Phylogeny and species delimitation of Near Eastern *Neurergus* newts (Salamandridae) based on genome-wide RADseq data analysis

Loïs Rancilhac[1], Forough Goudarzi[1,2], Marcelo Gehara[3], Mahmoud-Reza Hemami[2], Kathryn R. Elmer[4], Miguel Vences[1], Sebastian Steinfarz[1].

[1]Zoological Institute, Technische Universität Braunschweig, Mendelssohnstrasse 4, 38106 Braunschweig, Germany.

[2]Department of Natural Resources, Isfahan University of Technology, Isfahan, Iran.

[3] American Museum of Natural History, Department of Herpetology, Central Park West at 79th St, New York, NY 10024, USA.

[4] Institute of Biodiversity, Animal Health & Comparative Medicine, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK

Corresponding author: Loïs Rancilhac (loisrancilhac@gmail.com)

Zoological Institute, Technische Universität Braunschweig, Mendelssohnstrasse 4, 38106 Braunschweig, Germany

loisrancilhac@gmail.com

Short title: Phylogeny of *Neurergus* newts

Abstract

We reconstruct the molecular phylogeny of Near Eastern mountain brook newts of the genus *Neurergus* (family Salamandridae) based on newly determined RADseq data, and compare the outcomes of concatenation-based phylogenetic reconstruction with species-tree inference. Furthermore, we test the current taxonomy of *Neurergus (*with four species: *Neurergus strauchii, N. crocatus, N. kaiseri,* and *N. derjugini*) against coalescent-based species-delimitation approaches of our genome-wide genetic data set. While the position of *N. strauchii* as sister species to all other *Neurergus* species was consistent in all of our analyses, the phylogenetic relationships between the three remaining species changed depending on the applied method. The concatenation approach, as well as quartet-based species-tree inference, supported a topology with *N. kaiseri* as the closest relative to *N. derjugini*, while full-coalescent species-tree inference approaches supported *N. crocatus* as sister species of *N. derjugini*. Investigating the individual signal of gene trees highlighted an extensive variation among gene histories, most likely resulting from incomplete lineage sorting. Coalescent-based species-delimitation models suggest that the current taxonomy might underestimate the species richness within *Neurergus* and supports seven species. Based on the current sampling, our analysis suggests that *N. strauchii*, *N. derjugini* and *N. kaiseri* might each be subdivided into further species. However, as amphibian species are known to be composed of deep conspecific lineages that do not always warrant species status, these results need to be cautiously interpreted in an integrative taxonomic framework. We hypothesize that the rather shallow divergences detected within *N. kaiseri* and *N. derjugini* likely reflect an ongoing speciation process and thus require further investigation. On the contrary, the much deeper genetic divergence found between the two morphologically and geographically differentiated subspecies of *N. strauchii* leads us to propose that *N. s. barani* should be considered a distinct species, *N. barani* (Öz 1994).

Keywords: Neurergus, Multi-Species Coalescent, Species-tree inference, Species delimitation, ddRAD sequencing, *Neurergus barani*.

1. Introduction

Mountain brook newts of the genus *Neurergus* (Cope 1862) form a monophyletic group occurring in the Near East, from Turkey to Iraq and Iran. Although a few studies have focused on different aspects of *Neurergus* biology (Bogaerts et al., 2006; Hendrix et al., 2014; Schneider and Schneider, 2010; Sparreboom et al., 2000; Steinfartz et al., 2002), our knowledge on their ecology, behavior and evolutionary history remains – compared to other salamandrid genera such as European newts and salamanders – understudied, mainly because of their remote ranges in countries difficult to access (Hendrix et al., 2014). All *Neurergus* species inhabit geographically restricted areas (Fig. 1) and are included in threatened conservation categories, ranging from vulnerable to critically endangered (IUCN SSC Amphibian Specialist Group, 2016; Papenfuss et al., 2009; Papenfuss et al., 2009; Sharifi et al., 2009). Thus, it is important to elucidate their biology and ecology in order to inform effective conservation strategies. In this context, knowledge on the phylogeny, systematics and taxonomy of these threatened amphibians is key to delimiting where conservation efforts need to focus. Fine-scale phylogenetic and phylogeographic studies can be used to delimit such conservation units (Vogler and Desalle, 1994), and furthermore provide the means to analyse the genetic structure of populations as a basis for downstream conservation measures (e.g. Dincă et al., 2018). As conservation assessments and legal documents traditionally use species as principle conservation units, it is also important to have a correct species-level taxonomy, particularly when cryptic, or near-cryptic, species are involved (Delić et al., 2017).

In *Neurergus*, four species are currently recognised based on molecular (mitochondrial and nuclear genes) and morphological characters: *N. crocatus*, *N. derjugini* (including the subspecies *derjugini* and *microspilotus*), *N. kaiseri*, and *N. strauchii* (including

the subspecies *barani*, *strauchii* and *munzurensis*). The monophyly of the genus has been well established based on nuclear allozyme and mitochondrial DNA sequence data (Steinfartz et al., 2002) and phylogenetic reconstructions within the Salamandridae, based on complete mitogenomes and nuclear sequences (Kieren et al., 2018; Veith et al., 2018; Zhang et al., 2008), placed it as sister group to Near Eastern newts of the genus *Ommatotriton*. The relationships within *Neurergus* have also been studied using a variety of molecular markers (Hendrix et al., 2014; Steinfartz et al., 2002), but always based on a limited number of loci. Consequently, evolutionary relationships among some *Neurergus* species remain poorly resolved and the taxonomic status of the *N. derjugini* and *N. strauchii* subspecies *N. derjugini* and *N. strauchii* remain uncertain.

Considering the enormous methodological progress in the genomic era, it has become possible to resolve phylogenetic relationships and to investigate species boundaries based on markers representing evolutionary diversification processes at different parts of an organism's genome. A variety of approaches exist to sample loci at the genomic scale for such purposes. Among other methods, it is possible to sequence only the expressed regions of the genome (RNA sequencing or RNAseq, Ozsolak and Milos, 2011), to target specific regions of the genome (e.g. exon capture, Bragg et al., 2016; enrichment of Ultra Conserved Elements or UCEs, Faircloth et al., 2012), or to sequence randomly small DNA fragments (e.g. Restriction site Associated DNA sequencing, Miller et al., 2007; Genotyping By Sequencing, Elshire et al., 2011). However, it can be challenging to ensure homology of loci across taxa and to verify assumptions such as absence of selection and of intra-locus recombination. Thus, a careful choice of markers is paramount, particularly for poorly known non-model organisms.

Restriction site Associated DNA sequencing (RADseq) has become popular in recent years for phylogenetic, population genetic and systematic studies, as it can generate sequences of a large number of independent loci for a moderate cost (Andrews et al., 2016). In brief, genomes are digested with one or several restriction enzymes, and a short DNA fragment flanking the restriction site is sequenced, resulting in hundreds of small fragments that make a reduced representation of the whole genome available to analysis (Miller et al., 2007). Even though this method was originally used for inferences at the intraspecific level (Ree and Hipp, 2015), both empirical (e.g. Ebel et al., 2015; Manthey et al., 2016; Razkin et al., 2016) and *in silico* (e.g. Cariou et al., 2013; Rubin et al., 2012) studies have shown its usefulness to reconstruct well resolved phylogenies, even at relatively deep evolutionary scales (e.g. Lecaudey et al., 2018), with performances comparable to other genomic approaches (e.g. full genomes: Cariou et al., 2013; UCE targeting: Manthey et al., 2016; RNAseq: Rodríguez et al., 2017). As in more classical multilocus approaches, RADseq loci can be concatenated to infer phylogenies. The resulting concatenation matrices are usually very large (hundred of thousands to millions of bp), holding numerous informative sites, and thus generally lead to accurate inferences (Ree and Hipp, 2015). However, in cases of variation among gene histories, species-tree inference may become necessary for correct phylogenetic reconstruction, and different approaches are available for this purpose.

The more popular methods (e.g. *BEAST, Heled and Drummond, 2010; ASTRAL, Mirarab et al., 2014) use gene trees (which are jointly estimated or taken as input) to infer a species-tree that is consistent under the multi-species coalescent model. However, gene trees inferred from RADseq loci usually have very low resolution as the sequences are very short (Harvey et al., 2016) and the high number of loci makes the analyses very long and intensive. Methods inferring species-trees from bi-allelic markers, without explicitly estimating gene

trees (e.g. SNAPP, Bryant et al., 2012; SVDquartets, Chifman and Kubatko, 2014) could be more appropriate to RADseq data as these can sample unlinked SNPs at a genomic scale (Peterson et al., 2012). Yet, it is still unclear which method should be best applied for RADseq data at a given evolutionary scale. This uncertainty also applies to coalescent-based species-delimitation, as one of the most popular approaches, BPP (Yang, 2015), requires a resolved species-tree, either estimated from gene trees or specified by the user. This method has been used with RADseq data (e.g. Nieto-Montes de Oca et al., 2017), but its ability to reconstruct the true species-tree from these data, as well as the impact of errors in the species-tree on the species-delimitation, are not clear. Here, we apply a RADseq approach to sample several thousand loci across the genome of all described species of *Neurergus*. Using these data, we (i) infer the phylogeny of the genus from a matrix of concatenated loci and from (ii) various alternative species-tree methods, and (iii) investigate species boundaries using several coalescent-based species-delimitation approaches. The aim of this study is to resolve the phylogeny and systematics of this understudied genus, but also to explore new empirical insights on the performance and congruence of alternative coalescent-based species-tree reconstruction and species-delimitation methods on RADseq data.

