Molecular Phylogenetics and Evolution 61 (2011) 543–583

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians

R. Alexander Pyron [a,*], John J. Wiens [b]

[a] Dept. of Biological Sciences, The George Washington University, 2023 G St. NW, Washington, DC 20052, United States
[b] Dept. of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, United States

ARTICLE INFO

Article history:
Received 7 May 2011
Revised 10 June 2011
Accepted 12 June 2011
Available online 23 June 2011

Keywords:
Amphibia
Anura
Apoda
Caudata
Lissamphibia
Gymnophiona
Phylogeny
Supermatrix
Systematics

ABSTRACT

The extant amphibians are one of the most diverse radiations of terrestrial vertebrates (>6800 species). Despite much recent focus on their conservation, diversification, and systematics, no previous phylogeny for the group has contained more than 522 species. However, numerous studies with limited taxon sampling have generated large amounts of partially overlapping sequence data for many species. Here, we combine these data and produce a novel estimate of extant amphibian phylogeny, containing 2871 species (~40% of the known extant species) from 432 genera (~85% of the ~500 currently recognized extant genera). Each sampled species contains up to 12,712 bp from 12 genes (three mitochondrial, nine nuclear), with an average of 2563 bp per species. This data set provides strong support for many groups recognized in previous studies, but it also suggests non-monophyly for several currently recognized families, particularly in hyloid frogs (e.g., Ceratophryidae, Cycloramphidae, Leptodactylidae, Strabomantidae). To correct these and other problems, we provide a revised classification of extant amphibians for taxa traditionally delimited at the family and subfamily levels. This new taxonomy includes several families not recognized in current classifications (e.g., Alsodidae, Batrachylidae, Rhinodermatidae, Odontophrynidae, Telmatobiidae), but which are strongly supported and important for avoiding non-monophyly of current families. Finally, this study provides further evidence that the supermatrix approach provides an effective strategy for inferring large-scale phylogenies using the combined results of previous studies, despite many taxa having extensive missing data.

1. Introduction

With over 6800 known species (AmphibiaWeb; http://www.amphibiaweb.org/, accessed April, 2011; hereafter "AW") the extant amphibians (frogs, salamanders, and caecilians) are one of the most diverse radiations of terrestrial vertebrates. The number of known extant amphibians has increased rapidly in recent years, with over 2700 species (~40%) described in the last 26 years (Duellman, 1999; Lannoo, 2005). This newly discovered diversity includes dozens of new species from known genera in poorly studied tropical regions such as Madagascar (Vieites et al., 2009), but also new genera in relatively well-explored regions such as the southeastern United States (Camp et al., 2009), and even new families such as Nasikabatrachidae (Biju and Bossuyt, 2003). Unfortunately, much extant amphibian diversity is currently under extreme threat from pressures such as habitat loss, global climate change, and infectious disease, and many species have gone extinct in the last few decades (Blaustein and Wake, 1990; Stuart et al., 2004).

A phylogenetic framework is critical for discovering, understanding, and preserving extant amphibian diversity, but a large-scale phylogeny for extant amphibians is presently lacking. However, recent molecular and combined-data studies have made important contributions to higher-level phylogeny (Frost et al., 2006; Roelants et al., 2007; Wiens, 2007a, 2011) and to the phylogeny of many major groups, such as caecilians (San Mauro et al., 2009; Zhang and Wake, 2009b), hyloid frogs (Darst and Cannatella, 2004), ranoid frogs (e.g., Bossuyt et al., 2006; Wiens et al., 2009), microhylid frogs (van der Meijden et al., 2007), bufonid frogs (Pauly et al., 2004; Pramuk et al., 2008; Van Bocxlaer et al., 2009), centrolenid frogs (Guayasamin et al., 2009), dendrobatid frogs (Grant et al., 2006; Santos et al., 2009), hemiphractid frogs (Wiens et al., 2007a), hylid frogs (Faivovich et al., 2005, 2010; Wiens et al., 2005b, 2010), terraranan frogs (Hedges et al., 2008; Heinicke et al., 2009), and salamanders (Kozak et al., 2009; Vieites et al., 2011; Wiens et al., 2005a, 2007b; Zhang and Wake, 2009a).

The largest estimate of extant amphibian phylogeny to date is that of Frost et al. (2006). Those authors reconstructed amphibian phylogeny based on relatively intensive sampling of species (522) and characters (up to 4.9 kb of sequence data from 2 mitochondrial and 5 nuclear genes [mean = 3.5 kb], and 152 morphological

* Corresponding author.
E-mail addresses: rpyron@colubroid.org (R. Alexander Pyron), wiensj@life.bio.sunysb.edu (John J. Wiens).

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.06.012

characters). Those authors also proposed extensive changes in taxonomy, especially for taxa delimited at the family and genus level. However, that study has also been criticized on numerous grounds, including concerns about taxon sampling and methodological strategies (Marjanović and Laurin, 2007; Pauly et al., 2009; Wiens, 2007b, 2008). For example, those authors collected up to ~4900 characters per species, but their analysis is apparently based on 15,320 characters, suggesting that their controversial approach to sequence alignment (POY) dominates their results (Wiens, 2008). Although some of the changes made by Frost et al. (2006) have been widely adopted, others are more controversial, such as the partitioning of *Bufo* and *Rana* (Marjanović and Laurin, 2007; Pauly et al., 2009; AW). Indeed, many of these changes are no longer supported, even in Frost's (2011) taxonomic database of extant amphibians (e.g., the families Amphignathodontidae, Batrachophrynidae, Cryptobatrachidae, and Thoropidae recognized by Frost et al. (2006)). Much of the most unstable taxonomy involves the family-level assignment of many of the genera of hyloid frogs, particularly those traditionally assigned to the family Leptodactylidae.

