Molecular Phylogenetics and Evolution 53 (2009) 492–508

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Higher-level salamander relationships and divergence dates inferred from complete mitochondrial genomes

Peng Zhang [a,b,*], David B. Wake [a,*]

[a] Department of Integrative Biology, Museum of Vertebrate Zoology, 3101 Valley Life Sciences Building, University of California, Berkeley, CA 94720-3160, USA
[b] Key Laboratory of Gene Engineering of the Ministry of Education, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, People's Republic of China

ARTICLE INFO

Article history:
Received 19 February 2009
Revised 7 July 2009
Accepted 7 July 2009
Available online 10 July 2009

Keywords:
Salamander
Phylogeny
Mitochondrial genome
Molecular dating

ABSTRACT

Phylogenetic relationships among the salamander families have been difficult to resolve, largely because the window of time in which major lineages diverged was very short relative to the subsequently long evolutionary history of each family. We present seven new complete mitochondrial genomes representing five salamander families that have no or few mitogenome records in GenBank in order to assess the phylogenetic relationships of all salamander families from a mitogenomic perspective. Phylogenetic analyses of two data sets—one combining the entire mitogenome sequence except for the D-loop, and the other combining the deduced amino acid sequences of all 13 mitochondrial protein-coding genes—produce nearly identical well-resolved topologies. The monophyly of each family is supported, including the controversial Proteidae. The internally fertilizing salamanders are demonstrated to be a clade, concordant with recent results using nuclear genes. The internally fertilizing salamanders include two well-supported clades: one is composed of Ambystomatidae, Dicamptodontidae, and Salamandridae, the other Proteidae, Rhyacotritonidae, Amphiumidae, and Plethodontidae. In contrast to results from nuclear loci, our results support the conventional morphological hypothesis that Sirenidae is the sister-group to all other salamanders and they statistically reject the hypothesis from nuclear genes that the suborder Cryptobranchoidea (Cryptobranchidae + Hynobiidae) branched earlier than the Sirenidae. Using recently recommended fossil calibration points and a "soft bound" calibration strategy, we recalculated evolutionary timescales for tetrapods with an emphasis on living salamanders, under a Bayesian framework with and without a rate-autocorrelation assumption. Our dating results indicate: (i) the widely used rate-autocorrelation assumption in relaxed clock analyses is problematic and the accuracy of molecular dating for early lissamphibian evolution is questionable; (ii) the initial diversification of living amphibians occurred later than recent estimates would suggest, from the Late Carboniferous to the Early Permian (~294 MYA); (iii) living salamanders originated during the Early Jurassic (~183 MYA), and (iv) most salamander families had diverged from each other by Late Cretaceous. A likelihood-based ancestral area reconstruction analysis favors a distribution throughout Laurasia in the Early Jurassic for the common ancestor of all living salamanders.

1. Introduction

Salamanders (Caudata), one of three major groups of living amphibians, comprise 578 extant species, most commonly grouped into 67 genera and 10 families (AmphibiaWeb, 2009). Because salamanders are often used as model systems to assess fundamental issues of morphological, developmental and biogeographical evolution, robust phylogenetic hypotheses concerning relationships among the families of living salamanders are basic necessities.

There is a lack of consensus regarding family-level phylogenetic relationships for living salamanders (reviewed in Larson et al., 2003). Most studies support the monophyly of internally fertilizing salamanders, i.e., the families Ambystomatidae, Amphiumidae, Dicamptodontidae, Plethodontidae, Proteidae, Rhyacotritonidae, and Salamandridae (Duellman and Trueb, 1986; Larson and Dimmick, 1993; Hay et al., 1995; Wiens et al., 2005; Roelants et al., 2007), although this conclusion was challenged by three independent studies using both morphological and molecular data (Gao and Shubin, 2001; Weisrock et al., 2005; Frost et al., 2006). Earlier studies placed the family Sirenidae as the sister-group to all remaining salamanders (Goin et al., 1978; Duellman and Trueb, 1986; Milner, 1983, 1988, 2000), but recent analyses of nuclear gene sequences consistently favored the hypothesis that Cryptobranchoidea (Cryptobranchidae and Hynobiidae) branched

* Corresponding authors. Fax: +1 510 643 8238 (D.B. Wake).
E-mail addresses: alarzhang@gmail.com (P. Zhang), wakelab@berkeley.edu (D.B. Wake).

1055-7903/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2009.07.010

earlier than Sirenidae (Wiens et al., 2005; Frost et al., 2006; Roelants et al., 2007). Moreover, relationships within the large clade of internally fertilizing salamanders are not fully resolved and remain controversial.

Salamanders are known to have a long evolutionary age of at least 150 million years (Evans et al., 2005) and their initial diversification likely occurred within a relatively short window of time (Weisrock et al., 2005). When using DNA sequences to infer the phylogeny of salamanders, we face a major problem that the branches grouping multiple families are very short relative to the long terminal branches, which makes the phylogenetic relationships among the families of salamanders difficult to resolve (Wiens et al., 2008). To improve phylogenetic resolution, the most effective method is to increase the amount of phylogenetic signal (i.e., increase the quantity of DNA data). Compared with previous studies that used relatively small amounts of DNA data (Larson and Dimmick, 1993; Hedges and Maxson, 1993; Hay et al., 1995), recent efforts employing increasingly larger quantities of DNA data show better performance for tree resolution and higher levels of congruence with morphological studies (e.g., Roelants et al., 2007).

