Molecular Phylogenetics and Evolution 115 (2017) 16–26

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Inferring the shallow phylogeny of true salamanders (*Salamandra*) by multiple phylogenomic approaches

Ariel Rodríguez [a,b], James D. Burgon [c], Mariana Lyra [d], Iker Irisarri [e], Denis Baurain [f], Leon Blaustein [g], Bayram Göçmen [h], Sven Künzel [i], Barbara K. Mable [c], Arne W. Nolte [i,j], Michael Veith [k], Sebastian Steinfartz [a], Kathryn R. Elmer [c], Hervé Philippe [l,m], Miguel Vences [a,*]

[a] *Zoological Institute, Braunschweig University of Technology, Mendelssohnstr. 4, 38106 Braunschweig, Germany*
[b] *Institute of Zoology, Tierärztliche Hochschule Hannover, Bünteweg 17, 30559 Hannover, Germany*
[c] *Institute of Biodiversity, Animal Health & Comparative Medicine, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK*
[d] *Departamento de Zoologia, Instituto de Biociências, UNESP – Univ Estadual Paulista, Campus Rio Claro, Av 24A, N 1515, CEP 13506-900 Rio Claro, SP, Brazil*
[e] *Laboratory for Zoology and Evolutionary Biology, Department of Biology, University of Konstanz, Universitätsstr. 10, 78464 Konstanz, Germany*
[f] *InBioS – PhytoSYSTEMS, Eukaryotic Phylogenomics, University of Liège, Chemin de la Vallée 4, Bât. B22, 4000 Liège, Belgium*
[g] *Institute of Evolution and Department of Evolutionary and Environmental Biology, University of Haifa, Haifa 3498838, Israel*
[h] *Science Faculty, Biology Department, Zoology Section, Ege University, Turkey*
[i] *Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany*
[j] *Carl von Ossietzky University Oldenburg, Institute for Biology, Carl von Ossietzky Str. 9-11, 26111 Oldenburg, Germany*
[k] *Biogeography Department, Trier University, 54286 Trier, Germany*
[l] *Centre for Biodiversity Theory and Modelling, UMR CNRS 5321, Station of Theoretical and Experimental Ecology, 2 route du CNRS, 09200 Moulis, France*
[m] *Département de Biochimie, Centre Robert-Cedergren, Université de Montréal, Montreal, QC H3C 3J7, Canada*

ARTICLE INFO

Article history:
Received 3 February 2017
Revised 31 May 2017
Accepted 13 July 2017
Available online 14 July 2017

Keywords:
Amphibia
Caudata
Salamandridae
Phylogenomics
RNAseq
Transcriptomes
RADseq
Mitochondrial genomes
Concatenation
Species tree

ABSTRACT

The rise of high-throughput sequencing techniques provides the unprecedented opportunity to analyse controversial phylogenetic relationships in great depth, but also introduces a risk of being misinterpreted by high node support values influenced by unevenly distributed missing data or unrealistic model assumptions. Here, we use three largely independent phylogenomic data sets to reconstruct the controversial phylogeny of true salamanders of the genus *Salamandra*, a group of amphibians providing an intriguing model to study the evolution of aposematism and viviparity. For all six species of the genus *Salamandra*, and two outgroup species from its sister genus *Lyciasalamandra*, we used RNA sequencing (RNAseq) and restriction site associated DNA sequencing (RADseq) to obtain data for: (1) 3070 nuclear protein-coding genes from RNAseq; (2) 7440 loci obtained by RADseq; and (3) full mitochondrial genomes. The RNAseq and RADseq data sets retrieved fully congruent topologies when each of them was analyzed in a concatenation approach, with high support for: (1) *S. infraimmaculata* being sister group to all other *Salamandra* species; (2) *S. algira* being sister to *S. salamandra*; (3) these two species being the sister group to a clade containing *S. atra*, *S. corsica* and *S. lanzai*; and (4) the alpine species *S. atra* and *S. lanzai* being sister taxa. The phylogeny inferred from the mitochondrial genome sequences differed from these results, most notably by strongly supporting a clade containing *S. atra* and *S. corsica* as sister taxa. A different placement of *S. corsica* was also retrieved when analysing the RNAseq and RADseq data under species tree approaches. Closer examination of gene trees derived from RNAseq revealed that only a low number of them supported each of the alternative placements of *S. atra*. Furthermore, gene jackknife support for the *S. atra* - *S. lanzai* node stabilized only with very large concatenated data sets. The phylogeny of true salamanders thus provides a compelling example of how classical node support metrics such as bootstrap and Bayesian posterior probability can provide high confidence values in a phylogenomic topology even if the phylogenetic signal for some nodes is spurious, highlighting the importance of complementary approaches such as gene jackknifing. Yet, the general congruence among the topologies recovered from the RNAseq and RADseq data sets increases our confidence in the results,

* Corresponding author.
E-mail address: m.vences@tu-bs.de (M. Vences).

http://dx.doi.org/10.1016/j.ympev.2017.07.009
1055-7903/

and validates the use of phylotranscriptomic approaches for reconstructing shallow relationships among closely related taxa. We hypothesize that the evolution of *Salamandra* has been characterized by episodes of introgressive hybridization, which would explain the difficulties of fully reconstructing their evolutionary relationships.

1. Introduction

The rise of high-throughput sequencing techniques has provided molecular systematists with unprecedented opportunity to analyse controversial phylogenetic relationships in great depth (da Fonseca et al., 2016). In most organisms the sequencing of entire genomes is technologically within reach, but complexity-reduction approaches, such as restriction site associated DNA sequencing (RADseq), anchored hybrid enrichment, or sequencing of transcriptomes (i.e., the transcribed RNA; RNAseq), are more affordable and are increasingly being used to obtain markers that are representative of the nucleotide diversity across the genome (e.g., Emerson et al., 2010; Lemmon et al., 2012; Prum et al., 2015; Wen et al., 2015). Typically, phylogenomic approaches based on single nucleotide polymorphisms (SNPs) have been applied to inferences of population-level differentiation, phylogeography, and phylogenetic relationships among closely related species (Davey and Blaxter, 2011; Rubin et al., 2012; Peterson et al., 2012; Darwell et al., 2016), whereas those based on sequences of protein-coding genes derived from RNAseq or full genomes have been used for inferring deep nodes in the tree of life, often analyzed at the amino acid level (Bapteste et al., 2002; Chiari et al., 2012; Wickett et al., 2014; Jarvis et al., 2014; Chen et al., 2015; Irisarri and Meyer, 2016). More recently such phylotranscriptomic analysis of RNAseq-derived markers has proven to also provide valuable insights into shallow relationships between species (Wang et al., 2017; Brandley et al., 2015).

The ever-increasing amount of data available for phylogenomic analyses provides unprecedented opportunities to resolve the tree of life (Philippe et al., 2005), but also introduces novel risks of drawing misleading conclusions. In particular, current approaches of assessing node stability, such as non-parametric bootstrapping and Bayesian posterior probabilities, tend to provide very high node support with large amounts of data, but these might reflect artefacts such as unevenly distributed missing data rather than real phylogenetic signal (e.g., Dell'Ampio et al., 2014). Also, if model assumptions are not realistic, wrong phylogenetic relationships can be supported by high values thus representing systematic error. Accordingly, just adding more sequence information is not necessarily a route to increasing the quality of phylogenomic inference (Philippe et al., 2011). The scale of the problem might also differ by the type of data used. For example, SNP-based analyses cannot incorporate codon-based nucleotide substitution models into the phylogenetic inference process. Similarly, concatenating across genes in an RNAseq analysis precludes gene-specific estimation of transition matrices or accurate estimation of rate heterogeneity, because current software to estimate partition schemes (e.g., Lanfear et al., 2012) is still in its infancy when it comes to efficiently handling thousands of genes. It is also insufficiently known how sensitive phylogenomic resolution is to combining different types of data in the same analyses. While these issues are not new (e.g. "total evidence" debate from the 1990s; e.g. Bull et al., 1993), the "big data" typical of modern approaches could require new strategies for assessing confidence. For example, resampling genes or loci rather than individual nucleotide sites (e.g. gene jackknifing) in a concatenated analysis could be more informative than traditional bootstrap or posterior probability analyses (Irisarri et al., 2017). However, direct empirical comparisons between gene-based and SNP-based data obtained from the same set of samples, necessary to determine their relative sensitivity to phylogenetic error and to assess their performance in resolving shallow phylogenetic relationships, are scarce.

Here we empirically compare the use of RNAseq, RADseq, and whole mitochondrial genomes to test evolutionary hypotheses about relationships in a prominent group of amphibians, the 'true salamanders' of the genus *Salamandra* (family Salamandridae). This genus includes six recognised species (Speybroeck et al., 2010; Sillero et al., 2014) that vary in both color patterns and reproductive modes. Some species have a conspicuous yellow-black coloration, thought to be of aposematic function (*S. algira* from North Africa; *S. corsica* from Corsica; *S. infraimmaculata* from the Near East; and *S. salamandra* from Europe), whereas others have a uniformly black coloration (*S. atra* and *S. lanzai*; both distributed in the European Alps). They are also one of the few groups of vertebrates to vary in viviparity across their range, including instances of deposition of aquatic larvae or terrestrial juveniles reared on yolk nutrition, vs. release of fully metamorphosed young reared on maternal nutrition (Wake, 1993; Greven and Guex, 1994; Greven, 2003; Buckley et al., 2007; Caspers et al., 2014). Clarifying the phylogenetic relationships among these species is thus of interest for studies of biogeography, coloration and toxicity (function of aposematism), and the evolution of different reproductive modes.