2. Material and Methods

2.1 Sampling and laboratory work

Tissue samples were collected from all the species and subspecies of *Neurergus* (except *N. s. munzurensis*), for a total of 17 individuals from 13 locations (Fig. 1). Due to different sampling efforts, the distribution of samples is unequal among species, and especially samples of *N. kaiseri* are overrepresented. Additionally, one individual of both *Triturus cristatus* and *Ommatotrion vittatus* were sampled as outgroups. Genomic DNA was extracted

using the Macherey-Nagel NucleoSpin Tissue kit following the manufacturer's instructions. We performed double-digest Restriction Site Associated DNA sequencing (ddRADseq; Peterson et al., 2012) preparing the library 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 (NewEngland Biolabs); modified Illumina adaptors with unique barcodes for each individual were ligated onto this fragmented DNA; samples were multiplexed (pooled); and a PippinPrep was 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.

2.2 Assembly of RADseq data

De-multiplexed reads were assembled *de novo* using ipyrad (Eaton, 2014; Eaton, 2015). This method clusters reads depending on their similarity and then calls bases above a user-defined minimum depth. In order to avoid introducing errors in the dataset or clustering paralogous loci, we carefully investigated the impact of these parameters on the number of recovered loci and the phylogenetic accuracy of the obtained dataset. First, we assembled the dataset with a minimum depth ranging from 2 to 10, with the clustering threshold set as 85%. Then, once the minimum depth was selected, we performed 10 assemblies with clustering thresholds ranging from 85% to 95%. For the values of 85%, 90% and 95%, we filtered the datasets with different minimum numbers of individuals per locus (4, 8, 10, 14, 16 and 18, corresponding to 20%-95% of individual coverage) and inferred phylogenies using RAxML (Stamatakis, 2014)

with 100 rapid bootstraps under a GTR+GAMMA substitution model on the concatenation of all loci, and the concatenation of unlinked SNPs (i.e. one SNP sampled at each locus, corresponding either to the one with the highest coverage or, in case of equal coverage, randomly chosen). As some analyses do not require the inclusion of outgroups (e.g. some species-tree inference methods), we also assembled a dataset containing only the *Neurergus* samples. Indeed, removing distant species should increase the amount of data recovered and the accuracy of the analyses (Pante et al. 2015). For this second dataset, we performed the same series of tests than for the first one, with minimum individual coverage ranging from 4 to 16 (23%-94% of the total number of individuals). In the end, a total of 66 different datasets were assembled and trees inferred (Tab. S1). For each of the trees, the average bootstrap support of the "deep nodes" (i.e. relationship between the *Neurergus* species) and the "shallow nodes" (i.e. nodes at specific and intra-specific level) were calculated. The combination of parameters maximizing the bootstrap support of the phylogeny and the amount of loci recovered was then used for the final analyses. After performing the final assemblies of both datasets, ipyrad was used to format the data in different ways: concatenation of all the loci, sequences for each locus, and concatenation of unlinked SNPs extracted from the loci.

2.3 Concatenation analysis

The best model of substitution for the concatenated loci was selected using the Bayesian Information Criterion (BIC) implemented in ModelGenerator (Keane et al., 2006). Phylogenetic relationships were then inferred using RAxML with the new rapid hill-climbing algorithm on the concatenated matrix, with *Triturus cristatus* and *Ommatotriton vittatus* as outgroups. 100 bootstrap replicates were performed to assess the support of the nodes.

Because for extensive molecular datasets, bootstrap values can sometimes be misleading and provide high support for spurious nodes (Rodríguez et al., 2017; Rubin et al., 2012), we also performed gene jackknifing (Delsuc et al., 2008): loci were randomly sampled up to ca. 10,000, 100,000, 500,000 and 800,000 positions, and 100 random draws were made for each value. For each replicate, RAxML was run with the same setup as previously, and the proportion of replicates supporting each node of the complete tree was calculated using the bipartition model implemented in RAxML. In order to quantify the genetic divergence between the phylogenetic lineages, pairwise nucleotide diversities between all groups were calculated as the average across all loci using the R package PopGenome (Pfeifer et al., 2014).

2.3.2 Species tree inference

We inferred the species-tree using various approaches. First, we used SNAPP (Bryant et al., 2012), a method to infer a species-tree from unlinked SNPs without estimating gene trees. This model uses a Bayesian approach that is implemented in BEAST2 (Bouckaert et al., 2014). For this analysis, individuals were clustered into putative species on the basis of the concatenation tree. As this analysis implements the coalescent model which does not require rooting to polarize the topology, it was run on the dataset without outgroups, with 15,000,000 mcmc steps and sampling every 10,000 steps, and with a G (2, 1000) prior distribution on population size and default priors on mutation rates (mutation rate U = 1.0, mutation rate V = 1.0). The convergence of the analysis was checked using Tracer 1.7 (Rambaut et al., 2018). We then implemented two methods that reconstruct the species-tree from quartet trees. Tetrad (Eaton, 2015b) features a method similar to SVDquartets (Chifman and Kubatko, 2014): quartet trees are inferred from unlinked SNPs, and joined together into a species-tree that is

consistent under the multi-species coalescent. Thus, the analysis was first performed on the dataset including the two outgroups, with all possible quartets sampled, and with 500 non-parametric bootstrap replicates. To test for the effect of the amount of data on the support of the nodes, the method was also performed on the dataset without outgroups, with the same parameters. We also used ASTRAL-II (Mirarab and Warnow, 2015), which implements a similar method, but the quartet trees are calculated from gene trees. Thus, the gene trees for all loci were inferred from the *Neurergus*-specific dataset, using RAxML with the rapid hill-climbing algorithm and a JC69 substitution model, and with 100 bootstrap replicates for each gene tree. Then, the species-tree was inferred using ASTRAL-II, and again 100 bootstrap replicates were performed.

2.3.3 Phylogenetic signal of gene trees

To resolve controversial relationships, the signal of individual gene trees was investigated. The two conflicting topologies differ in the position of *N. kaiseri* as the sister taxon of *N. derjugini* and *N. d. microspilotus* (Concatenated, tetrad and ASTRAL-II; see Results), or as *N. crocatus* as the sister taxon to *N. derjugini* and *N. d. microspilotus* (BPP and SNAPP; see Results). Under ILS alone, one of the topologies described above should represent the true history and occur in higher frequency in the population of gene trees. Thus, the proportion of genes supporting each alternative topology, as well as a third alternative with *N. crocatus* as sister species to *N. kaiseri*, was calculated with a custom R script, using the package ape (Paradis et al., 2004). We used the same gene trees as for the ASTRAL-II analysis. For each gene tree we sampled one individual per species and verified the monophyly of each alternative group. We ran 100 independent replicates of this analysis to capture the variation generated by sampling only one individual per species.

We used ASTRAL-II to compute the quartet scores of the different competing topologies. This is another way to measure the overall agreement of the gene trees with a topology; the quartet score is the proportion of quartet trees induced by all the gene trees that are present in the species-tree (Mirarab et al., 2014).

2.4 Species delimitation

The putative species used in the SNAPP analysis were tested using the software BPP (Bayesian Phylogeny and Phylogeography; Yang, 2015; Yang and Rannala, 2010). The species-tree and species-delimitation were jointly inferred (analyze A11). A gamma distribution G(2, 1000) was used as a prior on the population-size parameter, and the distributions G(2, 100) and G(2, 10000) were also tested to check for the effect of the prior on the results. A gamma distribution G(10, 1000) was used as a prior on the age of the root of the tree, corresponding to the time of divergence of the *Neurergus* inferred from mitochondrial markers by the study of Steinfartz et al. (2002), and the other splitting times were inferred using a dirichlet process. The analysis was set with 100,000 burnin steps and 200,000 steps, with sampling every 10 steps. In case the species-tree inferred with BPP differed from the one inferred with tetrad, the latter was used to perform species-delimitation with a fixed guide tree with BPP (analysis A10) to test for the effect of uncertainty on the species-tree on the species-delimitation. Due to limited computational resources, the whole dataset could not be analyzed at once. Thus, 5 datasets of 500 randomly drawn loci were built and analyzed independently, a method first used by Nieto-Montes de Oca et al. (2017).

Species boundaries were also investigated using BFD* (Leaché et al., 2014). This method computes the marginal likelihood of species-trees inferred with SNAPP using the PathSampling algorithm implemented in BEAST2. Different models can be tested by

lumping and splitting populations or reassigning individuals, and their marginal likelihood can be compared to find the best fitting model. Alternative species delimitation hypotheses were defined using the concatenation tree (Table 3), and their marginal likelihood was calculated by sampling 48 steps with the following parameters: 400,000 MCMC steps with 50% burnin and 10,000 pre-burnin steps. MCMC parameters for species-tree inference were left as for the previous SNAPP analysis. For each model, two independent runs were performed, and the average Marginal Likelihood was retained.

