fastest 10%) was the only approach that recovered all 10 well-established clades (100%), provided the highest confidence in these clades (mean bootstrap of 97% and 99%, respectively), and provided relatively strong mean support for other relationships across the tree (mean bootstrap of 91% and 89%, respectively). NJst and ASTRAL analyses including only 4 supergenes provided similar but slightly weaker results based on all three criteria, including the percentage of well-established clades supported (90%), mean support for well-established clades (90%), and mean support for other clades (86–87%). NJst analyses including 40 and 400 supergenes (with all sites or only the fastest 10%) performed worse by all three criteria, recovering fewer well-established clades (70–80%), lower mean support for well-established clades (63–70%), and lower mean support for estimated clades across the tree (73–80%). NJst analyses with statistical binning also performed relatively poorly. NJst analyses with no bins and including all sites recovered a similar percentage of well-established clades (90%) as the 4 supergene NJst analyses, but with lower support for these relationships (79%) and for relationships across the tree (74%). Analyses using ASTRAL (all sites) without bins performed similarly to NJst without bins (both recover 9 of 10 well-established clades), but gave higher mean support for both well-established and unknown clades. Use of unweighted and weighted statistical binning with ASTRAL (all sites) reduced the number of well-established clades that were recovered to 7 of 10 (and reduced their mean support) but increased the mean support for the unknown clades. Disturbingly, these statistical binning analyses also showed strong support for clades that contradicted the well-established clades (i.e. Phyllomedusinae, Leptodactylidae). Overall, we found that the concatenated likelihood analyses performed best overall, followed closely by the NJst and ASTRAL analyses with the smallest

Table 2
Results from primary analyses, including loci with up to 60% missing taxa per locus. The 4, 40, and 400 bins refer to the naive binning datasets. An asterisk indicates that support is from local posterior probabilities (LPP; Sayyari and Mirabab, 2016) whereas other analyses used multi-locus bootstrapping (Seo, 2008). See also Fig. 3.

Dataset	Proportion of well-established clades recovered	Mean support for well-established clades	Mean support for unknown clades
RAxML (all sites)	1.0	97.0	91.4
RAxML (fast sites only)	1.0	98.9	88.7
NJst (no bins – all sites)	0.9	79.4	73.7
NJst (4 bins – all sites)	0.9	89.9	87.4
NJst (4 bins – fast sites)	0.9	90.0	85.6
NJst (40 bins – all sites)	0.7	69.6	80.3
NJst (40 bins – fast sites)	0.7	68.6	80.3
NJst (400 bins – all sites)	0.8	68.5	73.6
NJst (400 bins – fast sites)	0.8	63.1	73.4
NJst (unweighted statistical binning)	0.7	65.1	75.2
ASTRAL (no bins)*	0.9	88.3	83.8
ASTRAL (182 bins – unweighted statistical binning)*	0.7	69.1	85.8
ASTRAL (182 bins – weighted statistical binning)*	0.7	70.0	94.3
ASTRAL (4 bins – all sites)	0.9	90.0	87.5

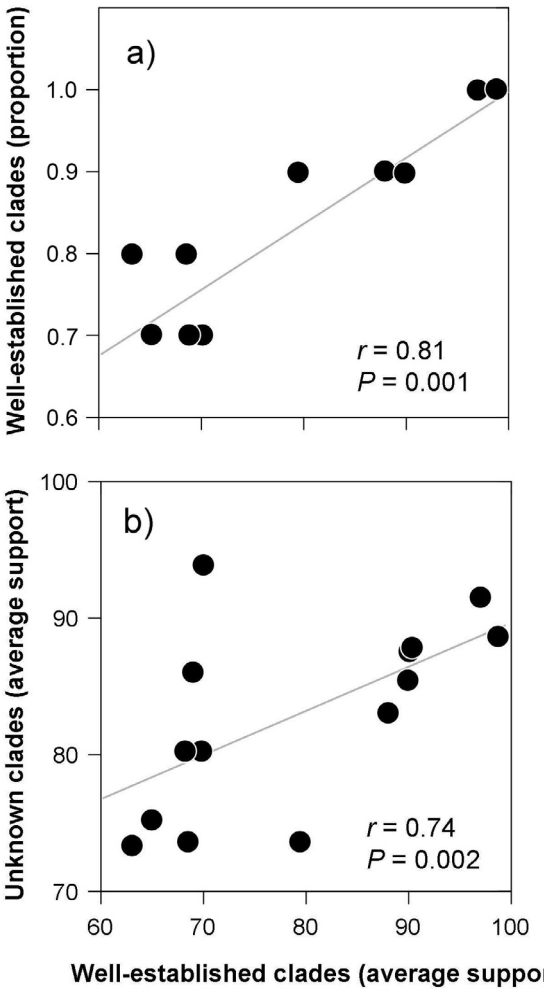

Fig. 3. Comparing three performance criteria for 14 different approaches applied to the primary hyloid dataset (2214 UCEs and up to 60% missing taxa per gene). Different approaches include concatenated likelihood (all sites, fastest 10%), NJst with all sites, and using 4, 40, and 400 supergenes (all sites and fastest 10%), NJst with statistical binning, and ASTRAL with no bins and with weighted and unweighted statistical binning. (a) The proportion of the 10 well-established clades that are supported by each approach is correlated with the branch support for these clades. (b) The mean support values for the 10 well-established clades are strongly correlated with mean support across other all clades. Non-parametric Spearman's rho statistics are included.

number of supergenes (Table 2).

Importantly, these different approaches sometimes gave very different phylogenetic estimates (apart from the well-established clades). The concatenated ML analyses (Fig. 4a) and NJst and ASTRAL analyses with 4 supergenes (Figs. S2, S3, S10–S12) gave largely concordant estimates. In these trees, Rhinodermatidae was the sister taxon to all other hyloids. Next, hyloids were divided into a clade containing four families from southern South America (Austral Clade hereafter; including Alsodidae, Batrachylidae, Cycloramphidae, and Hylodidae) and a clade containing all other hyloids. The clade including all other hyloids then consisted of Telmatobiidae as the sister taxon of two other clades. One of these clades contained the families Ceratophryidae, Hemiphractidae, and Hylidae. The other contained Terrarana (Brachycephalidae, Craugastoridae, and Eleutherodactylidae), and its sister group, consisting of two subclades, one with Allophrynidae + Centrolenidae and Dendrobatidae + Leptodactylidae, the other with Bufonidae + Odontophrynidae. The only difference among these estimates was that in the NJst trees and ASTRAL trees, Leptodactylidae was non-monophyletic (in contrast to the ML tree). Almost all of these relationships were strongly supported (Fig. 4).

In contrast, NJst analyses with 40 supergenes (fast and all sites) placed Rhinodermatidae with the Austral Clade (Alsodidae, Batrachylidae, Cycloramphidae, and Hylodidae) and showed non-monophyly of Bufonidae and Leptodactylidae (Figs. S13, S14). Many relationships among families remained strongly supported, but support for many others was relatively weak.

NJst analyses with 400 supergenes (both with fast sites and with all sites) also placed Rhinodermatidae with the Austral Clade and showed non-monophyly of Bufonidae and Leptodactylidae (Figs. S15, S16). They also no longer placed Telmatobiidae as the sister taxon to the main clade of nonaustral Hyloidea. Most relationships among families were only weakly supported.