Previous molecular phylogenetic studies based on DNA sequences of mitochondrial and nuclear genes, and complete mitochondrial genomes, placed the Asian *Lyciasalamandra*, another clade of viviparous salamanders, as sister taxon to *Salamandra*, and a clade comprising *Chioglossa* and *Mertensiella* sister to the *Salamandra*/*Lyciasalamandra* clade (Titus and Larson, 1995; Veith et al., 1998; Weisrock et al., 2001, 2006; Veith and Steinfartz, 2004; Frost et al., 2006; Steinfartz et al., 2007; Zhang et al., 2008; Pyron, 2014). However, despite combining DNA sequences of 10 mitochondrial and 13 nuclear genes (Vences et al., 2014), the relationships among species of *Salamandra* have remained poorly resolved. Several relationships were supported by both types of markers, such as a clade containing the black-colored alpine species (*S. atra*, *S. lanzai*) plus *S. corsica*, but most other relationships did not receive strong support.

In the present study, we newly sequenced and assembled three phylogenomic data sets to resolve the phylogenetic relationships among species of *Salamandra*: (1) 3070 protein-coding nuclear genes obtained from transcriptomes (RNAseq); (2) 7440 anonymous nuclear markers obtained via double-digest Restriction Site Associated DNA sequencing (ddRADseq); and (3) full mitochondrial genomes. We analyse these data using concatenation and "species tree" approaches, to assess the phylogeny of true salamanders. Further, we scrutinize the congruence of the different molecular data sets and analytical approaches for phylogenetic resolution.

2. Materials and methods

2.1. RNAseq analyses

Transcriptomic data from one individual each of *Salamandra salamandra* from Germany (Kottenforst near Bonn) and of *S.*

infraimmaculata from Israel were available from the study of Czypionka et al. (2015). A further transcriptomic data set was available for *S. salamandra* from France (Banyuls; geographical coordinates 42.479183, 3.101555) from the study of Figuet et al. (2014). New transcriptomic data were generated for single individuals of the other four species of *Salamandra* (*S. algira*, *S. atra*, *S. corsica*, *S. lanzai*), along with two species of *Lyciasalamandra* as an outgroup. We used pooled samples of different organs (skin, muscle and liver) preserved in RNAlater and frozen at −80 °C. RNA extraction from 100mg of tissue of each salamander was carried out using a trizol protocol (see Supplementary Material). RNA was prepared for sequencing following the Illumina TruSeq mRNA protocol. Sequencing was carried out on the Illumina MiSeq (2 × 250 bp paired-end) platform. Illumina reads were quality-trimmed and filtered using Trimmomatic v. 0.32 (Bolger et al., 2014) with default settings. Filtered reads, paired and unpaired, were used for *de novo* transcriptome assembly using Trinity v. 2.1.0 (Grabherr et al., 2011) following published protocols (Haas et al., 2013). Candidate coding regions within transcript sequences from the final assembly were identified and translated using Transdecoder 2.1.0 (Haas et al., 2013). Raw reads were submitted to the NCBI Short Read Archive database (Bioproject PRJNA385088).

As a basis for selecting nuclear protein-coding genes for analysis, we used a previously compiled alignment from Irisarri et al. (2017), in the following called reference alignment. For detailed methods of obtaining this reference alignment, see Irisarri et al. (2017). In brief, the reference alignment was assembled by first grouping 20 vertebrate proteomes into putative orthology using USEARCH (Edgar, 2010) and OrthoMCL (Li et al., 2003) and discarding orthogroups with missing data for major clades of jawed vertebrates. After aligning and custom paralog-splitting, the resulting protein clusters were complemented with 80 additional published genomic and transcriptomic data sets of vertebrates using the software Forty-Two (or "42"; D. Baurain; https://bitbucket.org/dbaurain/42/) that controls for orthology using several proteomes in strict three-way reciprocal best BLAST hit tests. Subsequently, following the methods outlined in Irisarri et al. (2017), the reference alignment went through several decontamination steps to remove: (1) all human and non-vertebrate sequences; (2) cross-contaminations; (3) highly incomplete genes; (4) genes with poor alignment or frame shifts; (5) genes resulting in extremely long branches in some taxa suggesting the possibility of contamination or undetected paralogy. The final vertebrate reference data set as used for the analysis of Irisarri et al. (2017) contained 4593 genes.

Sequences of *Salamandra* transcriptomes were aligned to this reference data set using the software Forty-Two. The resulting alignment was then submitted to a pipeline of thorough filtering and decontamination, composed of the following eight steps: (1) Sequences from non-vertebrate sources (e.g. Bacteria or Platyhelminthes) were detected by BLAST searches and 21,265 sequences were removed (almost exclusively from the previously published *S. infraimmaculata* and *S. salamandra* sequences). (2) To remove redundant and/or divergent sequences, for each sample represented by at least two sequences for a given gene (e.g., multiple short transcripts that could not be assembled together), every sequence was compared against all the other sequences in the alignment by BLAST. A total of 16,295 sequences were eliminated if their average bit score was at least 10% lower than the best average bit score of the redundant set and if there was a length overlap of ≥95% between the two sequences. (3) We then excluded genes providing unrealistic phylogenetic resolution in our target group (salamanders) by excluding such genes for which the genus *Salamandra* or the family Salamandridae were not recovered as monophyletic; for this analysis, phylogenetic trees were inferred using RAxML v8 (Stamatakis, 2014) and an LG + Γ model (Le and Gascuel, 2008) using only amphibian sequences. (4) To remove from the remaining genes those that might be affected by undetected ancient paralogy, we split the respective alignments by looking for the branch that maximizes taxonomic diversity (see Amemiya et al., 2013). Phylogenetic trees were inferred again on these split alignments and we again retained only genes for which the genus *Salamandra* or the family Salamandridae were monophyletic. In total, 3105 genes recovered a monophyletic genus *Salamandra* and among the remaining genes retained (i.e. with non-monophyletic *Salamandra*), 508 recovered Salamandridae as monophyletic. (5) We then retrieved nucleotide sequences for these genes from the original transcriptomic data; genes for which sequences were available for fewer than five *Salamandra* species or without *Lycisalamandra* data were discarded. The corresponding retained nucleotide sequences were both recovered and aligned according to the amino acid alignments using the software Leel (or "1331"; D. Baurain; https://bitbucket.org/dbaurain/42/). All subsequent analyses were based on these nucleotide alignments. (6) We checked for remaining contaminating or paralogous sequences by comparing the branch lengths in gene trees and in the concatenation tree as in Irisarri et al. (2017), and removed 445 sequences having a branch length ratio (gene tree vs. concatenation tree) >7. (7) To reduce stochastic error in gene trees, we removed all the codons that were present in less than 50% of the species and any sequence having less than 30 nucleotides. (8) The resulting nucleotide data set (available as Mendeley Research Data) was then concatenated using SCaFoS (Roure et al., 2007) and the 3070 genes with less than 3 missing species were retained and used for phylogenetic analysis.

Maximum likelihood (ML) phylogenetic inference was conducted on the concatenated nucleotide matrix using RAxML, partitioning the data by genes. For each gene, we estimated a separate general time reversible (GTR) model of nucleotide evolution (i.e. 6 substitution rates), with rate heterogeneity modelled according to a gamma distribution (shape parameter alpha) with four rate categories. We assessed node support with 1000 non-parametric bootstrap replicates. As conflicting genealogical histories often exist in different genes throughout the genome, concatenation methods can result in incorrect trees with high support (Kubatko and Degnan, 2007; Degnan and Rosenberg, 2009). We thus took several strategies for assessing potential conflicts.

First, node support was assessed using a gene jackknife approach (Delsuc et al., 2008) to determine what proportion of the data would need to be sampled to resolve the maximal number of nodes: one hundred alignment replicates were generated by randomly sampling genes up to *ca.* 10,000, 50,000, 100,000, 500,000, 1,000,000 and 3,000,000 nucleotide positions, respectively. For each replicate, unpartitioned ML trees were estimated using RAxML with a GTR + Γ model defined for the whole data set, and gene jackknife proportions estimated for each node.

Second, we also used ASTRAL II (Mirarab and Warnow, 2015), a statistically consistent algorithm to estimate the species tree topology under the multi-species coalescent model (Mirarab and Warnow, 2015). Clade support was evaluated by computing the local posterior probability, a feature of ASTRAL II that has shown high precision compared with multi-locus bootstrapping on a wide set of simulated and biological datasets (Sayyari and Mirarab, 2016). As species tree analyses do not require outgroups (Heled and Drummond, 2009) the ASTRAL II analyses was carried out with ingroup sequences (genus *Salamandra*) only, but an additional exploratory analysis including the outgroup was also performed.

2.2. RADseq analyses

Tissue samples were collected from two individuals of *Salamandra algira*, *S. atra*, *S. corsica*, *S. infraimmaculata* and *S. salamandra*, and one individual of *S. lanzai*. Tissue was also collected from one

individual each of *Lyciasalamandra billae* and *L. flavimembris* to provide an outgroup. Genomic DNA was extracted using the Macherey-Nagel NucleoSpin® Tissue kit following the manufacturer's instructions. We applied double-digest Restriction Site Associated DNA sequencing (ddRADseq; Peterson et al., 2012) and for simplification hereafter refer to the resulting sequences as RADseq data set. The library was prepared as follows (per Recknagel et al., 2015 with modification of Illumina adapters): 1 μg of DNA from each individual was double-digested using the *PstI-HF*® and *AclI* restriction enzymes (New England Biolabs); modified Illumina adaptors with unique barcodes for each individual were ligated onto this fragmented DNA; samples were multiplexed (pooled); and a PippinPrep used to size select fragments around a tight range of 383 bp, based on the fragment length distribution identified using a 2200 TapeStation instrument (Agilent Technologies). Finally, enrichment PCR was performed to amplify the library using forward and reverse RAD primers. Sequencing was conducted on an Illumina Next-Seq machine at Glasgow Polyomics to generate paired-end reads 75 bp in length. Raw reads were submitted to the NCBI Short Read Archive database (Bioproject PRJNA386146).