3. Results

3.1 RADseq dataset reconstruction

Processed reads ranged from 781,375 to 9,365,666 per sample. The characteristics of the 66 datasets assembled are given in Table S1. Varying the clustering threshold from 85% to 95% had mostly no impact on the topology inferred. Out of the 66 trees inferred, 63 of them yielded the topology given in Fig. 2, usually with high bootstrap supports, with the three inconsistent topologies resulting from assemblies with very few loci (unlinked SNPs with high individual coverage, Fig. S1). Among these three, two placed *N. crocatus* as sister species to *N. kaiseri*, and one failed to properly recover the intra-specific structure of *N. kaiseri*. The final dataset was assembled using a minimum depth of 8, a clustering threshold of 89%, and a filtering threshold that excluded all loci with fewer than 16 individuals (14 for the dataset without outgroups, respectively corresponding to 82% and 84% of the individuals). Finally, we recovered 8,314 loci with an average length of 115 bp (min: 109 bp; max 127 bp). The final matrix includes 958,256 bp, with a total of 12.24% missing data (see Table 1 for more details). As expected, removing the outgroup taxa significantly increased the

number of loci: the *Neurergus*-specific dataset contained 15,547 loci (1,791,783 bp). From these two datasets, respectively, 8,314 and 14,765 supposedly unlinked SNPs were extracted.

Table 1. Summary statistics for the two datasets assembled.

Dataset	Length of the concatenation	Number of loci	Number of unlinked SNP	Percentage of missing data	Number of variable sites	Number of parsimony informative sites
With outgroups	958,256 bp	8,314	8,314	12.24%	40,781 (4.3%)	25,799 (2.7%)
Without outgroups	1,791,783 bp	15,547	14,765	9.3%	52,795 (2.9%)	44,343 (2.5%)

3.2 Phylogenetic analyses

3.2.1 Concatenation analyses

The phylogenetic tree inferred from the concatenated matrix of all loci was fully resolved, with 100% bootstrap support at every node (Fig. 2). Gene jackknifing also recovered full support for all nodes from ca. 100,000 bp sampled. The tree can be roughly divided in two groups: *N. strauchii*, with a deep divergence between the subspecies *strauchii* and *barani* (average pairwise nucleotide diversity = 0.0056), and a monophyletic group including *N. crocatus*, *N. derjugini* and *N. kaiseri*. Within this clade, *N. derjugini* and *N. kaiseri* are sister lineages. The subspecies of *N. derjugini* show very shallow divergence (smaller than the intraspecific divergence between the two localities of *N. crocatus*; average pairwise nucleotide diversity = 0.0035), but are reciprocally monophyletic with high support. *N. kaiseri* displays a high genetic diversity between localities and can be split into two main clades, corresponding to locations found in the southern (Ab Palangi and Shevi fall) and northern (Bozorg Ab, Barik Ab, Tova, Shahbazan and Nergeseh) parts of the distribution

range. For the subsequent species-tree and species-delimitation inferences, all described species and subspecies, as well as both clades of *N. kaiseri*, were considered putative species. The matrix of the pairwise nucleotide diversities is displayed in Table 2.

Table 2. Nucleotide diversity between all pairs of lineages, averaged on all the loci

	N. s. strauchii	***N. s. barani***	***N. crocatus***	***N. d. derjugini***	***N. d. microspilotus***	***N. kaiseri* N**
N. s. strauchii	--	--	--	--	--	--
N. s. barani	0.0056	--	--	--	--	--
N. crocatus	0.0073	0.0074	--	--	--	--
N. d. derjugini	0.0075	0.0077	0.0071	--	--	--
N. d. microspilotus	0.0076	0.0077	0.0071	0.0035	--	--
***N. kaiseri* N**	0.0072	0.0073	0.0066	0.0067	0.0068	--
***N. kaiseri* S**	0.0074	0.0076	0.0070	0.0072	0.0073	0.0039

3.2.2 Species-tree inference

The topology of the species-tree inferred with tetrad was consistent with the concatenation tree (hereafter referred to as “topology A”; Fig. 3a). All the nodes received full support, except for the one at the root of the group including *N. crocatus*, *N. derjugini* and *N. kaiseri*, supported by only 74% of the bootstrap replicates. When running the analysis on the dataset without outgroups, the topology remained the same and the support of this node increased to

33894%. ASTRAL-II revealed the same topology, except that *N. d. derjugini* was paraphyletic.
339The tree inferred with SNAPP, on the other hand, showed a different topology (hereafter
340referred to as “topology B”; Fig. 3b), placing *N. kaiseri* as sister species of a clade constituted
341of *N. crocatus* and *N. derjugini*. All the nodes had a posterior probability of 1. However, even
342with 15,000,000 steps (running for a total of ca. 154 h using 10 cores), the MCMC run did
343not converge properly as the trace of the topology posterior probability showed a clearly
344increasing trend, and the ESS remained very low. When using BPP to infer the species-tree
345jointly with species-delimitation, we also obtained “topology B”, and with good convergence
346of the MCMC chain.
347
348*3.2.3 Phylogenetic signal of gene trees*
349To understand the support from single markers for each of the alternative topologies A and B,
350the signal of the gene trees was investigated. An average of 18% of the loci supported
351topology B, while 17% supported topology A (Fig. 4). Additionally, 13 % of loci supported
352the position of *N. crocatus* as sister species to *N. kaiseri*. Topology A obtained a quartet score
353of 0.601, and topology B of 0.636.
354
355*3.3 Species-delimitation*
356Using BPP, the seven *a priori* defined taxa were all delimited as different species with high
357posterior probability (i.e. $P > 0.95$) in 17 analyses out of the 20 performed. In the three other
358cases, the two subspecies of *N. derjugini* were lumped into a single species, sometimes with
359full support (P=1). Variations on the theta prior had no significant impact on the species
360delimitation. The species-tree inferred with BPP was identical to the one inferred with

SNAPP (topology B). Using the alternative topology as fixed guide tree to perform species-delimitation did not change the results.

The BFD* analyses supported the BPP results, as the seven-species model obtained the highest marginal likelihood (Table 3).

Table 3. Summary of species-delimitation models and their respective marginal likelihood values, tested using BFD*, ranked by decreasing marginal likelihood.

Number of species	Model	Marginal L
7	Split *N. strauchii*, *N. derjugini* and *N. kaiseri*	-94087.385
6	Split *N. strauchii* and *N. kaiseri*	-94616.512
6	Split *N. derjugini* and *N. kaiseri*	- 95608.314
5	Split *N. kaiseri*	-96079.767
6	Split *N. strauchii* and *N. derjugini*	-96290.751
5	Split *N. strauchii*	-96875.032
5	Split *N. derjugini*	-97900.825
4	Current taxonomy	-98270.815

4. Discussion

4.1 Phylogenetic inference and species-delimitation of Neurergus using RADseq data

Our first approach to infer phylogenetic relationships among species of *Neurergus* relied on a concatenation of all RAD loci, followed by the analysis of this supermatrix with classical tools such as RAxML. As expected, the resulting matrix was rather large and informative, and

its analysis yielded a fully resolved tree with 100% bootstrap support for every node (topology A; Fig. 2). Gene-jackknifed pseudo-replicates of the concatenated dataset also recovered full support for every node from ca. 100,000 bp sampled. We then used several species-tree reconstruction approaches as alternative means to assess whether this tree could be considered representative of the evolutionary history of these newt species. The two species-tree approaches based on quartet trees, tetrad and ASTRAL-II, resulted in the same topology as the concatenation analysis (Fig. 3a). Nodes were overall strongly supported, except for the one defining the *N. crocatus* + *N. derjugini* + *N. kaiseri* clade, which in the tetrad analysis received quite low support. However, when using the most complete dataset (without outgroup taxa, 14,765 SNPs), the support of this node increased substantially, thus suggesting that the lower support was due to the lack of informative loci and not to genuine variation among gene trees. On the contrary, two other coalescent-based methods for species-tree inference yielded a different, yet strongly supported topology (Fig. 3b). Indeed, in both SNAPP and BPP results, *N. derjugini* comes out as sister species to *N. crocatus*, and not to *N. kaiseri* as in the previous analyses.