Standard NJst analyses with all sites and no bins weakly placed the clade of Ceratophryidae, Hemiphractidae, and Hylidae as the sister taxon to all other hyloids (Fig. 4b). The Austral Clade was strongly supported. The sister group to the Austral Clade was Rhinodermatidae (with strong support for its placement), and the sister taxon to this pair of clades was Telmatobiidae. Most other between-family relationships across the tree were only weakly supported (except monophyly of Terrarana).

ASTRAL analyses without bins, and with unweighted and weighted statistical binning (each with 182 bins with an average length of 4,408 bp ± 337 S.D.; Figs. S8, S17, S18), also placed Rhinodermatidae with the Austral Clade, Telmatobiidae as the sister taxon to the remaining hyloids, and the clade of Hemiphractidae, Ceratophryidae, and Hylidae as the sister taxon to the clade including all other hyloids

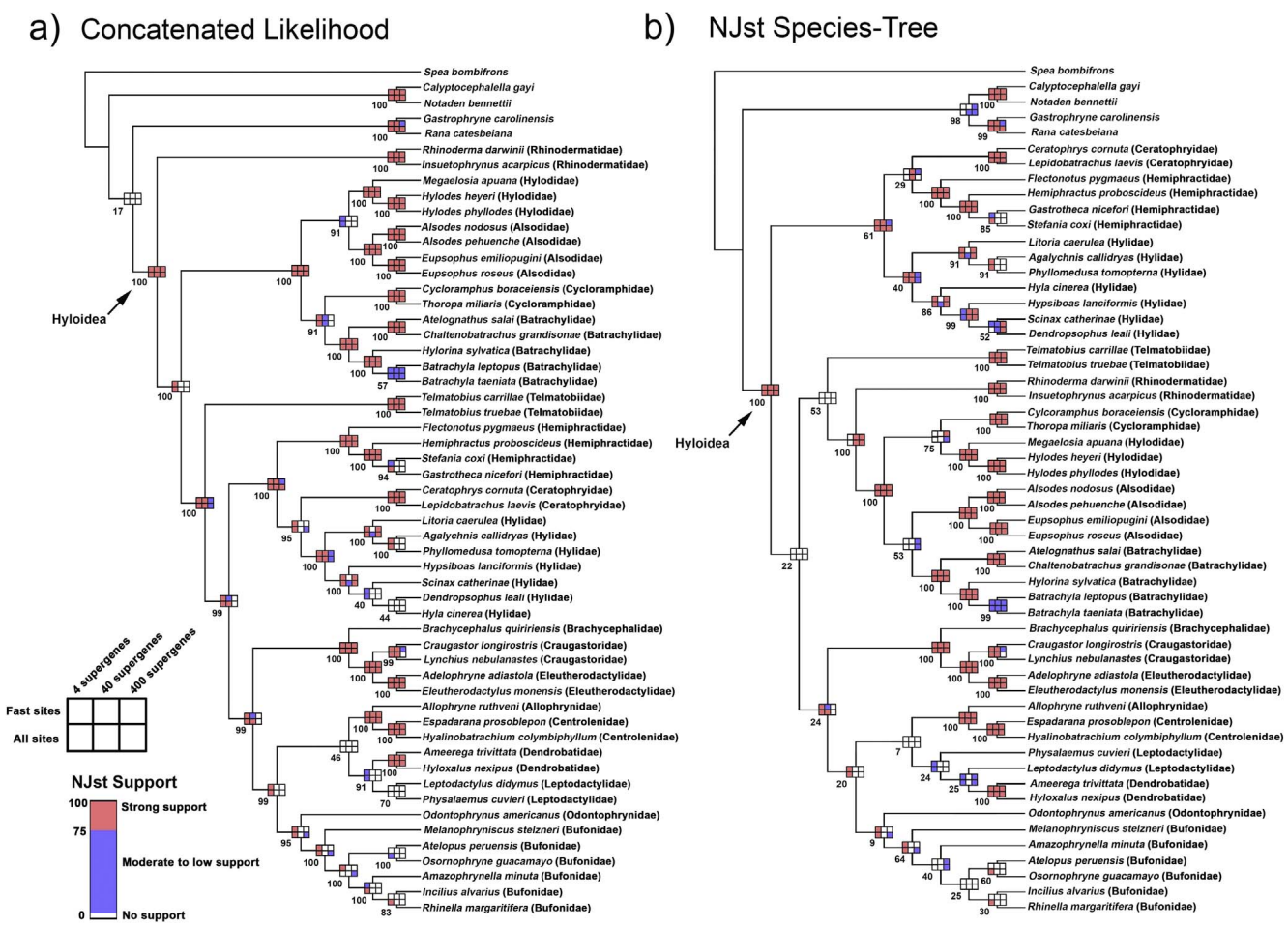

Fig. 4. Two representative phylogenies of hyloid frogs (summarizing support across different approaches), based on (a) concatenated likelihood analysis, including all sites, and (b) NJst analysis, including all sites and with no bins (both trees are based on including loci with up to 60% missing taxa per UCE). Numbers next to each branch indicate branch support. The boxes on both trees indicate congruence and support with naive binning NJst analyses, specifically those with different numbers of supergenes (4, 40, or 400) and either including only the fastest 10% of sites (top row) or including all sites (bottom row). Red boxes indicate support values > 75%, whereas blue boxes indicate values < 75%. White boxes indicate that the clade was not recovered in the specific analysis shown.

(similar to NJst with 40 supergenes). However, the statistical binning analyses gave problematic results in the latter clade, for example, showing non-monophyly of Bufonidae and Leptodactylidae. They also show non-monophyly of Phyllomedusinae. These patterns were also observed with the unweighted statistical binning analysis with NJst (Fig. S19). Thus, weighted statistical binning with ASTRAL gives the highest mean support for unknown relationships of any approach (94; albeit using an alternative support measure from most analyses, LPP), but this included strong support for clades that were seemingly incorrect.

All approaches generally agreed on several aspects of hyloid phylogeny, despite the previously described disagreements. These included the Austral Clade (Alsodidae, Batrachylidae, Cycloramphidae, and Hylodidae), the clade uniting Ceratophryidae, Hemiphractidae, and Hylidae, and the clade uniting Terrarana, Allophrynidae, Centrolenidae, Leptodactylidae, Dendrobatidae, Odontophrynidae, and Bufonidae. Thus, much of the disagreement about hyloid phylogeny in our results was restricted to the relationships among these three multi-family clades and the placement of Rhinodermatidae and Telmatobiidae.