Sequence reads were de-multiplexed, Illumina adaptors and barcodes removed, and reads truncated to 60 nucleotides using Stacks v. 1.35 (Catchen et al., 2013). Processed reads per sample ranged from 4.9 to 16.8 million, compared to 5–17 million raw reads per sample. Reads were assembled *de novo* into loci using pyRAD v. 3.0.6 (Eaton, 2014). Reads were first clustered within an individual at a minimum depth of 10 with a clustering threshold of 85%. The same clustering threshold was then used to assemble *de novo* loci across samples; final RAD loci ranged in length from 109 to 144 nucleotides, with an average of 111 nucleotides. As the performance of *de novo* assembled RADseq data matrices in phylogenetic reconstructions depends on the sample coverage and potential intra-locus paralogy (Huang and Knowles, 2014; Takahashi et al., 2014), we explored a range of thresholds for loci coverage between samples (4, 6, 8 10 and 11 individuals; equivalent to 31–100% of the in-group) and maximum number of SNPs per RAD locus (2, 4, 6, 8 and 10). RAxML, with a GTRGAMMA model and 100 rapid bootstraps, was used to explore the resulting concatenated sequence matrices in order to choose the filters that produced the most reliable trees (based on node resolution and support). All trees agreed in almost all aspects (except the relationships among *S. atra*, *S. lanzai*, and *S. corsica* that were left unresolved in some analyses), and we eventually chose a between-sample coverage of 6 with a maximum number of 2 SNPs per locus.

Phylogenetic analyses of the RADseq data set were conducted in BEAST 2.4.2 (Bouckaert et al., 2014). For the concatenated analysis of loci, a BEAST xml file was generated using BEAUTi 2.4.2. The best fitting evolutionary model inferred by jModeltest 2.1.10 was the transversion model (TVM; based on the Bayesian information criterion). As BEAST2 only has four base substitution models, which do not include TVM, a GTR substitution model with the alpha gamma rate parameter fixed at one was selected to simulate it. A relaxed clock (log normal) was used, with all other parameters left on default settings. A MCMC chain of 10 million generations was run (10% burn-in) with tree and parameter estimates sampled every 1000 MCMC generations. Tracer 1.6.0 (Rambaut and Drummond, 2007) was used to assess chain convergence. Eight prior operators (treeScaler; SubtreeSlide; RateAGScaler; RateATScaler; YuleModelTreeScaler; YuleModelSubtreeSlide; FixMeanMutationRatesOperator; and ucldStdevScaler) were optimised based on the output of this trial, and the analysis re-run. A maximum clade credibility tree was then generated from the output of the optimised analysis using TreeAnnotator 2.4.2. "Gene" jackknifing (i.e., jackknifing of RAD loci) was carried out as described for the RNAseq data, with replicates of *ca.* 10,000, 50,000, 100,000, 500,000 and 800,000 nucleotide positions.

In addition to the concatenation approach, we also used the coalescent-based program SNAPP (Bryant et al., 2012) to infer the species tree under a finite-sites model of mutation from unlinked biallelic SNPs extracted from the RADseq data set. The *Lyciasalamandra* outgroup was removed as no outgroup is required in species tree analyses (Heled and Drummond, 2009) and loci were re-filtered in pyRAD to extract a single SNP per locus, giving a final data set consisting of 3586 loci from across the 11 *Salamandra* samples. Using the SNAPP template in BEAUTi, 2.4.2 a BEAST xml file was generated. Given a lack of reliable prior information, mutation rates were sampled and a uniform distribution was used for the lambda parameter of the Yule prior; all other priors were left at default. BEAST was run with 10 million generations, 10% burn-in, and tree and parameter estimates sampled every 1000 MCMC generations. Convergence was assessed with Tracer 1.6.0 and the maximum clade credibility tree generated using TreeAnnotator 2.4.2.

2.3. Mitogenome analyses

We assembled mitogenomes of all *Salamandra* species from the quality-trimmed RNAseq data. We first randomly sampled 20% of the raw data and subsequently retrieved and assembled mitochondrial sequences with MIRA v4.0 (Chevreux et al., 1999) and MITObim v1.8 (Hahn et al., 2013) following Machado et al. (2016) and using default parameters. We used the complete mitochondrial sequence of *Salamandra infraimmaculata* (EU880331) as reference genome in the first MIRA step. Assemblies in CAF format were manually verified in Geneious software, v. 6 (Biomatters) to evaluate the coverage and quality of each mtDNA element. All positions with coverage lower than 4 were coded as ambiguous ("N"). Preliminary annotation of each sequence was done using the mitochondrial genome annotation server MITOS (Bernt et al., 2013) with default parameters. Validation of tRNA sequences were performed using tRNAscan-SE (Lowe and Chan, 2016). The resulting automatic annotation was confirmed and edited manually by comparison to *Salamandra infraimmaculata* EU880331. All newly determined sequences were submitted to Genbank (accession numbers MF043386–MF043393). We also included complete or almost complete mitochondrial genome sequences of *Chioglossa lusitanica* (EU880308) and *Mertensiella caucasica* (EU880319) as outgroups, added species of *Lyciasalamandra* (EU880318, AF154053) as hierarchical outgroups, and furthermore added one species of *Salamandra* for which a full mitogenome sequence was available from Genbank (EU880331). The latter sample was originally analyzed as *S. salamandra* (Zhang et al., 2008), but corresponds to a sample of *S. infraimmaculata* from Turkey.

We aligned mitochondrial sequences using MAFFT v.7 (Katoh and Standley, 2013) and determined the optimal among-gene partitioning scheme and model choice for dataset in PartitionFinder (Lanfear et al., 2012) under the Bayesian Information Criterion (BIC). Bayesian phylogenetic inference was performed with MrBayes v.3.2.6 (Ronquist et al., 2012) using two independent runs of eight chains. Chains were started from random trees and run for 10 million generations each, being sampled every 1000 generations. Twenty-five percent of the trees were discarded as 'burn-in' before generating a consensus tree. The full mitogenomic data set was also analyzed under the ML optimality criterion in RAxML v. 8. (Stamatakis, 2014), using the GTRGAMMA model of nucleotide substitution and a partitioned approach, with partitions and substitution models as defined by Partitionfinder (Lanfear et al., 2012). Node support was assessed using 1000 bootstrap replicates.

2.4. Independence of data sets

To understand whether the RNAseq and RADseq datasets were independent, we calculated sequence overlap between them. For each individual, sequences representing the RAD loci, were aligned to the RNAseq data set using Bowtie2 v.2.2.9. An overall alignment rate of 1.23% was found: of 56,987 paired reads that mapped, none aligned concordantly, 46 (0.08%) aligned discordantly once, and when single-end reads were aligned independently (113,882 in total), 817 (0.72%) aligned one time and 497 (0.44%) aligned >1 time. This confirms that the loci used for the RNAseq and RADseq analyses were almost completely non-overlapping and that the two analyses can be considered independent subsamples of the same underlying genomes. We also confirmed that no genes encoding mitochondrial proteins were present in the final RNAseq alignment used for analysis.

2.5. Gene ontology analyses

Given that mitochondrial and nuclear DNA sequences gave a contradicting signal regarding the monophyly of black or alpine salamanders (i.e. a grouping of *S. atra* + *S. corsica* vs. *S. atra* + *S. lanzai*; see Results) we tested whether this discordance could be explained by differences in the functional categories of genes. We first counted the total number of genes supporting a certain topology using Phylosort, v. 1.3 (Moustafa and Bhattacharya, 2008). We then tested whether incongruencies among analyses were due to particular nuclear genes coding for proteins whose functions interact with mitochondria and might have thus coevolved with mitochondrial genes (Hill, 2016). We first created a consensus protein sequence for each of the 3070 RNAseq genes and used BLAST to compare them to the UniProtKB/Swiss-Prot database (http://www.uniprot.org/). We selected the most similar sequence to represent the gene ontology term for the given protein. From the total list of genes, we selected those for which the phylogenetic tree supported either the *S. atra* + *S. corsica* or *S. atra* + *S. lanzai* sister group relationship. We then used the UniProt Retrieve/ID mapping web server (http://www.uniprot.org/uploadlists/) to classify genes into Gene Ontology domains (Table S1).

3. Results

3.1. RNAseq analyses

The final concatenated alignment derived from RNAseq contained 3,255,534 bp including 74,801 variable (2.30%) and 28,125 parsimony-informative (0.87%) positions. This corresponded to the nucleotide sequences of 3070 genes with mean alignment lengths of 1060 ± 566 bp SD (minimum–maximum = 228–7068 bp). Taxonomic coverage varied from 6 to 9 per gene, and the percentage of missing data per taxon ranged from 0 to 35.8%. The ML tree calculated from this concatenated matrix (Fig. 1A) provided a fully resolved tree of *Salamandra* species, with *S. infraimmaculata* sister to all remaining species, and a clade of *S. algira* and *S. salamandra* sister to a clade of *S. atra*, *S. corsica* and *S. lanzai*, with the two alpine species *S. atra* and *S. lanzai* forming a monophyletic group sister to *S. corsica*.