Discrepancies between concatenation trees and species-trees are well documented (Kubatko and Degnan, 2007; Roch and Steel, 2015) and are usually attributed to four main processes generating variation among gene genealogies: (i) Incomplete Lineage Sorting (Pollard et al., 2006); (ii) gene flow between lineages (Solís-Lemus et al., 2016); (iii) gene duplication and loss; and (iv) recombination (Degnan and Rosenberg 2006). In the present case, both competing topologies had a rather low quartet score, highlighting a high level of discordance among gene trees. Further investigations revealed that topology A was supported by a proportion of ca. 17 % of the gene trees, while ca. 18 % supported topology B. Under the coalescent, short internal branch lengths in the species-tree can lead to the predominance

of gene tree topologies that differ from the species-tree history. This phenomenon results in high support for the wrong topology if the genes are concatenated, generating a conflict between species-tree and concatenation methods. Such cases of rapid diversification in internal branches of the species-tree history are defined to be in the gene tree anomaly zone (Kubatko and Degnan 2007; Degnan 2013). As the branch at the root of the *N. crocatus* + *N. kaiseri* + *N. derjugini* group is rather short, suggesting a rapid divergence, we could expect a higher proportion of gene trees supporting the wrong species-tree topology. Here, topology B was supported by more gene trees than topology A, although the difference was very low. The theoretical expectation in the anomaly zone would be to have a high support for topology B if the loci are concatenated, which is contrary to our results. Furthermore, the small difference in gene tree support for each competing topologies (17% versus 18%) could be a consequence of other evolutionary mechanisms, like gene flow among the lineages. Given the isolated geographic distribution of the sampled populations, current gene flow is highly improbable and can be excluded with some confidence, but migration between ancestral populations could have happened. In this case, the tree-like representation of the relationships within the group would be misleading, and the evolutionary history of the *Neurergus* would be best represented as a network with a deep reticulation (Degnan 2018; Burbrink and Gehara 2018). As the coalescent methods applied here only account for ILS as a source of variation among gene trees, it is impossible to rule out other sources of discordance with certainty.

On the other hand, the discrepancy between species-tree inference methods is more complicated and could be related to a combination of biological processes and differences in the mechanics of each method, or intrinsic characteristics of the different datasets used. The impact of the assembly and filtering parameters on downstream analyses has been investigated in several studies (e.g. Leaché et al. 2015, Lee et al. 2018). A low clustering

threshold, for example, can have a strong impact on phylogenetic analyses, by clustering paralogous or non-homologous loci together (Rubin et al. 2012). Similarly, the amount of missing data in RADseq datasets, either due to low individual coverage of the loci, or loci dropout when distantly related species are included (Viricel et al., 2014), can be problematic, for example for species-tree inference (Schmidt-Lebuhn et al., 2017). However, in our case, the topology inferred was very constant when varying the clustering threshold, the minimum individual coverage of the loci, and the presence or absence of outgroups (Tab. S1, Fig. S1). Thus, the processing of the RADseq data seems not to be the cause of our conflicting results.

With respect to the differences in each analytical method, it is important to note that the MCMC of the SNAPP analysis did not mix properly and had low ESS values. Also, the BPP analysis was performed with fractions of the dataset due to computational limitations. These two methods apply the full Bayesian function, modeling the entire evolutionary process, from sequences or SNPs to species-trees, taking into account phylogenetic uncertainty in the gene tree topologies and parameters of the multispecies coalescent model. In theory, they are more robust than methods that do not full Bayesian, like ASTRAL-II and tetrad. Both SNAPP and BPP inferred a different topology and are both consistent with each other. However, these results should be taken with caution because of poor chain mixing in SNAPP and the smaller amount of data used in BPP. Therefore, additional analyses will be needed to understand the evolutionary processes within *Neurergus* newts. Particularly, less computationally intensive methods that jointly estimate gene trees and species-trees, and model ILS and other processes such as gene flow, will be necessary to better understand the reasons for this phylogenetic conflict. Regarding the species-delimitation, BPP gave consistent results even with an alternative guide tree. Thus, even in the case of phylogenetic

446uncertainty, this method seems to be a valid alternative to BFD* for species-delimitation with
447RADseq data, even if the amount of data is still an important limitation.
448
4494.2 *Phylogeny and systematics of* Neurergus *newts*
450Given their limited number of species and the rather well-supported previous mitochondrial
451phylogenies, *Neurergus* newts would at first glance appear an easy target for phylogenomics.
452However, unravelling their phylogenetic history turned out to be less simple than initially
453thought. While the position of *N. strauchii* as sister-species to all other *Neurergus* is
454unambiguous, and congruent with previous studies, the relationships within the *N. crocatus* +
455*N. kaiseri* + *N. derjugini* clade are more difficult to resolve. For these taxa our analyses
456recovered two alternative scenarios: (A) *N. crocatus* as sister-species to *N. kaiseri* + *N.*
457*derjugini* or (B) *N. kaiseri* as sister-species to *N. crocatus* + *N. derjugini*. This second
458scenario is the most likely from a biogeographic point of view, as the ranges of *N. crocatus*
459and *N. derjugini* are quite close, while *N. kaiseri* occurs further south. Thus, after the initial
460split between *N. strauchii* and the *N. crocatus* + *N. kaiseri* + *N. derjugini* clade, divergence
461within this group could have occurred along the Zagros mountains, with *N. kaiseri* first
462becoming isolated in the southern part, and *N. crocatus* and *N. derjugini* being separated
463subsequently. Although it is uncertain to what extent the ecology of these newts reflects
464phylogenetic relationships, it is worth noting that *N. crocatus* and *N. derjugni* live in similar
465habitats (mountain streams), while *N. kaiseri* has adapted to more pond-dwelling habits,
466facing the more arid conditions at the southern distribution range of the whole genus
467(Steinfartz et al., 2002).
468 The RADseq data also revealed a high degree of intraspecific diversity in
469*Neurergus*. Except for *N. kaiseri*, our geographical sampling per species is limited, yet we

uncovered a pattern of strong molecular divergence among populations, not only for *N. kaiseri* but also for *N. crocatus*. In *N. crocatus*, the individual from the north (Semdinli) turned out to be clearly separated from the southern ones (Shiwalok). This structure in *N. strauchii* corresponds to the two described subspecies, *N. s. strauchi* and *N. s. barani*. Also for *N. derjugini*, two genetically divergent subspecies (*N. d. derjugini* and *N. d. microspilotus*) are known, and previous analysis of two nuclear gene fragments failed to place these two lineages into a monophyletic group (Hendrix et al., 2014). The more informative RADseq data now strongly support the clade containing these morphologically very similar and geographically close lineages and indicate only shallow genetic divergences between them. Moreover, in *N. kaiseri*, a substantial divergence was detected between populations from the northern and southern parts of its range, but this divergence is not currently reflected in the species' taxonomy.

Using coalescent-based species-delimitation models on the RADseq data, a classification with at least seven *Neurergus* species was strongly supported. This would suggest that the species richness of the genus has so far been severely underestimated based on morphology, mitochondrial markers and a restricted number of nuclear genes (Hendrix et al., 2014; Steinfartz et al., 2002). Taxonomically, this would imply that the sampled subspecies of both *N. strauchii* and *N. derjugini* should all be considered distinct species, as well as the two main clades of *N. kaiseri*. However, simulation studies have shown that the MSC model does not necessarily infer true species boundaries but might rather reflect population structure (Sukumaran and Knowles, 2017). Empirical studies using several species-delimitation approaches revealed that the number of species can be overestimated in the presence of gene flow and/or intra-specific structure (Chan et al., 2017; Miralles and Vences, 2013). This is particularly relevant for the split of *N. d. derjugini* and *N. d.*

494*microspilotus*. Indeed, even though they are separated as distinct species using MSC species-
495delimitation, these two subspecies only show a very shallow divergence on the concatenation
496tree. We hypothesize that this split might reflect intraspecific structure, resulting from a
497decrease in gene flow between the populations, rather than completed speciation. Recent
498range reduction, leading to strict allopatry, could be the process underlying the observed
499weak differentiation. This hypothesis agrees with the lack of support that the split of *N.*
500*derjugini* received in some of the BPP analyses, and with lack of substantial increase in the
501marginal likelihood of the models in the BFD* analysis.

502 On the contrary, the split between *N. s. strauchii* and *N. s. barani* is always
503strongly supported. These two taxa show a deep divergence estimated at 3 million years ago
504based on allozyme data (Steinfartz et al., 2002). A large-scale genetic and morphological
505study revealed that the two subspecies are separated by the Euphrate River (Pasmans et al.,
5062006), which could act as a long-term barrier to gene flow and therefore promote speciation.
507Major water courses as primary reason for isolation have been identified also for other
508amphibians (Zeisset and Beebee, 2008). The analysis of morphological data from the whole
509range of *N. strauchii* has also revealed morphological differences between the two
510subspecies, particularly in a significantly higher number of yellow spots in *N. s. strauchii*
511(Pasmans et al., 2006). Thus, considering the concordant genomic and morphological
512divergence, we suggest treating *N. s. strauchii* and *N. s. barani* as distinct species, elevating
513the latter to species status: *N. strauchii* (Steindachner 1887), and *N. barani* (Öz 1994). A full
514revision of the species complex, however, will necessitate inclusion of *N. s. munzurensis*, a
515recently described subspecies (Olgun et al., 2016) not present in our data set. As it is
516morphologically distinct from both *strauchii* and *barani*, and separated from them by large
517water courses, it might also be characterized by a relevant genetic divergence.