3.2. Secondary analyses

We also performed several additional analyses to address the robustness of the main phylogenetic results (primarily using ML and unbinned NJst analyses). These included: (a) exclusion of loci with different levels of missing data, (b) exclusion of highly incomplete taxa, (c) use of the slowest sites alone, and (d) gene-tree resolution as it related to UCE length. First, we explored the effect of reducing the overall amount of missing data, by excluding loci with up to 40% missing taxa per locus (as opposed to including loci with up to 60% missing taxa per locus, as in the other analyses). These datasets included 49.1% and 55.5% missing data overall (overall proportion of gaps and undetermined characters) for the 40% and 60% datasets, respectively. Using the 40% criterion slightly reduced the percentage of missing data but greatly reduced the number of loci included (from 2214 to 807). Results from the concatenated likelihood and NJst analyses (no supergenes, all sites included) remained topologically similar to the matched analyses with up to 60% missing taxa per locus (Figs. S20, S21, S22). Results were also similar between unbinned ASTRAL analyses with up to 40% vs. 60% missing taxa per locus (Figs. S23, S24).

Results were also similar when nine highly incomplete taxa were removed (those with < 500 UCEs enriched per taxon; Figs. S25, S26). Removal of these taxa increased mean branch support (Table 3), and the overall missing data decreased from 55.5 to 52.2% (compared to the 60% missing dataset with all taxa included). Importantly, the highly incomplete taxa that were removed were ones that were misplaced in some analyses (e.g. *Osornophryne* [Bufonidae], *Phyllomedusa* [Phyllomedusinae]), and which reduced the proportion of known clades that were recovered by a given approach. Nevertheless, many of the

Table 3
Comparison of mean support for secondary analyses contrasting concatenated ML and multigene coalescent methods (ASTRAL and NJst) for different levels of missing data (40% and 60%) and taxon completeness (including vs. excluding taxa with less than 500 UCEs each). Values are average bootstrap support across all clades for a given tree. See Methods for explanation of why incomplete taxa were not excluded for the 40% missing data treatment.

Method	40% missing	60% missing
RAxML	93.19	92.50
RAxML (500+ UCEs)	N/A	95.14
ASTRAL	72.42	76.25
NJst	73.59	75.94
NJst (500+ UCEs)	N/A	83.09

differences between concatenated and NJst unbinned analyses observed in our primary analyses (Fig. 4) persisted when excluding highly incomplete taxa (e.g. the placement of Rhinodermatidae; Figs. S21, S22).

Use of only the slowest 90% of sites (86% of the total dataset, 40% missing, 275,315 out of 318,999 bp; 87% of the total dataset, 60% missing, 698,218 out of 802,167 bp) yielded poor results for concatenated ML (e.g. Figs. S4–S7). Specifically, most relationships among families were only weakly supported, and many well-established groups were non-monophyletic. For example, Bufonidae, Hylidae, Leptodactylidae, Ranoidea, and Terrarana were no longer monophyletic. Nevertheless, some clades remained strongly supported by both analyses, including Ceratophryidae, Dendrobatidae, Hemiphractidae, Telmatobiidae, Allophrynidae + Centrolenidae, and some Austral Clade families (e.g. Alsodidae, Batrachylidae). In summary, mean support decreased dramatically when excluding the 10% fastest sites (43,684 out of 318,999 bp). In contrast, analyses including only the 10% fastest sites and those including all sites were similar in both overall support and topology (Fig. 2b; S20, S21).

Our measures of gene-tree incongruence were higher on average than those of Betancur et al. (2014), likely because of missing taxa between UCE trees and the overall concatenated and species trees. Unlike Betancur et al. (2014), we observed a significant positive relationship between locus length and incongruence ($r^2 = 0.0738$ $P < .001$, both for concatenated and ASTRAL comparisons; Fig. 5). Thus, we interpret these findings as evidence that shorter UCEs were (on average) more similar to the concatenated and coalescent topologies. Although this indicates that sampling bias associated with short UCE contig lengths was not an issue for our coalescent analyses, it does reveal that longer contigs were (on average) somewhat less congruent with the concatenated and coalescent topologies. This pattern might be explained by sequencing-error coverage issues (i.e. smaller read depths with increasing distance from the UCE-anchor) or the distribution of missing taxa in gene trees. This should be explored in future studies.

3.3. Hyloid frog phylogeny

Our preferred estimate of hyloid frog phylogeny corresponds to the concatenated ML analysis, and the NJst and ASTRAL analyses with 4 supergenes (Fig. 6). This topology was chosen based on our criteria for evaluating the performance of different approaches described above (Table 2). This tree is also similar to those from the unbinned ASTRAL analysis and NJst with 40 supergenes, except for the placement of Rhinodermatidae.

Our preferred estimate shows several intriguing similarities and differences with previous estimates. Note that this comparison is not intended to be a comprehensive review of all previous studies of hyloid phylogeny, especially since very few previous studies have included all (or even most) hyloid families. We compared our results to four studies that were relatively comprehensive (Pyron and Wiens, 2011; Pyron, 2014; Feng et al., 2017; Hutter et al., in press).

Two of these studies yielded very similar estimates to each other (Pyron and Wiens, 2011; Pyron, 2014), presumably because they were based on nearly identical data and methods. These studies (Fig. 1d) showed weak support for placing Terrarana (e.g. Brachycephalidae, Craugastoridae, Eleutherodactylidae) as the sister taxon to all other hyloids, with Hemiphractidae then as the sister taxon to the remaining species. Our results (Fig. 6) instead showed strong support for placing Terrarana as the sister taxon of a clade including Bufonidae + Odontophrynidae, Allophrynidae + Centrolenidae, and Dendrobatidae + Leptodactylidae. We also found strong support for placing Hemiphractidae with Hylidae and Ceratophryidae. These previous studies (Pyron and Wiens, 2011; Pyron, 2014) showed weak support for a clade including the Austral Clade families (Alsodidae, Batrachylidae, Cycloramphidae, Hylodidae), but they also included in this clade Ceratophryidae, Odontophrynidae, Telmatobiidae, and Rhinodermatidae. Our analyses showed strong support for placing all four of these latter families elsewhere (although these previous analyses do place Rhinodermatidae with the Austral Clade). Finally, these previous analyses showed support for a clade uniting Allophrynidae and Centrolenidae, in a clade with Dendrobatidae, Leptodactylidae, and Bufonidae. Our analyses were consistent with this clade also, but with the inclusion of Odontophrynidae.

Interestingly, the placement of Odontophrynidae in a clade with Centrolenidae, Dendrobatidae, Leptodactylidae, and Bufonidae was also supported by Roelants et al. (2007; Fig. 1c). That study also placed Rhinodermatidae as the sister taxon to all other Hyloidea, and Ceratophryidae with some Hylidae. However, some aspects of their results disagreed with ours (e.g. non-monophyly of Hylidae, placement of Terrarana with some Hylidae), and that study also lacked many hyloid families included here (e.g. Alsodidae, Cycloramphidae, Hemiphractidae, Hylodidae).