Timing of phylogenetic events during the evolution history of salamanders has been estimated by earlier workers and is a matter of considerable interest to paleontologists and historical biogeographers. Using mitogenome data but incomplete taxon sampling, Zhang et al. (2005) suggested that the origin of living salamanders was no less than 197 million years ago (MYA). Based on data from the nuclear RAG1 gene, San Mauro et al. (2005) and Hugall et al. (2007) estimated the age of stem Caudata at about 270 million years ago. Another recent molecular study (Roelants et al., 2007), using four nuclear and a mitochondrial marker for representatives of all living families, provided a younger estimate of about 220–249 MYA. Marjanović and Laurin (2007) compiled a supertree including 223 extinct species of lissamphibians. Using paleontological data and inferences, they hypothesized that living salamanders arose in Mid-Late Jurassic (~162 MYA), a much younger date than any calculation based on the molecular data. This apparent discordance on divergence time estimates among different molecular studies and between molecular and fossil results is a focus of our analysis.

Mitochondrial DNA (mtDNA) is a useful marker system in phylogenetic analyses because of its maternal mode of inheritance and relative lack of recombination (Saccone et al., 1999). As a single, haploid, nonrecombining linkage unit, the mt genome of vertebrates represents only one-fourth of the effective population size compared with the nuclear (nc) genome, which results in a shorter expected coalescence time for mt loci compared with nc loci and a greater probability that the mt gene tree will accurately reflect the species tree (Moore, 1995). Moreover, mtDNA is a moderate-scale genome suitable for complete sequencing and thus provides substantial amounts of DNA data for phylogenetic analyses. Previous studies demonstrated that mitogenomic data recovered robust phylogenies (with high statistical support) for many taxa (Miya and Nishida, 2000; Miya et al., 2001; Mueller et al., 2004; Zhang et al., 2005, 2006), and thus may resolve questions of higher-level relationships of salamanders.

In order to re-examine the family-level relationships among living salamanders, we sequenced seven complete mitochondrial genomes of salamanders from five families, four previously not represented. By combining these sequences with published salamander mitochondrial genomes, we present a comprehensive molecular phylogenetic analysis for living salamanders. We also use various statistical tests to evaluate alternative phylogenetic hypotheses derived from previous studies as well as the hypotheses generated from our new phylogenetic results. Finally, we present estimates for the time tree of evolution in this clade using new analytical methods.

2. Materials and methods

2.1. Taxon sampling for mitochondrial genomes

Complete mitochondrial genomes for 83 salamanders were deposited in GenBank before this study began, representing 6 of 10 families. Our sampling strategy is to include all extant salamander families but also all key genera for each family, in order to reduce long-branch attraction and to more accurately date phylogenetic events. For the Plethodontidae, Salamandridae, Hynobiidae, Ambystomatidae, and Rhyacotritonidae, existing data deposited in GenBank are relatively abundant. For the Cryptobranchidae, the sole North American species, *Cryptobranchus alleganiensis,* was added to the Asian species of *Andrias* (previously studied). For the remaining families (Amphiumidae, Dicamptodontidae, Proteidae, and Sirenidae), we added species to include a total of six missing genera. Moreover, complete mitochondrial genomes of three frogs, three caecilians, one lungfish and one coelacanth were retrieved from GenBank to serve as outgroup taxa in the phylogenetic analyses. Data for four representative sauropsids (1 bird, 1 lizard, and 2 crocodiles) were retrieved from GenBank to be used in our molecular dating analyses. The details for all sequences used in this study are given in Table 1.

2.2. Laboratory protocols

Total DNA was purified from frozen or ethanol-preserved tissues (liver or muscle) using the Qiagen (Valencia, CA) DNeasy Blood and Tissue Kit. A suite of 22 primers (Table 2) was used to amplify contiguous and overlapping fragments that covered the entire mt genome (Fig. 1). PCRs were performed with AccuTaq LA DNA Polymerase (SIGMA) in total volumes of 25 μl, using the following cycling conditions: an initial denaturing step at 96 °C for 2 min; 35 cycles of denaturing at 94 °C for 15 s, annealing at 45–55 °C (see Table 2) for 60 s, and extending at 72 °C for 5 min; and a final extending step of 72 °C for 10 min. PCR products were purified either directly via ExoSAP (USB) treatment or gel-cutting (1% TAE agarose) using the gel purification kit (Qiagen). Sequencing was performed directly with the corresponding PCR primers using the BigDye Deoxy Terminator cycle-sequencing kit v3.1 (Applied Biosystems) in an automated DNA sequencer (ABI PRISM 3730) following the manufacturer's instructions. For some large PCR fragments, specific primers were designed according to newly obtained sequences to facilitate primer walking.

2.3. Sequence alignments, data partition, and model selection

We included all species listed in Table 1 except for the 4 sauropsid species (33 in total) for phylogenetic reconstruction. For estimates of divergence dates, all species (37 in total) were used. Ribosomal RNAs and tRNAs were aligned manually with reference to secondary structure, according to recommendations of Kjer (1995) and Gutell et al. (1994). Models for rRNA secondary structure came from the Comparative RNA Web (CRW) site. Length variable regions (mainly rRNA and tRNA loops) were excluded. All 22 tRNA alignments were then combined to generate a concatenated alignment. Several tRNA genes are incomplete in some mt genomes. For these, "Ns" were added to the corresponding alignments and treated as missing data. All 13 protein-coding genes were translated to amino acids and aligned using Clustal W (Thompson et al., 1997) implemented in the Megalign program (DNASTAR package) at default settings, and then shifted back to DNA sequences. Thus we obtained alignments for amino acids and nucleotides simultaneously. To avoid bias in refining the protein-coding gene alignments, we used Gblocks (Castresana, 2000) to extract