Finally, the status of *N. kaiseri* is more difficult to assess without additional analyses. MSC species-delimitation analyses strongly support that individuals from the North and the South of the distribution area are independent lineages. A recent study of the mitochondrial D-loop supports this result, as these two groups do not share any haplotype (Farasat et al., 2016). These results are in concordance with a recent morphological study that highlighted significant morphometric differentiation between the groups. However, even though the phylogenetic divergence between lineages seems to be high, their pairwise nucleotide diversity is low in comparison to that of *barani* and *strauchii*, for example. Thus, this seemingly high diversity in comparison to other *Neurergus* species could be an artefact due to the comparable larger sample size (7 individuals versus an average of 3.3 individuals included for the other species). However, in the BFD* analyses, the model splitting *N. kaiseri* alone had a higher marginal likelihood than the one splitting *N. strauchii*. Given our sampling of the former species, biased toward the northern clade (5 samples vs. 2 in the Southern clade), it is possible that we are underestimating their pairwise nucleotide diversity. Population-level analyses of a more extensive molecular dataset are needed to reveal whether there is still gene flow going on between the two lineages.

Overall, speciation in *Neurergus* appears to be driven mostly or exclusively by allopatric processes, with no sympatry known between species. In this respect, field exploration would be particularly important to detect possible geographically intermediate populations or even contact zones, e.g. of the two subspecies of *N. derjugini*. In the case of *N. kaiseri*, an in-depth look at the contact zone between the two lineages will permit identifying possible gene flow, or lack thereof, as well as ecological factors underlying their geographical separation. Finally, an extensive genomic sampling of all the *Neurergus* species would also be very interesting to know if they consistently display a genetic diversity

comparable to that of *N. kaiseri*, which could then be attributed to specific features of these newts such as possible low dispersal abilities. Otherwise, understanding the processes underlying the highest genetic richness of *N. kaiseri* could reveal very interesting elements of their evolutionary history, such as ecological shifts, or a particular biogeographic history.

Acknowledgments

We thank Jürgen Fleck, Willi and Christoph Schneider for providing tissue samples. Iker Irisarri provided scripts and suggestions for gene jackknifing. We thank A. Adam and H. Recknagel for support with ddRAD library preparation and Glasgow Polyomics for sequencing. Michael Hofreiter provided useful advice. LR and MV acknowledge support by the Deutsche Forschungsgemeinschaft (VE247/16-1 – HO 3492/6-1) in the framework of the "TaxonOmics" priority program.

Competing interest statement

The authors declare that there is no competing interest.

References

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat. Rev. Genet. 17, 81–92. https://doi.org/10.1038/nrg.2015.28

Bogaerts, S., Pasmans, F., Woeltjes, T., 2006. Ecology and conservation aspects of *Neurergus strauchii* (Amphibia: Salamandridae). Proc. 13th Congr. Soc. Eur. Herpetol. 15–18.

Bouckaert, R., Heled, J., Kühnert, 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. https://doi.org/10.1371/journal.pcbi.1003537

574Bragg, J.G., Potter, S., Bi, K., Moritz, C., 2016. Exon capture phylogenomics: efficacy across
575 scales of divergence. Mol. Ecol. Resour. 16, 1059–1068. https://doi.org/10.1111/1755-
576 0998.12449
577Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012.
578 Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a
579 full coalescent analysis. Mol. Biol. Evol. 29, 1917–1932.
580 https://doi.org/10.1093/molbev/mss086
581Burbrink, F. T., & Gehara, M. (2018). The biogeography of deep time phylogenetic
582 reticulation. *Systematic biology*, syy019. https://doi.org/10.1093/sysbio/syy019
583Cariou, M., Duret, L., Charlat, S., 2013. Is RAD-seq suitable for phylogenetic inference? An
584 in silico assessment and optimization. Ecol. Evol. 3, 846–852.
585 https://doi.org/10.1002/ece3.512
586Carstens, B.C., Knowles, L.L., 2007. Estimating species phylogeny from gene-tree
587 probabilities despite incomplete lineage sorting: an example from *Melanoplus*
588 grasshoppers. Syst. Biol. 56, 400–411. https://doi.org/10.1080/10635150701405560
589Chan, K.O., Alexander, A.M., Grismer, L.L., Su, Y.-C., Grismer, J.L., Quah, E.S.H., Brown,
590 R.M., 2017. Species delimitation with gene flow: a methodological comparison and
591 population genomics approach to elucidate cryptic species boundaries in Malaysian
592 Torrent Frogs. Mol. Ecol. 26, 5435–5450. https://doi.org/10.1111/mec.14296
593Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model.
594 Bioinformatics 30, 3317–3324. https://doi.org/10.1093/bioinformatics/btu530
595Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely
596 gene trees. PLoS Genet. 2, e68. https://doi.org/10.1371/journal.pgen.0020068
597Degnan, J. H. (2013). Anomalous unrooted gene trees. *Systematic biology*, *62*(4), 574-590.
598 https://doi.org/10.1093/sysbio/syt023
599Degnan, J. H. (2018). Modeling hybridization under the network multispecies coalescent.
600 *Systematic biology*, *67*(5), 786-799. https://doi.org/10.1093/sysbio/syy040
601Delić, T., Trontelj, P., Rendoš, M., Fišer, C., 2017. The importance of naming cryptic species
602 and the conservation of endemic subterranean amphipods. Sci. Rep. 7, 3391.
603 https://doi.org/10.1038/s41598-017-02938-z
604Delsuc, F., Tsagkogeorga, G., Lartillot, N., Philippe, H., 2008. Additional molecular support
605 for the new chordate phylogeny. genesis 46, 592–604. https://doi.org/10.1002/dvg.20450
606Dincă, V., Bálint, Z., Vodă, R., Dapporto, L., Hebert, P.D.N., Vila, R., 2018. Use of genetic,
607 climatic, and microbiological data to inform reintroduction of a regionally extinct
608 butterfly. Conserv. Biol. 0, 1–10. https://doi.org/10.1111/cobi.13111
609Eaton, D., 2015. Ipyrad: interactive assembly and analysis of RADseq data sets. [WWW
610 Document]. URL http://ipyrad.readthedocs.io/index.html (consulted 02.2018)
611Eaton, D., 2015b. Inferring species trees with tetrad [WWW Document]. URL
612 http://nbviewer.jupyter.org/github/dereneaton/ipyrad/blob/master/tests/cookbook-
613 quartet-species-tree.ipynb (consulted 03.2018)
614Eaton, D.A.R., 2014. PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses.
615 Bioinformatics 30, 1844–1849. https://doi.org/10.1093/bioinformatics/btu121

616Ebel, E.R., DaCosta, J.M., Sorenson, M.D., Hill, R.I., Briscoe, A.D., Willmott, K.R., Mullen,
617 S.P., 2015. Rapid diversification associated with ecological specialization in Neotropical
618 *Adelpha* butterflies. Mol. Ecol. 24, 2392–2405. https://doi.org/10.1111/mec.13168
619Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell,
620 S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high
621 diversity species. PLoS One 6, e19379. https://doi.org/10.1371/journal.pone.0019379
622Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn,
623 T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning
624 multiple evolutionary timescales. Syst. Biol. 61, 717–726.
625 https://doi.org/10.1093/sysbio/sys004
626Farasat, H., Akmali, V., Sharifi, M., 2016. Population genetic structure of the endangered
627 Kaiser's mountain newt, *Neurergus kaiseri* (Amphibia: Salamandridae). PLoS One 11,
628 1–16. https://doi.org/10.1371/journal.pone.0149596
629Harvey, M.G., Smith, B.T., Glenn, T.C., Faircloth, B.C., Brumfield, R.T., 2016. Sequence
630 capture versus Restriction Site Associated DNA Sequencing for shallow systematics.
631 Syst. Biol. 65, 910–924. https://doi.org/10.1093/sysbio/syw036
632Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data.
633 Mol. Biol. Evol. 27, 570–580. https://doi.org/10.1093/molbev/msp274
634Hendrix, R., Fleck, J., Schneider, W., Schneider, C., Geller, D., Avci, A., Olgun, K.,
635 Steinfartz, S., 2014. First comprehensive insights into nuclear and mitochondrial DNA
636 based population structure of Near East mountain brook newts (Salamandridae: Genus
637 *Neurergus*) suggest the resurrection of *Neurergus derjugini*. Amphib. Reptil. 35, 173–
638 187. https://doi.org/10.1163/15685381-00002939
639Huang, H., Knowles, L.L., 2014. Unforeseen consequences of excluding missing data from
640 next-generation sequences: simulation study of RAD sequences. Syst. Biol. 65, 357–
641 365. https://doi.org/10.1093/sysbio/syu046
642IUCN SSC Amphibian Specialist Group, 2016. *Neurergus kaiseri* [WWW Document]. IUCN
643 Red List Threat. Species 2016. https://doi.org/http://dx.doi.org/10.2305/IUCN.UK.2016-
644 3.RLTS.T59450A49436271.en
645Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J., McInerney, J.O., 2006.
646 Assessment of methods for amino acid matrix selection and their use on empirical data
647 shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. 6,
648 1–17. https://doi.org/10.1186/1471-2148-6-29
649Kieren, S., Sparreboom, M., Hochkirch, A., Veith, M., 2018. A biogeographic and ecological
650 perspective to the evolution of reproductive behaviour in the family Salamandridae.
651 Mol. Phylogenet. Evol. 121, 98–109. https://doi.org/10.1016/j.ympev.2018.01.006
652Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from
653 concatenated data under coalescence. Syst. Biol. 56, 17–24.
654 https://doi.org/10.1080/10635150601146041
655Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R., 2014. Species delimitation using
656 genome-wide SNP Data. Syst. Biol. 63, 534–542. https://doi.org/10.1093/sysbio/syu018