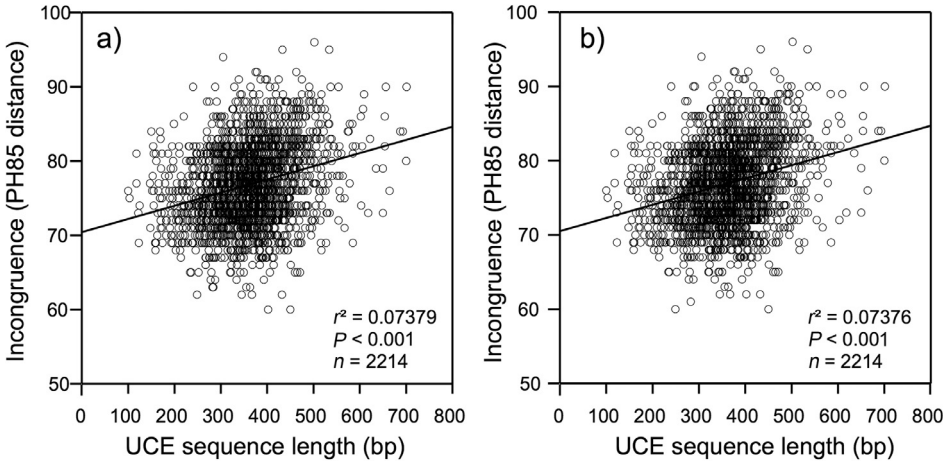

Fig. 5. Relationships between incongruence and sequence length (base pairs) based on the 2214 UCEs used in the 60% missing dataset. Incongruence is measured as the topological distance (PH85) between each UCE tree and (a) a concatenated maximum likelihood tree (Fig. S2) and (b) a ASTRAL multigene coalescent tree (Fig. S8).

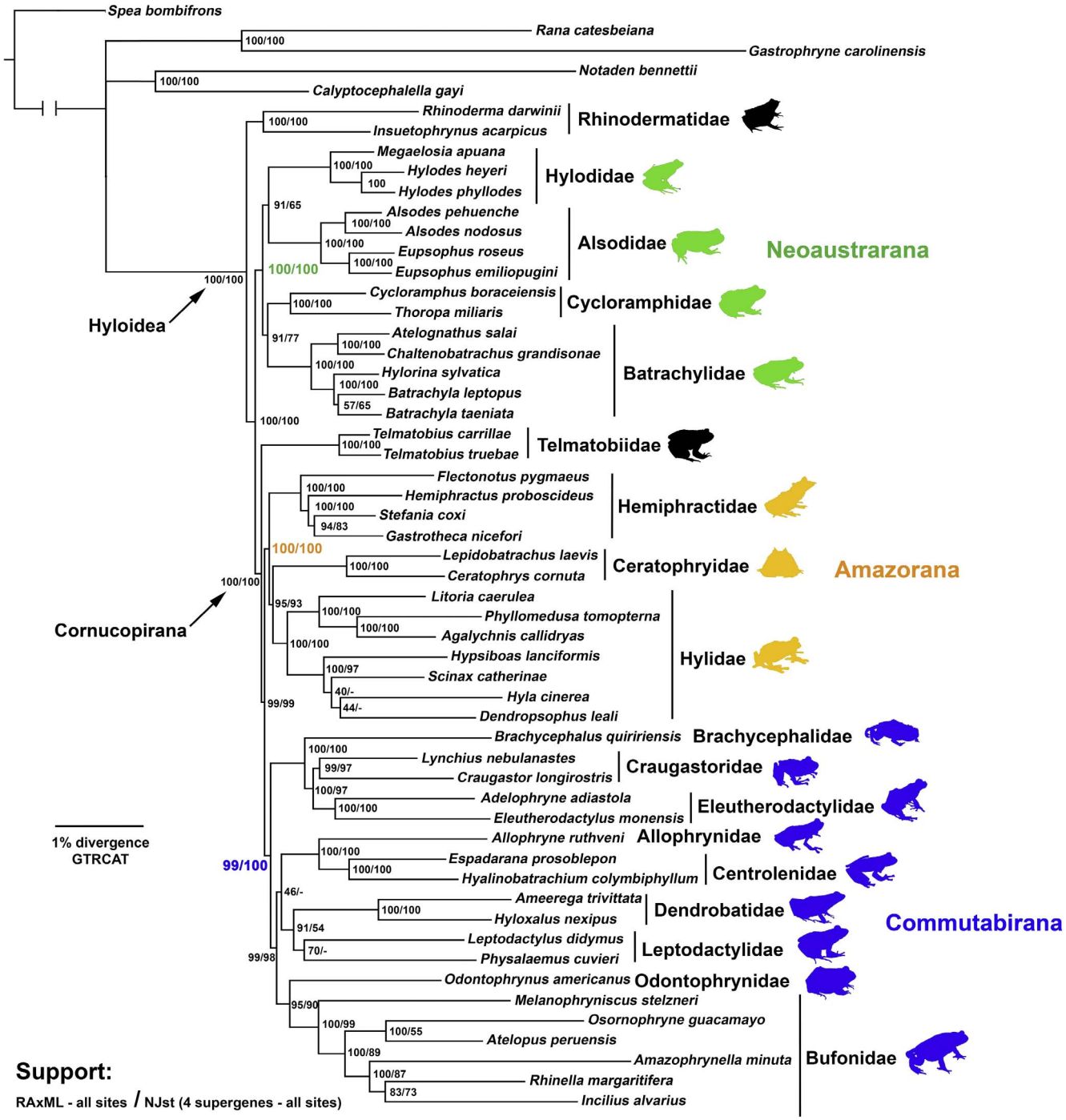

Fig. 6. Preferred phylogeny, based on concatenated likelihood analysis with all sites included, based on 2214 UCEs (802,167 base pairs), including loci with up to 60% missing taxa per UCE. The trees from the ASTRAL and NJst analyses with 4 supergenes (including all sites) have very similar topologies. One branch among the outgroups with low support values was collapsed. The two numbers next to each branch show support from concatenated analysis (left) and from NJst, all-sites, 4 supergenes (right). Dashes to the right of the slash indicate branches that are not supported by the NJst analysis with 4 supergenes and all sites.

Our results were also very dissimilar to those of Hutter et al. (in press). Those authors generated a "backbone" tree among hyloid genera based on 158 genera and 13 nuclear and 7 mitochondrial genes (but with some missing data). This tree showed strong support for a large clade including all hyloids except dendrobatids, hemiphractids, and terraranans (Fig. 1f), with dendrobatids weakly placed as the sister taxon to all other hyloids. Hylids were strongly supported as the sister group to all other members of this large clade. Like Pyron and Wiens (2011) and Pyron (2014), this tree showed weak support for a clade of eight relatively small South American families (Alsodidae, Batrachylidae, Ceratophryidae, Cycloramphidae, Hylodidae, Odontophrynidae, Rhinodermatidae, and Telmatobiidae). The tree of Hutter et al. (in press) also supported a clade uniting Bufonidae, Allophrynidae, Centrolenidae, and Leptodactylidae, similar to Pyron and Wiens (2011) and Pyron (2014), but excluding Dendrobatidae. Very few of these relationships were found in our tree, although we do strongly support placing alsodids, batrachylids, and hylodids in the Austral Clade (along with Cycloramphidae) and a clade including bufonids, allophrynids, centrolenids, and leptodactylids (but also including dendrobatids and odontophrynids in our tree).