Gene jackknife proportions of the RNAseq data set revealed that up to 3 million nucleotide positions are necessary to recover all final-tree bipartitions with high support (>75%; Fig. 2A). With 10,000 nucleotide positions, only one out of five nodes of interspecific relationships were recovered with high support (corresponding to the placement of the two *Lyciasalamandra* species as sister group); 50,000 nucleotide positions were sufficient to recover the close affinity of *S. algira* and *S. salamandra*; whereas the remaining three nodes required replicates of more than one million nucleotide positions to be resolved with high support.

The tree obtained from the ASTRAL II species tree analysis (Fig. 3A) partly agreed with the trees obtained by the analyses of the concatenated data set (Fig. 1) but placed *S. corsica* apart from the *atra-lanzai* clade. Repeating the same analysis with outgroup sequences (*Lyciasalamandra*) resulted in yet another topology, where *S. corsica* was placed sister to *S. atra* (Fig. S2).

3.2. RADseq analyses

From the RADseq data, we assembled an alignment of 7440 loci present in at least 6 samples (33.4% missing data), with a maximum number of 2 SNPs per locus, and containing 822,917 nucleotide positions, 17,985 SNPs, and 7189 parsimony-informative sites (0.87%). Bayesian inference analysis of this concatenated RADseq data set yielded a tree identical to that obtained with the concatenated RNAseq data (Fig. 1B), with all nodes showing posterior probabilities of 1. Exploratory analysis of a more stringently selected RADseq data set, with 1541 SNPs from 586 loci present in at least 11 samples and with at most 2 SNPs per locus (missing data 8.0%), recovered an identical topology, but with lower support values for several nodes.

A very similar jackknife pattern was observed for the RNAseq data (Fig. 2A) as compared to the RADseq data (Fig. 2B). Up to 500,000 nucleotide positions were needed to recover all nodes representing interspecific relationships with support values of 75% or higher.

The SNAPP analysis of the RADseq data led to a tree in agreement with the ASTRAL II tree of the RNAseq data, failing to group *S. corsica* with the clade containing *S. atra* and *S. lanzai* (Fig. 3B).

3.3. Mitogenome analyses

We recovered almost complete mitochondrial genomes for all samples analyzed (coding genes, rRNAs and tRNAs) from the RNA-seq raw sequence reads. The phylogenetic inference of the mitogenomic alignments (Fig. 4) was largely congruent with the nuclear data, with two exceptions. The mitogenome tree showed strong bootstrap support for different relationships among the three species in the clade including the two alpine salamanders (*S. atra* and *S. lanzai*) plus *S. corsica*; *S. atra* and *S. corsica* were identified as being sister groups, as compared to *S. atra* and *S. lanzai* in the RNA-seq and RADseq analyses. The placement of *S. infraimmaculata* also differed: it was inferred to be sister to *S. algira* and S. *salamandra* in the mitogenome tree (although this node was poorly resolved) but ancestral to all of the other species in the nuclear data analyses. The relationships among these three species varied according to partitioning scheme used for the mitogenomic sequences (Fig. S1).

3.4. Gene Ontology analyses

To investigate the origin of the discordance observed between the concatenated nuclear gene analyses compared to the mitogenome analyses, we specifically analyzed which sets of genes supported the two alternative topologies within the clade containing *S. atra*, *S. corsica* and *S. lanzai*. Altogether, out of a total of 3070 genes, only 680 genes supported the *atra-lanzai* clade and a similar number, 665 genes, supported the alternative *atra-corsica* clade. The clade of *algira-salamandra-atra-lanzai* to the exclusion of *corsica* (as recovered by the species tree analyses; Fig. 3) was supported by 279 genes.

A comparison of the ontology of genes supporting either the *atra-corsica* or the *atra-lanzai* clade revealed no clear pattern; for different functional properties, a similar proportion of gene trees supporting either topology were found (Table S1). Given the con-

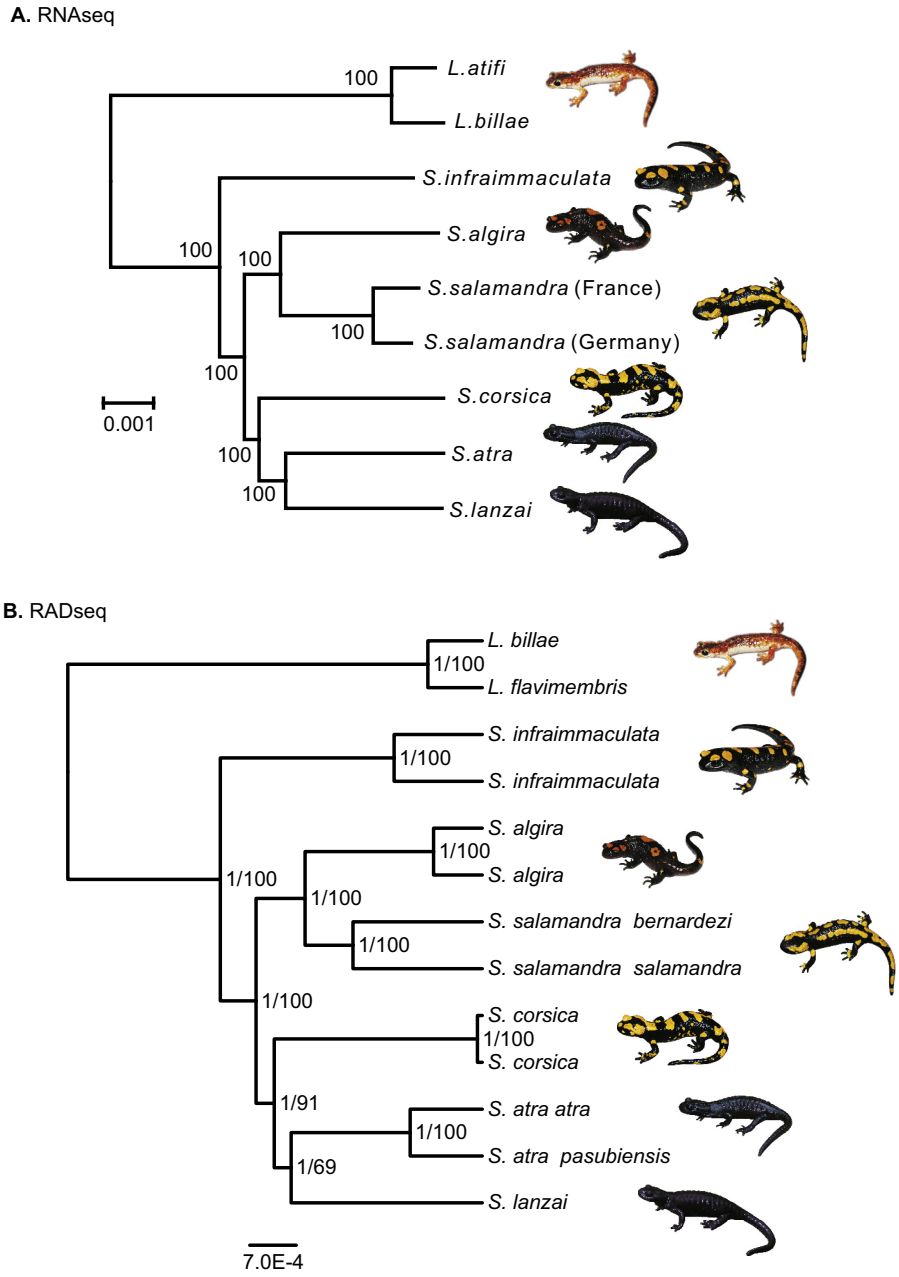

Fig. 1. (A) Phylogenetic tree resulting from the analysis of 3070 orthologous loci (3,256,500 bp) obtained from transcriptomes (RNAseq) of *Salamandra* species partitioned by genes and analyzed under a GTR + Γ model in RAxML; branch support was estimated with 1000 rapid bootstraps. (B) Phylogenetic tree based on a BEAST2 analysis of 7440 concatenated RADseq loci (17,985 SNPs) with a minimum number of 6 samples per locus and a maximum number of 2 SNPs per locus; branch support is based on Bayesian posterior probabilities (first number at nodes) and ML bootstrap analyses (second number at nodes; RaxML rapid bootstrapping, 100 replicates).

flicting phylogenetic resolution among concatenation and species tree approaches of the nuclear data, we tested whether the nuclear phylogenetic signal might have been influenced by genes functionally coupled to mitochondrial genes, e.g. in the respiratory chain. However, these genes again supported the *atra-corsica* vs. the *atra-lanzai* clade in similar proportions (4 vs. 8 genes tightly connected to mitochondrial functions, and 33 vs. 39 genes weakly connected to mitochondrial functions; Table S2).