Clearly, extant amphibian phylogeny and classification is still in need of additional study. Fortunately, the numerous studies referenced above (and many others) have produced a massive amount of data that are potentially suitable for a combined, supermatrix approach (e.g., de Queiroz and Gatesy, 2007; Driskell et al., 2004; Pyron et al., 2011; Thomson and Shaffer, 2010; Wiens et al., 2005b). This includes thousands of species represented in GenBank for numerous nuclear and mitochondrial genes, often with substantial overlap of genes among species.

Here, we present a large-scale estimate of amphibian phylogeny, including 2871 species (42% of the 6807 known, extant amphibian species) from 432 of the 504 currently recognized genera (86%), and representatives from every currently delimited, extant family and subfamily. This is 5.5 times more species and nearly twice as many genes as the largest previous study (Frost et al., 2006). The data matrix includes up to 12,712 bp for each species from 12 genes (three mitochondrial, nine nuclear). Importantly, rather than simply reanalyzing published data for relatively well-studied families (e.g., dendrobatids, hylids), we address the monophyly and relationships of many smaller groups that have not been the subject of focused studies (e.g., Ceratophryidae, Cylcoramphidae), as well as relationships among families. We produce a revised classification of extant amphibians, focusing on taxa traditionally ranked as families and subfamilies. This study also provides additional support for the value of the supermatrix approach to large-scale phylogenetic inference (e.g., de Queiroz and Gatesy, 2007; Driskell et al., 2004; Pyron et al., 2011; Thomson and Shaffer, 2010; Wiens et al., 2005b).

2. Materials and methods

2.1. Taxonomic reference

This analysis has been several years in the making. Our initial taxonomy was based on the September 2009 update of the AmphibiaWeb (AW) database. However, when we refer to current numbers, these are taken from the April, 2011 update. The AW list is fairly current in terms of recently described species, but more conservative than the Amphibian Species of the World (Frost, 2011; hereafter "ASW") regarding some of the more controversial of the recent taxonomic changes (e.g., *Bufo* and *Rana* maintain similar composition as they did prior to Frost et al., 2006). We note some instances where recent updates have modified our original classification. Note that even when not made explicit, we refer in all instances to the known extant diversity of Lissamphibia, given that the clade Amphibia includes numerous extinct stem-group members that are not lissamphibians. The gymnophionans, caudates, and anurans also contain numerous extinct taxa, many of which are grouped in separate genera, subfamilies, and families that are not addressed in our analyses or included in our discussion of phylogeny. See Marjanović and Laurin (2007), Carroll (2009), and Pyron (2011) for an overview of these taxa, their phylogenetic affinities, and the origins of Amphibia and Lissamphibia.

2.2. Molecular data

We identified 12 candidate loci that have been broadly sampled and successfully used in amphibian phylogenetics at both lower and higher taxonomic levels. These 12 genes included nine nuclear genes: C-X-C chemokine receptor type 4 (CXCR4), histone 3a (H3A), sodium–calcium exchanger (NCX1), pro-opiomelanocortin (POMC), recombination-activating gene 1 (RAG1), rhodopsin (RHOD), seventh-in-absentia (SIA), solute-carrier family 8 (SLC8A3), and tyrosinase (TYR). Three mitochondrial genes were also included: cytochrome *b* (cyt-b), and the large and small subunits of the mitochondrial ribosome genes (12S/16S; omitting the adjacent tRNAs as they were difficult to align and represented only a small amount of data). This selection of genes includes almost all of those genes used in the higher-level analyses by Frost et al. (2006) and Roelants et al. (2007), and most of those used in other large-scale studies (Faivovich et al., 2005; Grant et al., 2006; Wiens et al., 2009, 2010). However, we did not include the nuclear gene 28S (used by Frost et al. (2006)), as previous analyses of this gene region alone suggest that it contains relatively few informative characters and supports some relationships that are grossly inconsistent with other studies (see Wiens et al., 2006). We conducted GenBank searches by family and subfamily to gather all available sequences, using a minimum-length threshold of 200 bp (a somewhat arbitrary threshold of 1.5% of the total matrix length, to avoid including very short [e.g., <50 bp] fragments), and stopping in August of 2010. Only species in the taxonomic database were included in the sequence matrix, which excluded numerous named taxa of ambiguous status, and many sequences labeled 'sp.' We removed a few (<10) taxa with identical sequence data for all genes (arbitrarily retaining the first in alphabetical order), to avoid potentially misidentified or otherwise confounded specimens or sequences.

For the protein-coding genes, alignment was relatively straightforward. Conceptual translations were used to ensure an open reading frame, and sequences were aligned using the translation-alignment algorithm in the program Geneious v4.8.4 (GeneMatters Corp.), with the default cost matrix (Blosum62) and gap penalties (open = 12, extension = 3). For the ribosomal RNA sequences (12S and 16S sequences), alignment was more challenging. Preliminary global alignments using the MUSCLE (Edgar, 2004) and CLUSTAL (Larkin et al., 2007) algorithms under a variety of gap-cost parameters yielded low-quality results (i.e., alignments with large numbers of gaps and little overlap of potentially homologous characters).

We subsequently employed a two-step strategy for these data. First, we identified sequence clusters of similar length and coverage from the global alignment. These were subsequently aligned separately using the MUSCLE algorithm with the default high-accuracy parameters, which have been shown to outperform CLUSTAL in a variety of settings (Edgar, 2004). These alignments were subsequently refined using the MUSCLE refinement algorithm, and then adjusted manually and trimmed for quality and maximum coverage (i.e., end sequences with low overlap and poor apparent alignment were deleted using the alignment editor in Geneious). These length and position-based sequence groups were then aligned to each other using the profile-profile alignment algorithm in MUSCLE. The resulting final global alignment was