Our results were more similar to those of Feng et al. (2017). This makes sense, since their dataset (95 nuclear loci) is the most similar to ours. Their coalescent-based tree (ASTRAL) was generally strongly supported, but lacked several hyloid families (Fig. 1g). Their concatenated tree was consistent with their coalescent tree, but generally weakly supported (Fig. 1h). For taxonomic completeness, we compare our results to this concatenated tree. Our preferred tree is congruent with theirs in placing rhinodermatids at the base of Hyloidea, in placing some members of the Austral Clade near the base (Alsodidae, Batrachylidae, Hylodidae), and in placing Telmatobiidae above those two clades as the sister taxon to the remaining hyloids. Our tree is also congruent with theirs in supporting a clade consisting of hylids, ceratophryids, and hemiphractids, and a clade that unites terraranans, dendrobatids, leptodactylids, bufonids, odontophrynids, allophrynids, and centrolenids.

However, our trees differ dramatically in the placement of the family Cycloramphidae. Our results showed strong support for placing this family in the Austral Clade, with Alsodidae, Batrachylidae, and Hylodidae (Fig. 6). Their tree showed weak support for placing this family as the sister group to Odontophrynidae, in a clade with bufonids, allophrynids, centrolenids, leptodactylids, dendrobatids, and terraranans (Fig. 1).

There were also several less dramatic differences between our preferred tree (Fig. 6) and the tree of Feng et al. (2017). First, we placed Hylodidae with Alsodidae with moderately strong ML support (bs = 91%), whereas Feng et al. (2017) placed Hylodidae with Batrachylidae with weak support. Second, we placed Ceratophryidae and Hylidae as the sister taxon to Hemiphractidae with strong support, whereas Feng et al. (2017) placed Hemiphractidae and Hylidae as the sister taxon to Ceratophryidae (with strong ASTRAL support but weak support from the concatenated analysis). Third, within the clade that includes dendrobatids, terraranans, bufonids, allophrynids, centrolenids, leptodactylids, and odontophrynids, we placed terraranans as the sister taxon to all other families in this clade (with strong support), whereas Feng et al. (2017) placed dendrobatids in this position (with only weak support). We also found strong concatenated support for placing dendrobatids with leptodactylids. Fourth, we found strong support for a clade uniting Bufonidae and Odontophrynidae, whereas Feng et al. (2017) placed Leptodactylidae with Odontophrynidae, not Bufonidae. Overall, despite some important clades shared between our trees, there were actually several differences. Many of these differences may reflect weak support in the tree of Feng et al. (2017), due to the lower number of loci (95 vs. 2214) and the fact that several taxa were included in their concatenated tree based on relatively few genes (e.g. Allophrynidae, Centrolenidae, Cycloramphidae, Hylodidae).

4. Discussion

In this study, we evaluated different approaches for analyzing phylogenomic datasets using new empirical data for hyloid frogs. We then used the best-performing approach to generate a new hypothesis for hyloid relationships. Our results suggest that the optimal approach, based on our three performance criteria, is concatenated likelihood, followed closely by coalescent methods (NJst, ASTRAL) that utilized a small number of naive bins (supergenes). The topologies obtained using these three approaches were also very similar to each other. We also found that results were similar either including all the data or only the fastest 10% of the sites (support values were statistically indistinguishable; Table 4). Thus, even though 86–87% of the sites seemed to contain relatively little useful information, there also seemed to be no negative consequence to including all sites. Finally, our results provide a new estimate of hyloid frog phylogeny that is mostly well supported and that is quite different from most previous estimates. In the sections that follow, we discuss the implications of our results for phylogenomic studies in general, and the implications for frog phylogeny in particular.

Table 4
Results of Mann-Whitney U tests for whether support was significantly different between "all" and "fast" site analyses (see also Fig. 4). Note that none of the tests were significant (also see Fig. 2, C and D).

Dataset	Mann-Whitney U	*P*
4 supergene NJst	1330.5	0.875
40 supergene NJst	1364.5	0.931
400 supergene NJst	1371.0	0.898
Concatenated likelihood RAxML	1313.0	0.754

Table 5
Naive binning strategies for 2214 ultraconserved elements sequenced from hyloid frogs.

Rate*	Number of sites	Number of bins	Bin sizes (base pairs)
All	802,167	4	200,542
Fast	103,949	4	25,987
All	802,167	40	20,054
Fast	103,949	40	2599
All	802,167	400	2005
Fast	103,949	400	260

* Rates were inferred using TIGER (Cummins and McInerney, 2011) to partition site rate variation into 10 categories. Fast sites are those in the 10th or fastest category.

Table 6
Character state versus TIGER-inferred rate category comparison to support the accuracy of this method for UCE data from 2214 loci (802,167 base pairs (bp) total).

	Number of sites with N character states			
Rate categories	1 state	2 state	3 state	4 state
Lower 50% (636,682 bp)	634,643 (99%)	1867 (0.2%)	161 (0.002%)	11 (0.0002%)
Upper 50% (165,485 bp)	0 (0%)	135,291 (82%)	26,828 (16%)	3366 (2%)
Upper 10% (103,949 bp)	0 (0%)	81,209 (78%)	20,120 (19%)	2620 (3%)

4.1. Implications for other phylogenomic studies

Our results suggest that performing a coalescent species-tree analysis with no bins (or with statistical binning) may give suboptimal results when using UCEs. Based on our criteria for method performance, better results can be obtained from concatenated analyses and also from coalescent analyses that used a limited number of naive bins (possibly because this is more similar to concatenation). We found that when using a large number of bins (400) or no bins, NJst performed relatively poorly at recovering well-established relationships, along with providing only weak support for both well-established and unknown relationships. Our results also suggest that using statistical binning methods (here with ASTRAL) can yield strong support for seemingly incorrect relationships. However, we note that we used a single statistical binning threshold (50%) and that the use of different thresholds should be explored in future studies.

Perhaps the most important question arising from this aspect of our study is: how general are these results? Of course, this is an issue for every empirical case study. However, the improved performance from binning (and the relatively strong performance of concatenated analyses by our criteria) does have precedents in simulation studies assessing method accuracy with phylogenomic data (Bayzid and Warnow, 2013). The simulations of Liu et al. (2015) showed that using many bins can reduce phylogenetic accuracy, whereas use of a small number of bins (in their case, 5) might improve accuracy relative to an unbinned analysis (their Fig. 2a). Intriguingly, we also found optimal results from coalescent methods when using the smallest number of bins, rather than many bins or no bins at all (Table 2). However, Liu et al. (2015) used MP-EST rather than NJst or ASTRAL as a coalescent method, and the

relative accuracies of NJst, ASTRAL, and MP-EST are controversial (Mirabab and Warnow, 2015; Edwards et al., 2016; Gatesy et al., 2017). Liu and Edwards (2015) used simulations of 5-taxon trees to suggest that both weighted and unweighted statistical binning may give problematic coalescent estimates. Interestingly, our empirical results also suggest that these methods reduce accuracy (relative to unbinned analyses using the same coalescent method), and that they may give strong support for incorrect relationships.