Leache, A. D., Chavez, A. S., Jones, L. N., Grummer, J. A., Gottscho, A. D., & Linkem, C. W. (2015). Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. Genome biology and evolution, 7(3), 706-719. https://doi.org/10.1093/gbe/evv026

Lecaudey, L.A., Schliewen, U.K., Osinov, A.G., Taylor, E.B., Bernatchez, L., Weiss, S.J., 2018. Inferring phylogenetic structure, hybridization and divergence times within Salmoninae (Teleostei: Salmonidae) using RAD-sequencing. Mol. Phylogenet. Evol. 124, 82–99. https://doi.org/10.1016/j.ympev.2018.02.022

Lee, K. M., Kivelä, S. M., Ivanov, V., Hausmann, A., Kaila, L., Wahlberg, N., & Mutanen, M. (2018). Information Dropout Patterns in Restriction Site Associated DNA Phylogenomics and a Comparison with Multilocus Sanger Data in a Species-Rich Moth Genus. Syst. Biol., 67(6) : 925-939. https://doi.org/10.1093/sysbio/syy029

Maddison, W.P., Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55, 21–30. https://doi.org/10.1080/10635150500354928

Manthey, J.D., Campillo, L.C., Burns, K.J., Moyle, R.G., 2016. Comparison of target-capture and restriction-site associated DNA sequencing for phylogenomics: a test in cardinalid Tanagers (Aves, Genus: *Piranga*). Syst. Biol. 65, 640–650. https://doi.org/10.1093/sysbio/syw005

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., Johnson, E.A., 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17, 240–248. https://doi.org/10.1101/gr.5681207

Miralles, A., Vences, M., 2013. New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in *Madascincus* Lizards. PLoS One 8. https://doi.org/10.1371/journal.pone.0068242

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30, i541–i548. https://doi.org/10.1093/bioinformatics/btu462

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. https://doi.org/10.1093/bioinformatics/btv234

Nieto-Montes de Oca, A., Barley, A.J., Meza-Lázaro, R.N., García-Vázquez, U.O., Zamora-Abrego, J.G., Thomson, R.C., Leaché, A.D., 2017. Phylogenomics and species delimitation in the knob-scaled lizards of the genus *Xenosaurus* (Squamata: Xenosauridae) using ddRADseq data reveal a substantial underestimation of diversity. Mol. Phylogenet. Evol. 106, 241–253. https://doi.org/10.1016/j.ympev.2016.09.001

Olgun, K., Avci, A., Bozkurt, E., Üzüm, N., Olgun, H., Ilgaz, Ç., 2016. A new subspecies of Anatolia newt, *Neurergus strauchii* (Steindachner, 1887) (Urodela : Salamandridae), from Tunceli, Eastern Turkey. Russ. J. Herpetol. 23, 271–277. https://doi.org/2304-0271

Ozsolak, F., Milos, P.M., 2011. RNA sequencing: advances, challenges and opportunities. Nat. Rev. Genet. 12, 87–98. https://doi.org/10.1038/nrg2934

Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S.C., Boisselier, M.C., Samadi, S., 2015. Use of RAD sequencing for delimiting species. Heredity (Edinb). 114, 450–459. https://doi.org/10.1038/hdy.2014.105

Papenfuss, T., Sparreboom, M., Tok, V., Ugurtas, I.H., Sevinç, M., Kuzmin, S., Anderson, S., Eken, G., Kiliç, T., Engin Gem, 2009. *Neurergus strauchii* [WWW Document]. IUCN Red List Threat. Species 2009. https://doi.org/http://dx.doi.org/10.2305/IUCN.UK.2009.RLTS.T14735A4458797.en

Papenfuss, T., Sparreboom, M., Ugurtas, I.H., Rastegar-Pouyani, N., Kuzmin, S., Anderson, S., Eken, G., Kiliç, T., Gem, E., Kaya, U., 2009. *Neurergus crocatus* (errata version published in 2016) [WWW Document]. IUCN Red List Threat. Species 2009. https://doi.org/http://dx.doi.org/10.2305/IUCN.UK.2009.RLTS.T14734A4458480.en

Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20, 289–290. https://doi.org/10.1093/bioinformatics/btg412

Pasmans, F., Bogaerts, S., Woeltjes, T., Carranza, S., 2006. Biogeography of *Neurergus strauchii barani* Öz, 1994 and *N. s. strauchii* (Steindachner, 1887) (Amphibia: Salamandridae) assessed using morphological and molecular data. Amphibia-Reptilia 27, 281–288. https://doi.org/10.1163/156853806777239878

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. https://doi.org/10.1371/journal.pone.0037135

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S.E., Lercher, M.J., 2014. PopGenome: an efficient swiss army knife for population genomic analyses in R. Mol. Biol. Evol. 31, 1929–1936. https://doi.org/10.1093/molbev/msu136

Pollard, D.A., Iyer, V.N., Moses, A.M., Eisen, M.B., 2006. Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. PLoS Genet. 2, e173. https://doi.org/10.1371/journal.pgen.0020173

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. https://doi.org/10.1093/sysbio/syy032

Razkin, O., Sonet, G., Breugelmans, K., Madeira, M.J., Gómez-Moliner, B.J., Backeljau, T., 2016. Species limits, interspecific hybridization and phylogeny in the cryptic land snail complex *Pyramidula*: the power of RADseq data. Mol. Phylogenet. Evol. 101, 267–278. https://doi.org/10.1016/j.ympev.2016.05.002

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. Resour. 15, 1316–1329. https://doi.org/10.1111/1755-0998.12406

Ree, R., Hipp, A., 2015. Inferring phylogenetic history from restriction site associated DNA (RADseq), in: Next-Generation Sequencing in Plant Systematics, Eds. E. Hörland and M. Appelhans, Koeltz Scientific Books, pp. 181–204.

737Roch, S., Steel, M., 2015. Likelihood-based tree reconstruction on a concatenation of aligned
738 sequence data sets can be statistically inconsistent. Theor. Popul. Biol. 100, 56–62.
739 https://doi.org/10.1016/j.tpb.2014.12.005
740Rodríguez, A., Burgon, J.D., Lyra, M., Irisarri, I., Baurain, D., Blaustein, L., Göçmen, B.,
741 Künzel, S., Mable, B.K., Nolte, A.W., Veith, M., Steinfartz, S., Elmer, K.R., Philippe,
742 H., Vences, M., 2017. Inferring the shallow phylogeny of true salamanders
743 (*Salamandra*) by multiple phylogenomic approaches. Mol. Phylogenet. Evol. 115, 16–
744 26. https://doi.org/10.1016/j.ympev.2017.07.009
745Rubin, B.E.R., Ree, R.H., Moreau, C.S., 2012. Inferring phylogenies from RAD sequence
746 data. PLoS One 7, e33394. https://doi.org/10.1371/journal.pone.0033394
747Schmidt-Lebuhn, A.N., Aitken, N.C., Chuah, A., 2017. Species trees from consensus single
748 nucleotide polymorphism (SNP) data: testing phylogenetic approaches with simulated
749 and empirical data. Mol. Phylogenet. Evol. 116, 192–201.
750 https://doi.org/10.1016/j.ympev.2017.07.018
751Schneider, C., Schneider, W., 2010. Fieldnotes on the ecology and distribution of *Neurergus*
752 *crocatus* Cope, 1862 and *Neurergus strauchii strauchii* (Steindachner, 1887) in Turkey
753 (Amphibia: Caudata: Salamandridae). Herpetozoa 23, 59–69.
754Sharifi, M., Shafiei Bafti, S., Papenfuss, T., Anderson, S., Kuzmin, S., Rastegar-Pouyani, N.,
755 2009. *Neurergus microspilotus* (errata version published in 2016) [WWW Document].
756 IUCN Red List Threat. Species 2009.
757 https://doi.org/10.2305/IUCN.UK.2009.RLTS.T59451A11944058.en
758Solís-Lemus, C., Yang, M., Ané, C., 2016. Inconsistency of species tree methods under gene f
759 low. Syst. Biol. 65, 843–851. https://doi.org/10.1093/sysbio/syw030
760Sparreboom, M., Steinfartz, S., Schultschik, G., 2000. Courtship behaviour of *Neurergus*
761 (Caudata: Salamandridae). Amphib. Reptil. 21, 1–
762 11.https://doi.org/10.1163/156853800507219
763Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
764 large phylogenies. Bioinformatics, 30(9), 1312-1313.
765 https://doi.org/10.1093/bioinformatics/btu033
766Steinfartz, S., Hwang, U.W., Tautz, D., Öz, M., Veith, M., 2002. Molecular phylogeny of the
767 salamandrid genus *Neurergus*: evidence for an intrageneric switch of reproductive
768 biology. Amphib. Reptil. 23, 419–431. https://doi.org/10.1163/15685380260462338
769Sukumaran, J., Knowles, L.L., 2017. Multispecies coalescent delimits structure, not species.
770 Proc. Natl. Acad. Sci. 114, 1607–1612. https://doi.org/10.1073/pnas.1607921114
771Veith, M., Bogaerts, S., Pasmans, F., Kieren, S., 2018. The changing views on the
772 evolutionary relationships of extant Salamandridae (Amphibia: Urodela). 1. Plos One
773 13, e0198237. https://doi.org/10.1371/journal.pone.0198237
774Viricel, A., Pante, E., Dabin, W., Simon-Bouhet, B., 2014. Applicability of RAD-tag
775 genotyping for interfamilial comparisons: empirical data from two cetaceans. Mol. Ecol.
776 Resour. 14, 597–605. https://doi.org/10.1111/1755-0998.12206
777Vogler, A.P., Desalle, R., 1994. Diagnosing units of conservation management. Conserv. Biol.
778 8, 354–363. https://doi.org/10.1046/j.1523-1739.1994.08020354.x