4. Discussion

4.1. Phylotranscriptomic analysis of shallow phylogenetic relationships

In this study, we used nucleotide sequences of nuclear protein-coding genes derived from various phylogenomic data sets to reconstruct shallow phylogenetic relationships among closely related species of amphibians. Among these were sequences of nuclear protein-coding genes obtained by RNAseq, a kind of data set typically used to resolve deep phylogenies, with amino acid sequences as phylogenetic characters and taxa often separated for hundreds of millions of years (e.g., Misof et al., 2014; Jarvis et al., 2014; Chen et al., 2015). Here we followed an approach that recovered the nucleotide sequences of those transcripts that passed a stringent decontamination pipeline at the amino acid level (to remove sequences and genes potentially affected by sample contamination, paralogy, sequencing errors, or other artefacts; see Methods), and used the resulting concatenated nucleotide alignment of expressed genes for phylogenetic reconstruction. The results fully agreed with those derived from a concatenated alignment of RADseq-derived SNPs (Fig. 1), which represented a

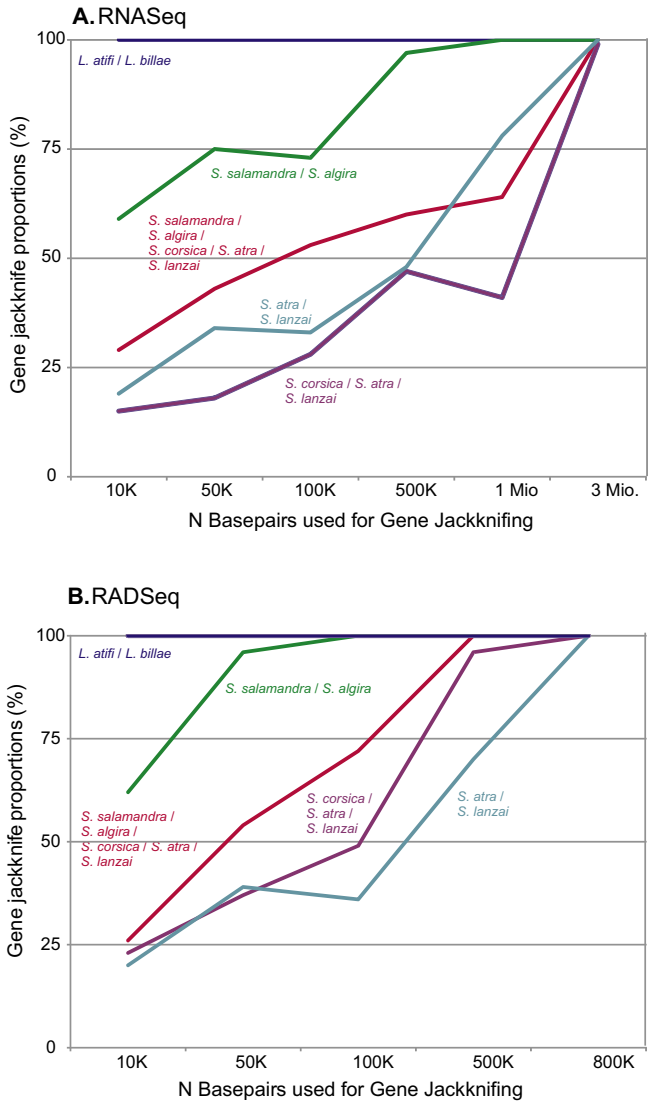

Fig. 2. Gene jackknifing results for (A) the RNAseq and (B) the RADseq data. All results including those of the largest data sets are based on gene resampling without replacement. For RADseq, the procedure treated each locus as "gene" for jackknifing purposes.

largely independent subsample of the salamander's genomes and is a kind of data set typically used for shallow phylogenetic inferences. This validates our usage of a phylotranscriptomic approach, and confirms that such RNAseq-derived data hold promise for reconstructing not only deep nodes of the tree of life but also shallow phylogenetic relationships among taxa probably characterized by recent gene flow.

It is remarkable that with double-digest RAD sequencing it was possible to obtain a substantial number of homologous RADseq loci, despite the enormous genome size of the target species, with haploid C-values estimated between 27 and 41 pg (Gregory, 2016). The overall number of nucleotide positions and phylogenetically informative sites in the RNAseq data set was almost fourfold that of the RADseq data set, but the RADseq data sets still led to a highly resolved tree and most nodes stabilized at shorter gene jackknife replicates as compared to the RNAseq dataset (Fig. 2). RADseq loci represent a relatively random subsample of the entire genome, potentially capturing a wider range of evolutionary signals than protein coding genes; i.e. it should contain fast as well as slowly evolving loci that might broaden the phylogenetic spectrum covered. By contrast, the RNAseq transcript sequences used in our analysis were restricted to those loci that are consistently expressed and conserved across vertebrates, and hence potentially more limited in phylogenetic resolution. Another potential shortcoming is that alleles at heterozygous positions are not called in the RNAseq analysis pipeline and a considerable amount of nucleotide variation is therefore neglected, with possible influences on the RNAseq-derived species tree. Interestingly, despite these very different characteristics of the two data sets, the proportion of phylogenetically informative sites was identical in both of them (0.87%) and they resolved the same relationships among the ingroup taxa. Together these results emphasise that both types of data might be useful for phylogenomics of closely related species and congruence analyses comparing the two can increase confidence in relationships resolved.

4.2. Conflict between concatenation and species tree approaches

Massive phylogenomic data sets, such as those obtained from RNAseq, certainly have the potential to lead to improved phylogenetic inference. However, simply adding more sequences to the data set is not enough (Philippe et al., 2011). Our analysis clearly exemplifies the limitations of large amounts of sequence data, as different analysis methods can result in opposing phylogenetic hypotheses, each with strong support using classical statistical metrics such as non-parametric bootstrap or Bayesian posterior probabilities. For the RADseq data, the species tree analysis with SNAPP (Fig. 3) placed *S. corsica* away from the *atra-lanzai* clade, with maximum posterior probability, conflicting with the analysis of the concatenated alignment that placed these three taxa in one clade (Fig. 1). For the ASTRAL II analysis of RNAseq data, the same species tree topology was found, albeit with partly weaker support.

These differences between the species tree analyses vs. concatenated analyses are surprising, given that in studies on other organisms, congruent results were obtained from the two approaches (e.g., Herrera and Shank, 2016; Tucker et al., 2016). The incongruence in our study could be caused by shortcomings of one of the approaches in (1) dealing with a clade of closely related species, probably affected by incomplete lineage sorting (ILS) and introgression, or (2) dealing with the kind of data, i.e. long protein-coding sequences derived from RNAseq vs. SNPs derived from RADseq.

It has been hypothesized that in the presence of introgression or incomplete lineage sorting, as it can be expected in the case of the closely related *Salamandra* species, multispecies coalescent species tree analyses should provide a more realistic phylogenetic resolution than concatenation (Liu et al., 2009; Leaché and Rannala, 2011; Mirarab et al., 2016). On the other hand, strong advocates for concatenation approaches in phylogenomics are often concerned with resolving deep nodes in the tree of life where ILS should be less of an issue (Gatesy and Springer, 2014). Still, concatenated data sets also appear to perform well for shallower nodes (Wang et al., 2017) and have correctly recovered relationships in studies with simulated sequence data sets (Rubin et al., 2012; Cariou et al., 2013; Tonini et al., 2015; Rivers et al., 2016). Based on these previous studies, we assume that in principle, both coalescence and concatenation approaches should be effective in reconstructing *Salamandra* relationships, given sets of DNA sequences appropriate for the respective method.

However, it is questionable whether the *Salamandra* data sets are equally appropriate for being analyzed with the two methods. It has been contended that coalescence methods should not be applied to complete protein-coding loci because they amalgamate potentially recombining genomic regions with different evolutionary histories, therefore violating important assumptions of the multispecies coalescent model (Springer and Gatesy, 2016). In

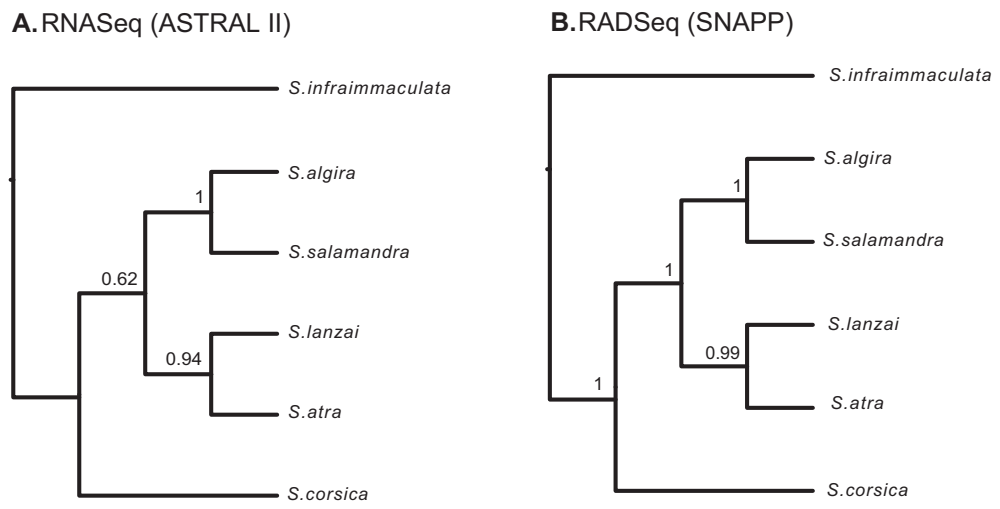

Fig. 3. Results of species tree analyses of *Salamandra*. (A) Species tree obtained from ML gene trees of each of the 3070 orthologous loci from the RNAseq analysis, summarized with ASTRAL II. Branch support was estimated by computing the local posterior probability (not calculated by ASTRAL II for the basalmost node which however is strongly supported in an analysis including *Lyciasalamandra* as outgroup; see Fig. S2). (B) Maximum clade credibility tree (cladogram representation) obtained from a SNAPP analysis of 3586 unlinked SNP loci identified from RADseq data.