Furthermore, even though our results here are based on one empirical case study, previous analyses of iguanian lizard relationships also suggested that concatenated analyses might give optimal results for UCE data (Streicher et al., 2016). That study used similar performance criteria to those used here, and also revealed considerable congruence between the concatenated results and the results of some of the coalescent analyses (Streicher et al., 2016). That study did not include binning or exclusion of slow sites, but neither of those approaches appear to be optimal based on our results here. We note that the well-established frog clades utilized here have previously-estimated crown ages that are 28–111 million years old (mean = 61; e.g. Pyron and Wiens, 2013). Similarly, the well-established clades analyzed by Streicher et al. (2016) have estimated crown ages ranging from 22–120 million years old (mean = 49; Zheng and Wiens, 2016). Therefore, based on these time estimates alone, our results should minimally be relevant to the numerous other empirical studies that address phylogenetic problems over these timescales. They might also apply to other timescales as well, but their relevance might also depend on the markers used. Overall, we acknowledge that our conclusions may not apply to every empirical dataset in every group of organisms. Nevertheless, many of our conclusions here do have some support from both simulations and other empirical datasets.

We note that our data may differ from other phylogenomic datasets in some ways. The first is that we included some loci with substantial missing data (although this may be strongly preferable to minimizing missing data and thereby drastically reducing the number of loci; Streicher et al., 2016). Yet, the basic differences between concatenated and unbinned coalescent analyses remained across different levels of missing data (see Figs. S2 and S20, concatenated versus S27 and S22, unbinned NJst, respectively). However, even the most complete datasets analyzed here still had considerable missing data.

Similarly, the basic differences in results were maintained when excluding taxa that were highly incomplete (Figs. S25, S26). The weak performance of some methods appeared to be associated with the incorrect placement of a few highly incomplete taxa (e.g. *Osornophryne, Phyllomedusa;* Figs. S13–S16, S18, S19). Yet, other highly incomplete taxa seemed to be placed correctly by all methods (e.g. *Hyla, Incilius;* Figs. S13–S16, S18, S19), and some taxa were more complete but still misplaced in many analyses (e.g. *Physalaemus*). Our results suggest that some incomplete taxa can be highly unstable, and different approaches vary in how well these unstable taxa are placed. Nevertheless, simulation studies that included no missing data also supported many of our conclusions (see above). Furthermore, simply excluding all taxa with missing data seems problematic and unwarranted.

Another question is the extent to which our results apply to other coalescent-based methods. Of course, binning analyses were initially designed to allow the use of more computationally intensive methods (e.g. *BEAST; Heled and Drummond, 2010) on larger datasets (Bayzid and Warnow, 2013). Evaluating our results on every method would be difficult, especially our comparisons of binned vs. unbinned analyses and our comparisons of different bin sizes (i.e. analyses with large numbers of bins/supergenes would be difficult). Nevertheless, this could be a useful area for future research.

Another important issue is whether our empirical analyses of method performance might be biased. Two of our criteria focused on the ability of methods to recover and strongly support clades that are relatively well established, based on the congruence between molecular and morphological data in previous studies. Therefore, a relevant question is whether these clades represent an unbiased sample of all clades throughout the tree. In particular, branches that are well supported by both morphological and molecular data may be longer than those that we are most interested in resolving (i.e. short branches with only weak support in earlier studies). Previous studies suggest that shorter branches are: (1) more likely to be poorly supported and have more conflicts among genes (Wiens et al., 2008, 2012), (2) be masked by non-phylogenetic signal (Philippe et al., 2011), and (3) have more frequent conflicts between estimates from concatenated and coalescent methods (Lambert et al., 2015). Thus, focusing on the ability of methods to recover longer, "easier" branches might give a biased picture of relative method performance for evaluating the shorter, harder branches. On the other hand, we found that across different approaches, the support for established and unknown clades was very strongly correlated (Fig. 3). Overall, it seems that approaches that cannot recover and strongly support these well-established clades are clearly problematic, but we acknowledge that approaches that can strongly resolve well-established clades are not guaranteed to resolve unknown clades. This may be a fruitful question for future simulation studies.

Finally, our results show that most sites used in our dataset (and possibly others) contributed little to resolving phylogenetic questions (Fig. 2b). Our results also show that there may be little harm in including these sites (Fig. 2c and d; Table 4). Yet, they also suggest that we might have been able to obtain similar results with only a small fraction of the data. Therefore, one important implication of our results is that it might be useful to focus future studies only on those UCE loci that yield the largest proportion of fast evolving sites. However, we also acknowledge that (1) the categorization of sites as fast or slow depends on the overall set of sequences included in a study: one study's "slow" sites might be another study's "fast" sites and (2) examining the distribution of parsimony-informative sites among fast and slow categories would be a useful exercise to explain the limited contribution of most sites used in our dataset.

4.2. Implications for frog phylogeny

Our results (Fig. 6) suggest a hyloid frog phylogeny that is quite different from previous estimates (Fig. 1; Darst and Cannatella, 2004; Frost et al., 2006; Roelants et al., 2007; Pyron and Wiens, 2011; Zhang et al., 2013; Pyron, 2014; Hutter et al., in press; Feng et al., 2017). An obvious question to ask is: why should anyone believe our results instead of these previous estimates? There are several reasons. First, most previous estimates of relationships among hyloid frog families were only weakly supported and were inconsistent with each other (Fig. 1). In contrast, our analyses revealed strong support for several clades, and many of these clades remained well-supported across diverse analyses (e.g. the Austral Clade, the clade consisting of Ceratophryidae, Hemiphractidae, and Hylidae; the clade comprising Bufonidae, Dendrobatidae, and their closest relatives). Second, many previous studies had less comprehensive taxon sampling of hyloid families (e.g. Darst and Cannatella, 2004; Roelants et al., 2007; Zhang et al., 2013). Third, most previous analyses were based on either mitochondrial data alone or combined mitochondrial and nuclear data, and may have been heavily influenced by the mitochondrial markers. Here, our analyses are based on thousands of nuclear loci. Fourth, although our data matrix contained considerable missing data, the matrices of Pyron and Wiens (2011) and Pyron (2014) actually had considerably more (~80% in those studies vs. ~55% or less here). Therefore, overall levels of missing data should not be an argument for favoring those trees over the one here (but note that there was no evidence of negative effects of missing data on those analyses; Pyron and Wiens, 2011). Finally, our results show several points of congruence with a recent phylogenomic analysis of higher-level frog relationships (based on up to 95 loci), which also supports a very different hyloid phylogeny than previous estimates (Feng et al., 2017).

We take this opportunity to propose new names for some of the

more well-supported clades of families (note that taxa above the family level do not require formal diagnoses). Even if these clades prove to be incorrect in the future, these names at least allow us to reference these groups. First, we propose the name Neoaustrarana for the Austral Clade described above, containing the families Alsodidae, Batrachylidae, Cycloramphidae, and Hylodidae. This name is a combination of the prefix *neo* meaning new, the Latin *australis* meaning southern, and the Latin *rana* meaning frog. This refers to the southern New World distribution of this clade. Second, we propose the name Cornucopirana for all hyloid families exclusive of Rhinodermatidae and Neoaustrarana. This name is a combination of the Latin *cornu copiae* meaning horn of plenty (a reference to the 'abundance' of diversity contained within this clade) and the Latin *rana* meaning frog. We propose the name Amazorana for the clade of Ceratophryidae, Hemiphractidae, and Hylidae (a clade also found by Feng et al., 2017). This name is a combination of the word Amazonia (where much of the clade's diversity occurs; AmphibiaWeb, 2016) and the Latin *rana* meaning frog. We propose the name Commutabirana for the clade including Terrarana and the families Allophrynidae, Bufonidae, Centrolenidae, Dendrobatidae, Leptodactylidae, and Odontophrynidae. This name is a combination of the Latin *commutabilis* meaning variable and the Latin *rana* meaning frog. This name references the variable natural histories and reproductive strategies observed in this clade (including direct developers, foam nesters, and species with parental care; Gomez-Mestre et al., 2012).