Yang, Z., 2015. The BPP program for species tree estimation and species delimitation. Curr. Zool. 61, 854–865. https://doi.org/10.1093/czoolo/61.5.854

Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. 107, 9264–9269. https://doi.org/10.1073/pnas.0913022107

Zeisset, I., Beebee, T.J.C., 2008. Amphibian phylogeography: a model for understanding historical aspects of species distributions. Heredity (Edinb). 101, 109–119. https://doi.org/10.1038/hdy.2008.30

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. https://doi.org/10.1016/j.ympev.2008.08.020

Figures captions :

Figure 1. Map of the sampling localities of *Neurergus* newts. Dashed lines show the whole range of the different taxa (distribution data from Steinfartz et al. 2002). The number of samples included in this study from each site is indicated in parentheses.

Figure 2. Maximum Likelihood tree of *Neurergus* newts inferred from the concatenation of 8,314 RAD loci (long branches have been artificially shortened). Stars at nodes indicate 100% bootstrap support and 100% jackknife support from ca. 100,000 positions sampled.

Figure 3. Species-tree of *Neurergus* newts inferred (a) using tetrad, with 8,314 SNPs, with bootstrap support from 500 replicates. When bootstrap supports differs on the tree without outgroup (14,765 SNPs), both values are indicated; (b) using SNAPP on the dataset without outgroup, with posterior probabilities of the nodes indicated.

Figure 4. Kernel density of the proportion of gene trees supporting each topology for 100 random draws of 1 individual per species, with dashed lines showing the average support.

The blue curve represent the Topology B, the green one the Toplogy A and the red one an alternative Topology with *N. kaiseri* and *N. crocatus* as sister species.