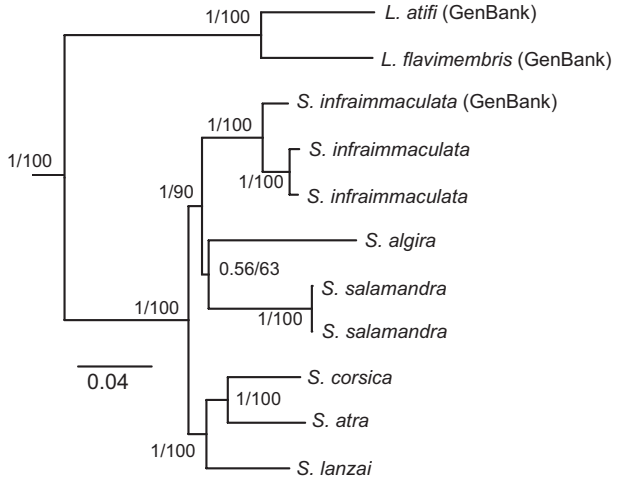

Fig. 4. Majority-Rule consensus tree obtained by partitioned Bayesian Inference from complete or almost complete mitochondrial genomes (sequences of protein-coding genes only; see Fig. S1 for trees based on data sets including also non-coding rRNA and tRNA genes). Numbers at nodes are Bayesian posterior probabilities, followed by bootstrap proportions in percent from a ML (1000 replicates).

our analyses, it is obvious that most of the *Salamandra* RNAseq derived loci corresponded to multiple exons that are distant from each other in the genome, separated by long intronic sequences that were not included in our data set (as they are not translated into mRNA).

Furthermore, coalescence methods can be less accurate than concatenation when the gene trees have poor phylogenetic signal (Mirarab et al., 2016). The example of *S. atra* relationships indicates that only a minority of all gene trees supports the favored placement of the species, in agreement with other examples of lacking phylogenetic congruence among single-gene trees (e.g., Dikow and Smith, 2013). This indicates that the protein-coding nuclear loci derived from RNAseq might not be sufficiently informative for reconstructing single-gene trees. Also the RADseq loci correspond to very short sequences, each with few SNPs, that might not be suitable for calculating single-locus trees with adequate phylogenetic resolution among the various species. For these loci we therefore used a SNP-based species tree approach which however has not yet been extensively tested and thus might require methodological refinement.

Because our data sets might thus not be optimal for being analyzed with species tree approaches, we consider the phylogenies obtained by the concatenated analyses (Fig. 1) to be more reliable. As a further cause for the conflicting topologies, we emphasize that in *Salamandra*, the phylogenetic signal supporting relationships within the *atra-corsica-lanzai* clade is at best very weak, given that in RNAseq gene jackknifing, over 1 million base pairs are needed to stabilize the preferred concatenated topology with bootstrap support values >60%.

4.3. Evolutionary history and biogeography of Salamandra

We hypothesize that the phylogenies placing *S. atra* sister to *S. lanzai* (e.g., Fig. 1) represent most accurately the evolutionary history of the genus *Salamandra*. We base this hypothesis on: (1) the congruence of the trees obtained from separate concatenated analyses of the RNAseq and RADseq datasets (Fig. 1); and (2) the phenotypic similarity of *atra* and *lanzai* in many key traits. Both are species occurring in the Alps, entirely black-colored (except the subspecies *S. atra aurorae* and *S. atra pasubiensis*), and pueriparous, i.e., giving birth to fully metamorphosed juveniles (Fig. 5). In contrast, *S. corsica* has a geographic distribution restricted to the island of Corsica, is larviparous and yellow-black colored. If our preferred phylogenetic hypothesis is correct, then the alternative clade (*atra-corsica*) as strongly supported by the mitogenomic data probably reflects ancient hybridization, with introgression of the mitochondrial genome of an ancestral *S. atra* population into the ancestor of *S. corsica*, with replacement of the original mitochondrial genome of that species.

As pointed out by Vences et al. (2014), one of the main sources of disagreement in previous molecular studies of *Salamandra* phylogeny was the placement of the root: while the unrooted topology was almost fully congruent among all analyses published to date, the outgroup (*Lyciasalamandra*) in previous studies was connected alternatively to almost every branch in the *Salamandra* tree (see Fig. 3 in Vences et al., 2014), leading to radically different phylogenetic scenarios. Here, we provide rather strong evidence that the position of *S. infraimmaculata* as the sister group of all other *Salamandra* species is most likely the one correctly reflecting the evolution of these salamanders. This topology is stable across all

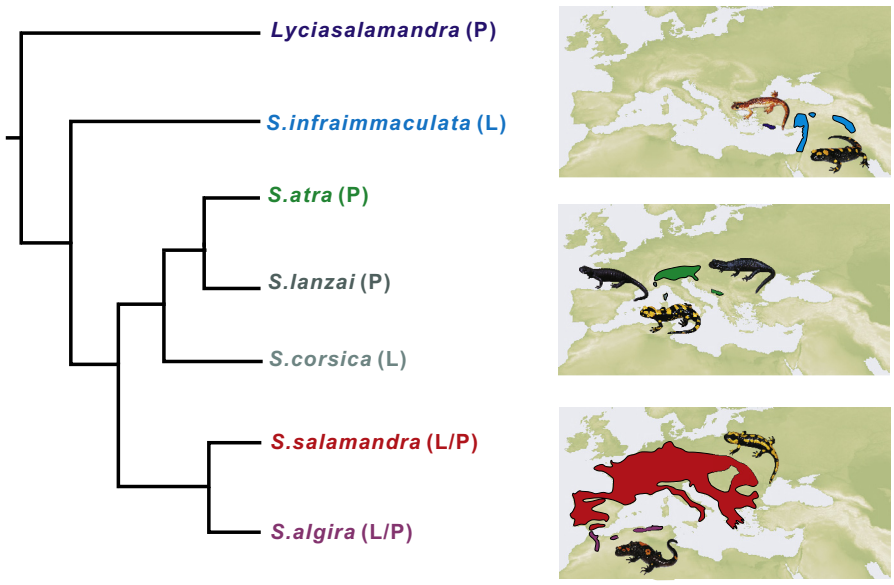

Fig. 5. Phylogenetic hypothesis of the genus *Salamandra*, based on concatenation analyses of nuclear gene data (RNAseq and RADseq), and maps showing approximate distribution of each species (summarized for all species of the outgroup clade *Lyciasalamandra)*. L and P after species names indicate whether the respective specie sare larviparous (L; giving birth to larvae), pueriparous (P; giving birth to fully metamorphosed juveniles), or comprise populations with both reproductive modes (L/P). Inset pictures are representative individuals of each taxon.

of the nuclear phylogenomic trees we resolved (Figs. 1 and 3) and was previously recovered in a phylogeny based on a small set of nuclear genes (Vences et al., 2014). In gene jackknifing this topology (i.e., the clade grouping all *Salamandra* to the exclusion of *S. infraimmaculata*) received support of 60% with a data set of 500,000 bp (RNAseq) and of 72% with 100,000 bp (RADseq), confirming it is highly supported by the data. Accepting this relationship of *S. infraimmaculata* as the sister species of all other *Salamandra*, the consensus topology of the nuclear gene data (Fig. 5) also suggests an origin of the genus in the Near East, considering that the two earliest branching clades (*Lyciasalamandra* and *S. infraimmaculata*) are restricted to this region.

4.4. Conclusion

The data presented here have shown the potential of phylogenomic data sets to elucidate shallow relationships among closely related taxa, even if these have probably been characterized by past episodes of introgression. It is encouraging that different commonly used phylogenomic approaches, such as RNAseq and RADseq, result in data sets that yield congruent results, despite having very distinctive properties. Yet, our results also confirm the need for caution in interpreting high bootstrap proportions or Bayesian posterior probability values: with an increase in quantity of phylogenomic data, high values of these classical support metrics can be misleading as they do not necessarily reflect a strong phylogenetic signal for a certain branch.

We are convinced that further improvement of analytical tools is of highest importance to deal with phylogenomic and phylotranscriptomic data sets, because model violations are to some degree inherent to all the methods used herein. For example, identification of the best fitting partition and substitution models is currently a computational hurdle for such large data sets and additional analytical tools are needed to better unravel, and critically assess, node support in both concatenation and species tree methods.

Although the massive data sets discussed herein provide a well-founded evolutionary and biogeographic hypothesis for the genus *Salamandra*, some doubts still remain on the relationships within this interesting group of terrestrial salamanders. Full genome sequences, currently prohibitive in costs, would allow a more conclusive understanding of past demography and possible episodes of introgressive hybridization among species of *Salamandra*, by identifying contiguous parts of genomic sequences affected by introgression. However, this might not necessarily result in a stronger phylogenetic signal. By the analysis of different comprehensive molecular data sets, we have definitely approached a limit to resolve the phylogenetic relations of these amphibians.

Acknowledgments

We are grateful to Marion Ballenghien and Nicolas Galtier for locality information on the sample used for their published *S. salamandra* transcriptome, to Martin Llewellyn for making computing resources available, to Alan Templeton for support with our research on *S. infraimmaculata*, to Franco Andreone and Philine Werner for help with obtaining crucial samples, and to David Donaire for continued support. We acknowledge the use of the Altamira Supercomputer (IFCA-CSIC), and support of the Alexander von Humboldt Foundation to II and AR, the European Molecular Biology Organization (EMBO) to II, the German Research Foundation (DFG) (STE1130/8-1 and BL 1271/1-1) as part of a German-Israeli project cooperation (DIP) grant to SS, AN and LB, and of the TULIP Laboratory of Excellence (ANR-10-LABX-41) to HP. This research was supported in part by a Natural Environment Research Council PhD studentship NE/L501918/1 (with KRE and BM) and a Systematics Research Fund award from the Linnean Society of London and the Systematics Association to JB. Support was provided by CalculQuébec (www.calculquebec.ca) and Compute Canada (www.computecanada.ca).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2017.07.009.