Finally, we note one intriguing implication of our results for the biogeography of hyloid frogs. Similar to Feng et al. (2017), our preferred trees (which place Rhinodermatidae as the sister taxon to all hyloids), strongly suggest that hyloids initially evolved in southern South America. This is where rhinodermatids occur, along with many members of the Neoaustrarana (i.e. Alsodidae, Batrachylidae), which is the sister group to all other hyloids exclusive of Rhinodermatidae. Other members of the Neoaustrarana also occur in southern South America, specifically in southeastern Brazil (Cycloramphidae, Hylodidae). Telmatobiidae is the sister group to all other hyloids, exclusive of Rhinodermatidae and the Neoaustrarana. Telmatobiids are also found in cooler climates, occurring at higher elevations in the Andes from Ecuador to Argentina and Chile (AmphibiaWeb, 2016). Most other hyloids occur more broadly in the Neotropical region (with hylids and bufonids occurring globally). Thus, our results suggest that the early evolution of hyloids occurred in regions of southern South America that currently have temperate climates (and the nearby highlands of southeastern Brazil), before they radiated across tropical regions of South America (including the Amazon Basin). We also note that an origin in southern South America is consistent with the putative sister group of hyloids, the clade including the Australian Myobatrachidae and the southern South American family Calyptocephalellidae. Future analyses should test this idea with time-calibrated phylogenies and explicit methods for biogeographic reconstruction (e.g. LaGrange and related methods; Ree and Smith, 2008). Our biogeographic findings (coupled with those of Feng et al. (2017)) are inconsistent with those of Pyron (2014), because the topology of that study placed these southern South American clades far from the base of Hyloidea, and explicitly reconstructed hyloids as originating in northern, tropical South America.

4.3. Implications for future phylogenomic studies in amphibians

Our study is among the first to report capture success rates for UCEs across multiple amphibian families. Among the 5040 UCEs targeted, we successfully captured 3429, with 2843 of these UCEs enriched for 10 or more species. The mean number of UCEs captured per species was 1477 (S.E. ± 92.00), and we captured > 2000 UCEs for several species (Table 1). In many cases, we captured nearly twice the number predicted based on *Xenopus tropicalis* (Faircloth et al., 2012). Thus, our demonstration that 68% of the targeted UCEs were captured is a promising result for the utility of this probe set in amphibians. In light of this, we provide a list of the successfully captured UCEs as a supporting document (Table S2) for use in future targeted sequence-capture studies of amphibians (particularly frogs) that may wish to refine the UCE probes we used herein.

Acknowledgements

We are indebted to Lucas Barrientos for laboratory assistance and Christian Cox for helpful discussion regarding methods. We thank Jill Castoe and Kimberly Bowles at the University of Texas at Arlington Genomics Core Facility for their assistance with MiSeq runs. We thank the many curators, collectors, and naturalists that donated tissues that were used in this study (many of whom contributed tissues to Wiens et al., 2005), but especially Rafael de Sá, Tod Reeder, and Miguel Trefaut Rodrigues (see Table S1). Claudio Correa thanks Javiera Cisternas and Álvaro Zúñiga for their help obtaining some samples. We thank Katrina Dlugosch and Michael Worobey for access to laboratory equipment. Peter Foster kindly assisted with access to the Natural History Museum clusters. We used an online R tutorial written by Katie Everson to construct saturation plots. We thank A. Larson and two anonymous reviewers for comments on the manuscript. For financial support we thank grants from Chile's Fondecyt 1160583 (to Pablo Guerrero) and FONDECYT 79130032 (to Claudio Correa), an NSF Graduate Fellowship (U.S.A.) to E.C. Miller, NSF grant DEB 1655690 to J.J. Wiens (U.S.A.), and the University of Arizona.

Appendix A. Supplementary material

Supplementary Tables S1–S2. Supplementary Figures S1–S27. Command Summaries 1–4 (Available at http://datadryad.org). Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2017.10.013.

References

AmphibiaWeb: Information on amphibian biology and conservation. (web application). 2016. Berkeley, California: AmphibiaWeb. Available: < http://amphibiaweb.org/ > . (accessed: 05 July 2016).
Bayzid, M.S., Warnow, T., 2013. Naive binning improves phylogenomic analyses. Bioinformatics 29, 2277–2284.
Bayzid, M.S., Mirabab, S., Boussau, B., Warnow, T., 2015. Weighted statistical binning: enabling statistically consistent genome-scale phylogenetic analyses. PLoS ONE 10, e0129183.
Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., Haussler, D., 2004. Ultraconserved elements in the human genome. Science 304, 1321–1325.
Betancur, R.R., Naylor, G.J.P., Orti, G., 2014. Conserved genes, sampling error, and phylogenomic inference. Syst. Biol. 63, 257–262.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.
Cummins, C., McInerney, J.O., 2011. A method for inferring the rate of evolution of homologous characters that can potentially improve phylogenetic inference, resolve deep divergence and correct systematic biases. Syst. Biol. 60, 833–844.
Darst, C.R., Cannatella, D.C., 2004. Novel relationships among hyloid frogs inferred from 12S and 16S mitochondrial DNA sequences. Mol. Phylogenet. Evol. 31, 462–475.
De Lisle, S.P., Rowe, L., 2015. Independent evolution of the sexes promotes amphibian diversification. Proc. R. Soc. Lond. B 282, 20142213.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nuc. Acid. Res. 32, 1792–1797.
Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M., Lemmon, A.R., Leaché, A.D., Liu, L., Davis, C.C., 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94, 447–462.
Faircloth, B.C., 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32, 786–788.
Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61, 717–726.
Faircloth, B.C., 2013. Illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. http://dx.doi.org/10.6079/J9ILL.
Feng, Y.-J., Blackburn, D.C., Liang, D., Hillis, D.M., Wake, D.B., Cannatella, D.C., Zhang, P., 2017. Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous-Paleogene boundary. Proc. Natl. Acad. Sci. USA 114, E5864–E5870.
Ford, L.S., Cannatella, D.C., 1993. The major clades of frogs. Herpetol. Mon. 7, 94–117.

Frost, D.R., Grant, T., Faivovich, J., Bain, R.H., Haas, A., Haddad, C.F.B., de Sá, R.O., Channing, A., Wilkinson, M., Donnellan, S.C., Raxworthy, C.J., Campbell, J.A., Blotto, B.L., Moler, P., Drewes, R.C., Nussbaum, R.A., Lynch, J.D., Green, D.M., Wheeler, W.C., 2006. The amphibian tree of life. Bull. Am. Nat. Hist. Mus. 297, 1–370.

Gatesy, J., Meredith, R.W., Janecka, J.E., Simmons, M.P., Murphy, W.J., Springer, M.S., 2017. Resolution of a concatenation/coalescence kerfuffle: partitioned coalescence support and a robust family-level tree for Mammalia. Cladistics 33, 295–332.