Figure 1

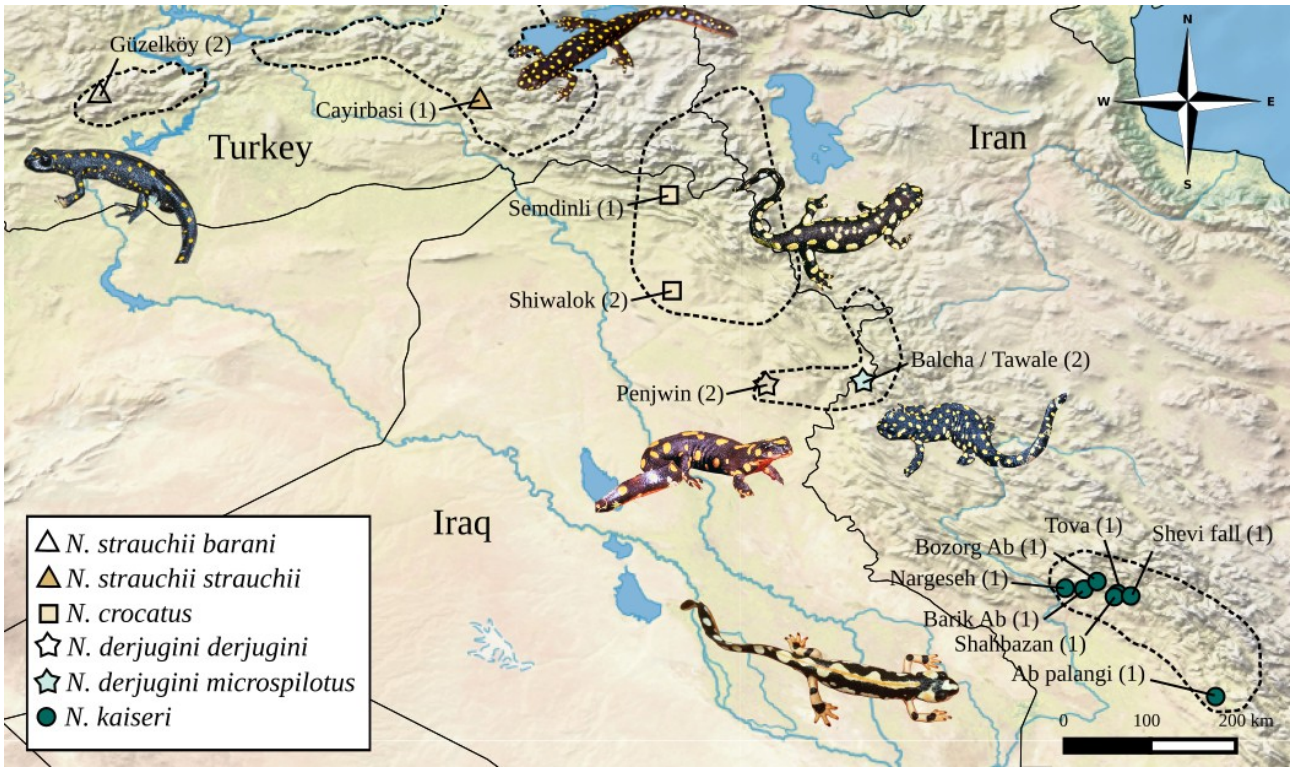

Figure 2

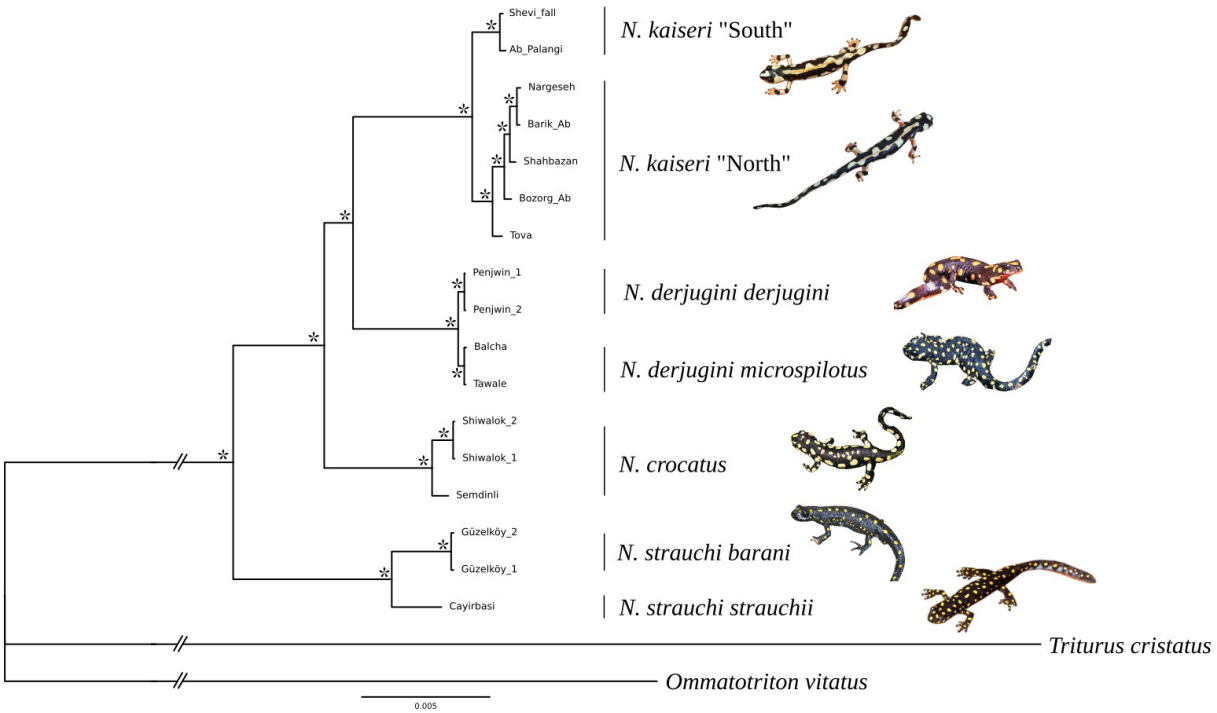

Figure 3

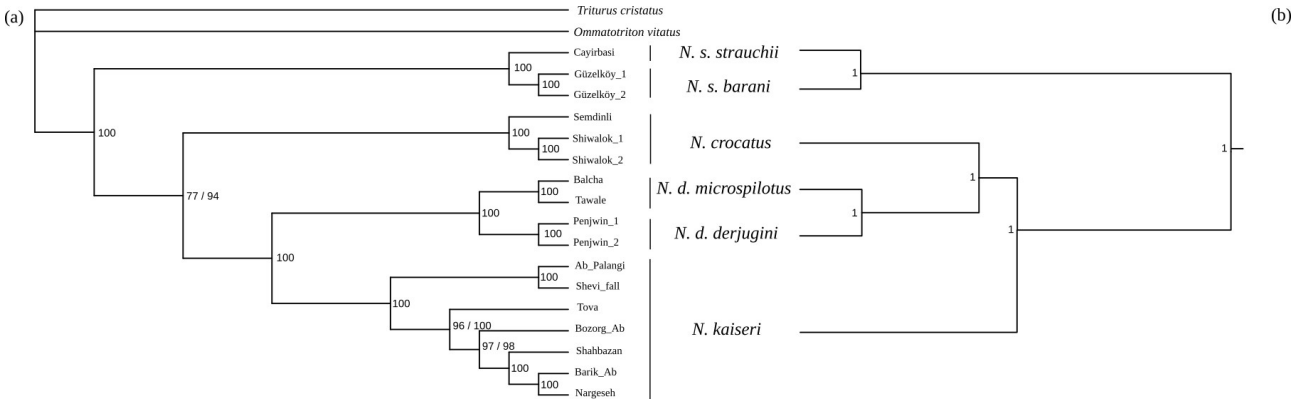

Figure 4

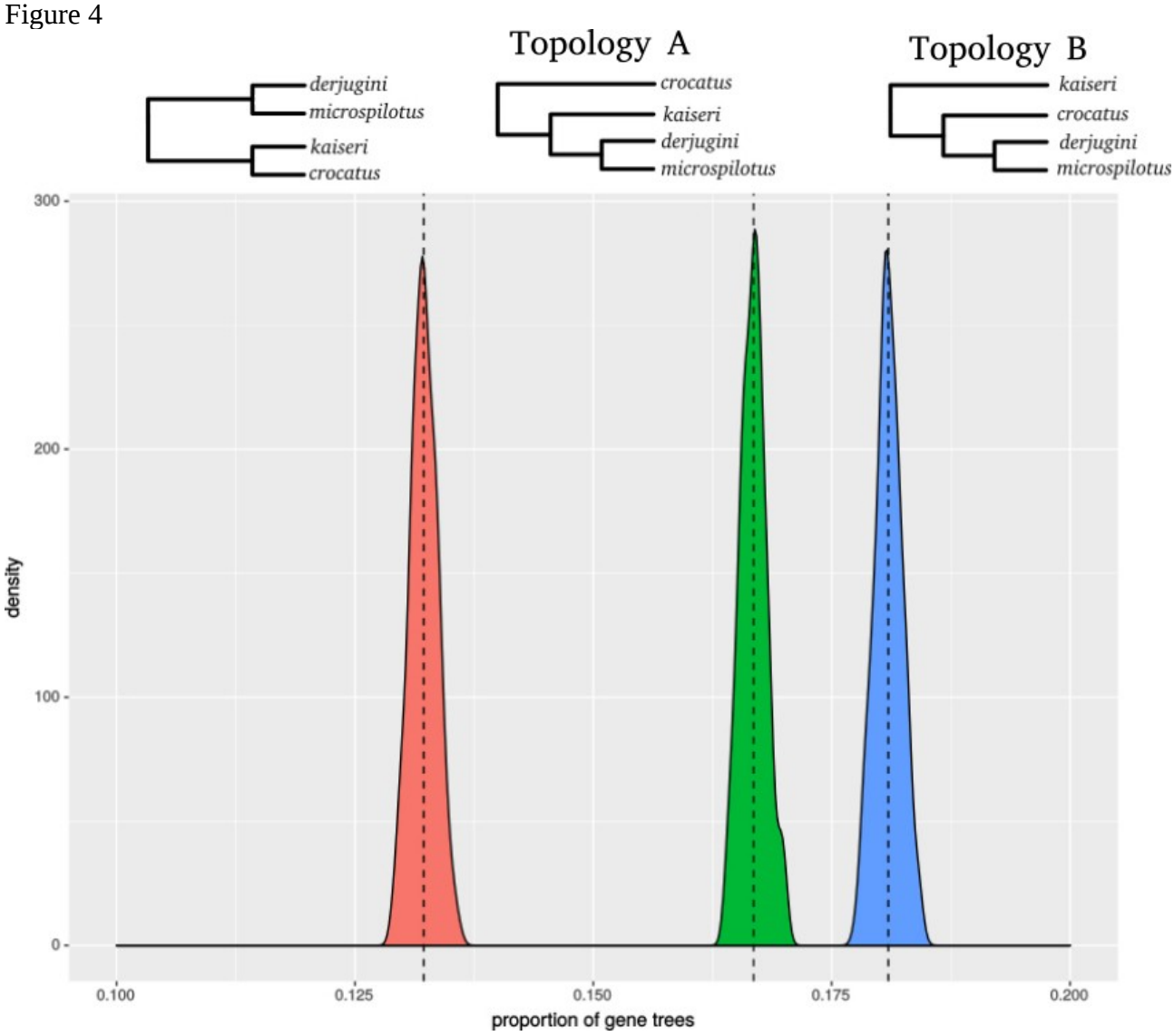

Table S1. Summary of the 66 dataset used to assess the effect of clustering threshold and minimum individual coverage on phylogenetic reconstruction. Concat. = concatenation of the loci; USNPs = concatenation of the unlinked SNPs (one per locus). The total length of the concatenation is given with the percentage of parsimony informative sites between parentheses, as well as the total number of unlinked SNPs. The datasets that yielded a topology that differs from the majority one (given in Figure 2) are in bold if the difference is in the deep nodes (higher than specific level) or underlined if the difference is in intra-specific nodes.

	Min. individual coverage	**Clustering threshold = 85%**		**Clustering threshold = 90%**		**Clustering threshold = 95%**	
		Concat.	**USNPs**	**Concat.**	**USNPs**	**Concat.**	**USNPs**
With outgroup (19 individuals)	4	9,300,135 bp (1.51%)	66,106 SNPs	9,820,748 bp (1.39%)	68,962 SNPs	10,947,732 bp (1.11%)	74,132 SNPs
	8	5,343,799 bp (2.10%)	43,072 SNPs	5,540,924 bp (1.97%)	44,421 SNPs	5,718,397 bp (1.60%)	44,844 SNPs
	10	4,214,578 bp (2.64%)	34,390 SNPs	4,335,433 bp (2.13%)	35,198 SNPs	4,271,838 bp (1.72%)	33,873 SNPs
	14	1,892,185 bp (2.61%)	15,830 SNPs	1,882,126 bp (2.46%)	15,682 SNPs	1,583,472 bp (1.96%)	12,909 SNPs
	16	991,030 bp (2.81%)	8,379 SNPs	955,241 bp (2.65%)	8,052 SNPs	702,497 bp (2.09%)	5,802 SNPs
	18	157,833 bp (2.91%)	**1,366 SNPs**	104,396 bp (2.52%)	**904 SNPs**	*20,951 bp (1.79%)*	<u>181 SNPs</u>
Without outgroup (17 individuals)	4	9,269,619 bp (1.49%)	64,449 SNPs	9,841,319 bp (1.39%)	67,912 SNPs	11,025,240 bp (1.10%)	74,201 SNPs
	8	5,337,230 bp (2.07%)	42,377 SNPs	5,564,653 bp (1.96%)	44,071 SNPs	5,759,113 bp (1.60%)	44,948 SNPs
	10	4,189,661 bp (2.83%)	33,703 SNPs	4,343,776 bp (2.12%)	34,863 SNPs	4,304,922 bp (1.72%)	33,979 SNPs
	14	1,812,605 bp (2.57%)	14,941 SNPs	1,835,854 bp (2.45%)	15,114 SNPs	1,582,491 bp (1.96%)	12,841 SNPs
	16	918,189 bp (2.75%)	7,651 SNPs	912,210 bp (2.65%)	7,603 SNPs	697,197 bp (2.09%)	5,729 SNPs

Figure S1. Bootstrap support for the trees inferred from the 66 alternative datasets. The average support is given for the deep nodes (relationships between species, triangles), and for the intra-specific nodes (squares) inferred from the concatenation of all the loci (blue) and only unlinked SNPs (red). The standard deviations are represented as error bars around the average values.

Legend :
Concatenation of all loci
Concatenation of unlinked SNPs
Deep nodes
Shallow nodes

With outgroups

Without outgroups

85%

90%

95%

Bootstrap support

Minimum individual coverage