References

Amemiya, C.T., Alföldi, J., Lee, A.P., Fan, S., Philippe, H., Maccallum, I., Braasch, I., Manousaki, T., Schneider, I., Rohner, N., Organ, C., Chalopin, D., Smith, J.J., Robinson, M., Dorrington, R.A., Gerdol, M., Aken, B., Biscotti, M.A., Barucca, M., Baurain, D., Berlin, A.M., Blatch, G.L., Buonocore, F., Burmester, T., Campbell, M. S., Canapa, A., Cannon, J.P., Christoffels, A., De Moro, G., Edkins, A.L., Fan, L., Fausto, A.M., Feiner, N., Forconi, M., Gamieldien, J., Gnerre, S., Gnirke, A., Goldstone, J.V., Haerty, W., Hahn, M.E., Hesse, U., Hoffmann, S., Johnson, J., Karchner, S.I., Kuraku, S., Lara, M., Levin, J.Z., Litman, G.W., Mauceli, E., Miyake, T., Mueller, M.G., Nelson, D.R., Nitsche, A., Olmo, E., Ota, T., Pallavicini, A., Panji, S., Picone, B., Ponting, C.P., Prohaska, S.J., Przybylski, D., Saha, N.R., Ravi, V., Ribeiro, F.J., Sauka-Spengler, T., Scapigliati, G., Searle, S.M., Sharpe, T., Simakov, O., Stadler, P.F., Stegeman, J.J., Sumiyama, K., Tabbaa, D., Tafer, H., Turner-Maier, J., van Heusden, P., White, S., Williams, L., Yandell, M., Brinkmann, H., Volff, J.N., Tabin, C.J., Shubin, N., Schartl, M., Jaffe, D.B., Postlethwait, J.H., Venkatesh, B., Di Palma, F., Lander, E.S., Meyer, A., Lindblad-Toh, K., 2013. The African coelacanth genome provides insights into tetrapod evolution. Nature 496, 311–316.

Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Duruflé, L., Gaasterland, T., Lopez, P., Müller, M., Philippe, H., 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. Proc. Natl. Acad. Sci. USA 99, 1414–1419.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69, 313–319.

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

Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537.

Brandley, M.C., Bragg, J.G., Singhal, S., Chapple, D.G., Jennings, C.K., Lemmon, A.R., Lemmon, E.M., Thompson, M.B., Moritz, C., 2015. Evaluating the performance of anchored hybrid enrichment at the tips of the tree of life: a phylogenetic analysis of Australian *Eugongylus* group scincid lizards. BMC Evol. Biol. 15, 62.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29, 1917–1932.

Buckley, D., Alcobendas, M., García-París, M., Wake, M.H., 2007. Heterochrony, cannibalism, and the evolution of viviparity in *Salamandra salamandra*. Evol. Dev. 9, 105–115.

Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L., Waddell, P.J., 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42, 384–397.

Cariou, M., Duret, L., Charlat, S., 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecol. Evol. 3, 846–852.

Caspers, B.A., Krause, E.T., Hendrix, R., Koop, M., Rupp, O., Rosentreter, K., Steinfartz, S., 2014. The more the better - polyandry and genetic similarity are positively linked to reproductive success in a natural population of terrestrial salamanders (*Salamandra salamandra*). Mol. Ecol. 23, 239–250.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22, 3124–3140.

Chen, M.Y., Liang, D., Zhang, P., 2015. Selecting question-specific genes to reduce incongruence in phylogenomics: a case study of jawed vertebrate backbone phylogeny. Syst. Biol. 64, 1104–1120.

Chevreux, B., Wetter, T., Suhai, S., 1999. Genome sequence assembly using trace signals and additional sequence information. Comput. Sci. Biol. (GCB) 99, 45–56.

Chiari, Y., Cahais, V., Galtier, N., Delsuc, F., 2012. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). BMC Biol. 10, 65.

Czypionka, T., Krugman, T., Altmüller, J., Blaustein, L., Steinfartz, S., Templeton, A.R., Nolte, A.W., 2015. Ecological transcriptomics - a non-lethal sampling approach for endangered fire salamanders. Methods Ecol. Evol. 6, 1417–1425.

de Fonseca, R.R., Albrechtsen, A., Themudo, G.E., Ramos-Madrigal, J., Sibbesen, J.A., Maretty, L., Zepeda-Mendoza, M.L., Campos, P.F., Heller, R., Pereira, R.J., 2016. Next-generation biology: sequencing and data analysis approaches for non-model organisms. Mar. Genomics 30, 3–13.

Darwell, C.T., Rivers, D.M., Althoff, D.M., 2016. RAD-seq phylogenomics recovers a well-resolved phylogeny of a rapid radiation of mutualistic and antagonistic yucca moths. Syst. Entomol. 4, 672–682.

Davey, J.W., Blaxter, M.L., 2011. RADSeq: next-generation population genetics. Brief. Funct. Genomics 9, 416–423.

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetics inference and the multispecies coalescent. Trends Ecol. Evol. 24, 332–340.

Dell'Ampio, E., Meusemann, K., Szucsich, N.U., Peters, R.S., Meyer, B., Borner, J., Petersen, M., Aberer, A.J., Stamatakis, A., Walzl, M.G., Minh, B.Q., von Haeseler, A., Ebersberger, I., Pass, G., Misof, B., 2014. Decisive data sets in phylogenomics: lessons from studies on the phylogenetic relationships of primarily wingless insects. Mol. Biol. Evol. 31, 239–249.

Delsuc, F., Tsagkogeorga, G., Lartillot, N., Philippe, H., 2008. Additional molecular support for the new chordate phylogeny. Genesis 46, 592–604.

Dikow, R.B., Smith, W.L., 2013. Complete genome sequences provide a case study for the evaluation of gene-tree thinking. Cladistics 29, 672–682.

Eaton, D.A.R., 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30, 1844–1849.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461.

Emerson, K.J., Merz, C.R., Catchen, J.N., Hohenlohe, P.A., Cresko, W.A., Bradshaw, W. E., Holzapfel, C.M., 2010. Resolving postglacial phylogeography using high-throughput sequencing. Proc. Natl. Acad. Sci. USA 107, 16196–16200.

Figuet, E., Ballenghien, M., Romiguier, J., Galtier, N., 2014. Biased gene conversion and GC-content evolution in the coding sequences of reptiles and vertebrates. Genome Biol. Evol. 7, 240–250.

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. Bul. Am. Mus. Nat. Hist. 297, 1–370.

Gatesy, J., Springer, M.S., 2014. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. 80, 231–266.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Palma, F.D., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 29, 644–652.

Gregory, T.R., 2016. Animal genome size database. <http://www.genomesize.com> (accessed: December 2016).

Greven, H., Guex, G.D., 1994. Structural and physiological aspects of viviparity in *Salamandra salamandra*. Mertensiella 4, 139–160.

Greven, H., 2003. Larviparity and pueriparity. In: Sever, D.M. (Ed.), Reproductive Biology and Phylogeny, Reproductive Biology and Phylogeny of Urodela, vol. 1. Science Publisher Inc, Enfield, pp. 447–475.

Heled, J., Drummond, A.J., 2009. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nat. Protoc. 8, 1494–1512.

Hahn, C., Bachmann, L., Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads-a baiting and iterative mapping approach. Nucleic Acids Res. 41, e129.

Herrera, S., Shank, T.M., 2016. RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. Mol. Phylogenet. Evol. 100, 70–79.

Hill, G.E., 2016. Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. Ecol. Evol. 6, 5831–5842.

Huang, H., Knowles, L.L., 2014. Unforeseen consequences of excluding missing data from Next-Generation Sequences: simulation study of RAD sequences. Syst. Biol. 65, 357–365.

Irisarri, I., Meyer, A., 2016. The identification of the closest living relative(s) of tetrapods: phylogenomic lessons for resolving short ancient internodes. Syst. Biol. 65, 1057–1075.

Irisarri, I., Baurain, D., Brinkmann, H., Delsuc, F., Sire, J. Yy., Kupfer, A., Petersen, J., Jarek, M., Meyer, A., Vences, M., Philippe, H., 2017. Phylotranscriptomic consolidation of the jawed vertebrate timetree. Nat. Ecol. Evol. http://dx.doi.org/10.1038/s41559-017-0240-5. in press.

Leaché, A.D., Rannala, B., 2011. The accuracy of species tree estimation under simulation: a comparison of methods. Syst. Biol. 60, 126–137.

Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y., Faircloth, B.C., Nabholz, B., Howard, J.T., Suh, A., Weber, C.C., da Fonseca, R.R., Li, J., Zhang, F., Li, H., Zhou, L., Narula, N., Liu, L., Ganapathy, G., Boussau, B., Bayzid, M.S., Zavidovych, V., Subramanian, S., Gabaldon, T., Capella-Gutierrez, S., Huerta-Cepas, J., Rekepalli, B., Munch, K., Schierup, M., Lindow, B., Warren, W.C., Ray, D., Green, R.E., Bruford, M.W., Zhan, X., Dixon, A., Li, S., Li, N., Huang, Y., Derryberry, E.P., Bertelsen, M.F., Sheldon, F.H., Brumfield, R.T., Mello, C.V., Lovell, P.V., Wirthlin, M., Schneider, M.P., Prosdocimi, F., Samaniego, J.A., Vargas Velazquez, A.M., Alfaro-Nunez, A., Campos, P.F., Petersen, B., Sicheritz-Ponten, T., Pas, A., Bailey, T., Scofield, P., Bunce, M., Lambert, D.M., Zhou, Q., Perelman, P., Driskell, A.C., Shapiro, B., Xiong, Z., Zeng, Y., Liu, S., Li, Z., Liu, B., Wu, K., Xiao, J., Yinqi, X., Zheng, Q., Zhang, Y., Yang, H., Wang, J., Smeds, L., Rheindt, F.E., Braun, M., Fjeldsa, J., Orlando, L., Barker, F.K., Jonsson, K.A., Johnson, W., Koepfli, K.P., O'Brien, S., Haussler, D., Ryder, O.A., Rahbek, C., Willerslev, E., Graves, G.R., Glenn, T.C., McCormack, J., Burt, D., Ellegren, H., Alstrom, P., Edwards, S.V., Stamatakis, A., Mindell, D.P., Cracraft, J., Braun, E.L., Warnow, T., Jun, W., Gilbert, M.T., Zhang, G., 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346, 1320–1331.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56, 17–24.

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.

Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61, 727–744.

Le, S.Q., Gascuel, O., 2008. LG: an improved, general amino-acid replacement matrix. Mol. Biol. Evol. 25, 1307–1320.
Li, L., Stoecker, C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189.
Liu, L., Yu, L., Kubatko, L.S., Pearl, D.K., Edwards, S.V., 2009. Coalescent methods for estimating phylogenetic trees. Mol. Phylogenet. Evol. 53, 320–328.
Lowe, T.M., Chan, P.P., 2016. TRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 44, W54–W57.
Machado, D.J., Lyra, M., Grant, T., 2016. Mitogenome assembly from genomic multiplex libraries: comparison of strategies and novel mitogenomes for five species of frogs. Mol. Ecol. Res. 16, 686–693.
Mirarab, S., Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, i44–i52.
Mirarab, S., Bayzid, M.S., Warnow, T., 2016. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Syst. Biol. 65, 366–380.
Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspock, U., Aspock, H., Bartel, D., Blanke, A., Berger, S., Bohm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B. M., Schutte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K.M., Zhou, X., 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346, 763–767.
Moustafa, A., Bhattacharya, D., 2008. PhyloSort: a user-friendly phylogenetic sorting tool and its application to estimating the cyanobacterial contribution to the nuclear genome of *Chlamydomonas*. BMC Evol. Biol. 8, 6.
Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7, e37135.
Philippe, H., Delsuc, F., Brinkmann, H., Lartillot, N., 2005. Phylogenomics. Annu. Rev. Ecol. Evol. Syst. 36, 541–562.
Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G., Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9, e1000602.
Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., Lemmon, A.R., 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526, 569–573.
Pyron, R.A., 2014. Biogeographic analyses reveals ancient continental vicariance and recent oceanic dispersal in Amphibia. Syst. Biol. 63, 779–797.
Rambaut, A., Drummond, A.J., 2007. Tracer: MCMC Trace Analysis Tool. Institute of Evolutionary Biology, University of Edinburgh.
Recknagel, H., Jacobs, A., Herzyk, P., Elmer, K.R., 2015. Double-digest RAD sequencing using Ion Proton semiconductor platform (ddRADseq-ion) with nonmodel organisms. Mol. Ecol. Res. 15, 1316–1329.
Rivers, D.M., Darwell, C.T., Althoff, D.M., 2016. Phylogenetic analysis of RAD-seq data: examining the influence of gene genealogy conflict on analysis of concatenated data. Cladistics 32, 672–681.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.
Roure, B., Rodriguez-Ezpeleta, N., Philippe, H., 2007. SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. BMC Evol. Biol. 7, S2.
Rubin, B.E.R., Ree, R.H., Moreau, C.S., 2012. Inferring phylogenies from RAD sequence data. PLoS ONE 7, e33394.
Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33, 1654–1668.
Sillero, N., Campos, J., Bonardi, A., Corti, C., Creemers, R., Crochet, P.-A., Isailovic, J.C., Denoël, M., Ficetola, G.F., Gonçalves, J., Kuzmin, S., Lymberakis, P., Pous, P.d., Rodríguez, Sindaco, R., Speybroeck, J., Toxopeus, B., Vieites, D.R., Vences, M., 2014. Updated distribution and biogeography of amphibians and reptiles of Europe. Amphibia-Reptilia 35, 1–31.
Speybroeck, J., Beukema, W., Crochet, P.-A., 2010. A tentative species list of the European herpetofauna (Amphibia and Reptilia). Zootaxa 2492, 1–27.
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, 1312–1313.
Steinfartz, S., Vicario, S., Arntzen, J.W., Caccone, A., 2007. A bayesian approach on molecules and behavior: reconsidering phylogenetic and evolutionary patterns of the Salamandridae with emphasis on *Triturus* Newts. J. Exp. Zool. B Mol. Dev. Evol. 308B, 139–162.
Takahashi, T., Nagata, N., Sota, T., 2014. Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. Mol. Phylogenet. Evol. 80, 137–144.
Titus, T.A., Larson, A., 1995. A molecular phylogenetic perspective on the evolutionary radiation of the salamander family Salamandridae. Syst. Biol. 44, 125–151.
Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., Ortí, G., 2015. Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. PLoS Currents Tree of Life, first ed. http://dx.doi.org/10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be.
Tucker, D.B., Colli, G.R., Giugliano, L.G., Hedges, S.B., Hendrye, C.R., Lemmon, E.M., Lemmon, A.R., Sites, J.W., Pyron, R.A., 2016. Methodological congruence in phylogenomic analyses with morphological support for teiid lizards (Sauria: Teiidae). Mol. Phylogenet. Evol. 103, 75–84.
Veith, M., Steinfartz, S., 2004. When non-monophyly results in taxonomic consequences—the case of *Mertensiella* within the Salamandridae (Amphibia: Urodela). Salamandra 40, 67–80.
Veith, M., Steinfartz, S., Zardoya, R., Seitz, A., Meyer, A., 1998. A molecular phylogeny of "true" salamanders (family Salamandridae) and the evolution of terrestriality of reproductive modes. J. Zool. Syst. Evol. Res. 37, 7–16.
Vences, M., Sanchez, E., Hauswaldt, J.S., Eikelmann, D., Rodríguez, A., Carranza, S., Donaire, D., Gehara, M., Helfer, V., Lötters, S., Werner, P., Schulz, S., Steinfartz, S., 2014. Nuclear and mitochondrial multilocus phylogeny and survey of alkaloid content in true salamanders of the genus *Salamandra* (Salamandridae). Mol. Phylogenet. Evol. 73, 208–216.
Wake, M.H., 1993. Evolution of oviductal gestation in amphibians. J. Exp. Zool. 266, 394–413.
Wang, H.-J., Li, W.-T., Liu, Y.-N., Yang, F.-S., Wang, X.-Q., 2017. Resolving interspecific relationships within evolutionarily young lineages using RNA-seq data: an example from Pedicularis section Cyathophora (Orobanchaceae). Mol. Phylogenet. Evol. 107, 345–355.
Weisrock, D.W., Macey, J.R., Ugurtas, I.H., Larson, A., Papenfuss, T.J., 2001. Molecular phylogenetics and historical biogeography among salamandrids of the "true" salamander clade: rapid branching of numerous highly divergent lineages in *Mertensiella luschani* associated with the rise of Anatolia. Mol. Phylogenet. Evol. 18, 434–448.
Weisrock, D.W., Papenfuss, T.J., Macey, J.R., Litvinchuk, S.N., Polymeni, R., Ugurtas, I. H., Zhao, E., Jowkar, H., Larson, A., 2006. A molecular assessment of phylogenetic relationships and lineage accumulation rates within the family Salamandridae (Amphibia, Caudata). Mol. Phylogenet. Evol. 41, 368–383.
Wen, J., Egan, A.N., Dikow, R.B., Zimmer, E.A., 2015. Utility of transcriptome sequencing for phylogenetic inference and character evolution. In: Hörandl, E., Appelhans, M.S. (Eds.), Next-generation sequencing in plant systematics. International Association for Plant Taxonomy (IAPT), pp. 51–91.
Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., Ruhfel, B.R., Wafula, E., Der, J.P., Graham, S.W., Mathews, S., Melkonian, M., Soltis, D.E., Soltis, P.S., Miles, N.W., Rothfels, C.J., Pokorny, L., Shaw, A.J., DeGironimo, L., Stevenson, D.W., Surek, B., Villarreal, J.C., Roure, B., Philippe, H., dePamphilis, C.W., Chen, T., Deyholos, M.K., Baucom, R.S., Kutchan, T.M., Augustin, M.M., Wang, J., Zhang, Y., Tian, Z., Yan, Z., Wu, X., Sun, X., Wong, G.K., Leebens-Mack, J., 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. USA 111, E4859–4868.
Zhang, P., Papenfuss, T.J., Wake, M.H., Qu, L., Wake, D.B., 2008. Phylogeny and biogeography of the family Salamandridae (Amphibia: Caudata) inferred from complete mitochondrial genomes. Mol. Phylogenet. Evol. 49, 586–597.