Gomez-Mestre, I., Pyron, R.A., Wiens, J.J., 2012. Phylogenetic analyses reveal unexpected patterns in the evolution of reproductive modes in frogs. Evolution 66, 3687 3687–1700.

Hansmann, S., Martin, W., 2000. Phylogeny of 33 ribosomal and six other proteins encoded in an ancient gene cluster that is conserved across prokaryotic genomes: influence of excluding poorly alignable site from analysis. Int. J. Syst. Evol. Microbiol. 50, 1655–1663.

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

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580.

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L., Kimball, R.T., 2016. Phylogenomic inference of landfowl (Aves: Galliformes): effects of missing data in concatenated and coalescent inference and evidence for a bias in gene tree reconciliation approaches. Mol. Biol. Evol. 33, 1110–1125.

Hutter, C.R., Lambert, S.M., Wiens, J.J., 2017;al., in press. Rapid diversification and time explain amphibian species richness at different scales in the Tropical Andes, Earth's most biodiverse hotspot. Am. Nat (in press).

Lambert, S.M., Reeder, T.W., Wiens, J.J., 2015. When do species tree and concatenated estimates disagree? An empirical analysis with higher-level scincid lizard phylogeny. Mol. Phylogenet. Evol. 82, 146–155.

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

Liu, L., Edwards, S.V., 2015. Comment on "Statistical binning enables an accurate coalescent-based estimation of the avian tree". Science 350, 171.

Liu, L., Xi, Z., Wu, S., Davis, C.C., Edwards, S.V., 2015. Estimating phylogenetic trees from genome-scale data. Ann. NY Acad. Sci. 1360, 36–53.

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

Meiklejohn, K.A., Faircloth, B.C., Glenn, T.C., Kimball, R.T., Braun, E.L., 2016. Analysis of a rapid evolutionary radiation using ultraconserved elements: evidence for a bias in some multispecies coalescent methods. Syst. Biol. 65, 612–627.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE). IEEE, New Orleans (LA), pp. 1–8.

Mirabab, S., Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 15, 44–52.

Mirabab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014a. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30, i541–i548.

Mirabab, S., Bayzid, M.S., Boussau, B., Warnow, T., 2014b. Statistical binning enables an accurate coalescent-based estimation of the avian tree. Science 346, 1250463.

Moen, D.S., Wiens, J.J., 2017. Microhabitat and climatic-niche change explain patterns of diversification among frog families. Am. Nat. 190, 29–44.

Moen, D.S., Morlon, H., Wiens, J.J., 2016. Testing convergence versus history: convergence dominates phenotypic evolution for over 150 million years in frogs. Syst. Biol. 65, 146–160.

Paradis, E., Claude, J., Strimmer, K., 2004. APE: analysis of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.

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

Penny, D., Hendy, M.D., 1985. The use of tree comparison metrics. Syst. Zool. 34, 75–82.

Philippe, H., Sörhannus, U., Baroin, A., Perasso, R., Gasse, F., Adoutte, A., 1994. Comparison of molecular and paleontological data in diatoms suggests a major gap in the fossil record. J. Evol. Biol. 7, 247–265.

Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, T.J., Manuel, M., Wörheide, G., Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9, e1000602.

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

Pyron, R.A., Wiens, J.J., 2011. A large-scale phylogeny of Amphibia including over 2,800 species, and a revised classification of extant frogs, salamanders, and caecilians. Mol. Phylogenet. Evol. 61, 543–583.

Pyron, R.A., Wiens, J.J., 2013. Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. Proc. R. Soc. Lond. B 280, 20131622.

R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria3-900051-07-0. http://www.R-project.org/.

Rasmussen, M.D., Kellis, M., 2007. Accurate gene-tree reconstruction by learning gene- and species-specific substitution rates across multiple complete genomes. Genome Res. 17, 1932–1942.

Ree, R.H., Smith, S.A., 2008. Maximum-likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57, 4–14.

Roelants, K., Gower, D.J., Wilkinson, M., Loader, S.P., Guillaume, K., Moriau, L., Bossuyt, F., 2007. Global patterns of diversification in the history of modern amphibians. Proc. Natl. Acad. Sci. USA 104, 887–892.

Salichos, L., Rokas, A., 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 440, 341–345.

Sayyari, E., Mirabab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33, 1654–1668.

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

Shaw, T., Ruan, Z., Glenn, T., Liu, L., 2013. STRAW: species TRee analysis Web Server. Nuc. Acid. Res. 41, W230–W241.

Simmons, M.P., Gatesy, J., 2016. Biases of tree-independent-character-subsampling methods. Mol. Phylogenet. Evol. 100, 424–443.

Springer, M.S., Gatesy, J., 2016. The gene tree delusion. Mol. Phylogenet. Evol. 94, 1–33.

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

Streicher, J.W., Wiens, J.J., 2016. Phylogenomic analyses reveal novel relationships among snake families. Mol. Phylogenet. Evol. 100, 160–169.

Streicher, J.W., Wiens, J.J., 2017. Phylogenomic analyses of more than 4,000 nuclear loci resolve the origin of snakes among lizard families. Biol. Lett. 13, 20170393.

Streicher, J.W., Schulte II, J.A., Wiens, J.J., 2016. How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in iguanian lizards. Syst. Biol. 65, 128–145.

Sukumaran, J., Holder, M.T., 2010. DendroPy: a python library for phylogenetic computing. Bioinformatics 24, 1569–1571.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Wiens, J.J., 2007. Global patterns of species richness and diversification in amphibians. Am. Nat. 170, S86–S106.

Wiens, J.J., 2011. Re-evolution of lost mandibular teeth in frogs after more than 200 million years, and re-evaluating Dollo's law. Evolution 65, 1283–1296.

Wiens, J.J., Tiu, J., 2012. Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling. PLoS ONE 7, 42925.

Wiens, J.J., Fetzner, J.W., Parkinson, C.L., Reeder, T.W., 2005. Hylid frog phylogeny and sampling strategies for speciose clades. Syst. Biol. 54, 719–748.

Wiens, J.J., Kuczynski, C.A., Smith, S.A., Mulcahy, D., Sites Jr, J.W., Townsend, T.M., Reeder, T.W., 2008. Branch length, support, and congruence: testing the phylogenomic approach with 20 nuclear loci in snakes. Syst. Biol. 57, 420–431.

Wiens, J.J., Hutter, C.R., Mulcahy, D.G., Noonan, B.P., Townsend, T.M., Sites Jr, J.W., Reeder, T.W., 2012. Resolving the phylogeny of lizards and snakes (Squamata) with extensive sampling of genes and species. Biol. Lett. 8, 1043–1046.

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assemble using de Bruijn graphs. Genome Res. 18, 821–829.

Zhang, P., Liang, D., Mao, R., Hillis, D.M., Wake, D.B., Cannatella, D.C., 2013. Efficient sequencing of anuran mtDNAs and a mitogenomic exploration of the phylogeny and evolution of frogs. Mol. Biol. Evol. 30, 1899–1915.

Zheng, Y., Wiens, J.J., 2016. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4,162 species. Mol. Phylogenet. Evol. 94, 537–547.