



Universiteit
Leiden
The Netherlands

Data concatenation, Bayesian concordance and coalescent-based analyses of the species tree for the rapid radiation of Triturus newts

Wielstra, B.M.; Arntzen, J.W.; van der Gaag, K.J.; Pabijan, M.; Babik, W.

Citation

Wielstra, B. M., Arntzen, J. W., Van der Gaag, K. J., Pabijan, M., & Babik, W. (2014). Data concatenation, Bayesian concordance and coalescent-based analyses of the species tree for the rapid radiation of Triturus newts. *Plos One*, 9(10). doi:10.1371/journal.pone.0111011

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3281016>

Note: To cite this publication please use the final published version (if applicable).



Data Concatenation, Bayesian Concordance and Coalescent-Based Analyses of the Species Tree for the Rapid Radiation of *Triturus* Newts

Ben Wielstra^{1,2*}, Jan W. Arntzen¹, Kristiaan J. van der Gaag³, Maciej Pabijan^{4,5}, Wieslaw Babik⁴

1 Naturalis Biodiversity Center, Leiden, The Netherlands, **2** Department of Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom, **3** Leiden University Medical Center, Forensic Laboratory for DNA Research, Leiden, The Netherlands, **4** Institute of Environmental Sciences, Jagiellonian University, Kraków, Poland, **5** Institute of Zoology, Jagiellonian University, Kraków, Poland

Abstract

The phylogenetic relationships for rapid species radiations are difficult to disentangle. Here we study one such case, namely the genus *Triturus*, which is composed of the marbled and crested newts. We analyze data for 38 genetic markers, positioned in 3-prime untranslated regions of protein-coding genes, obtained with 454 sequencing. Our dataset includes twenty *Triturus* newts and represents all nine species. Bayesian analysis of population structure allocates all individuals to their respective species. The branching patterns obtained by data concatenation, Bayesian concordance analysis and coalescent-based estimations of the species tree differ from one another. The data concatenation based species tree shows high branch support but branching order is considerably affected by allele choice in the case of heterozygotes in the concatenation process. Bayesian concordance analysis expresses the conflict between individual gene trees for part of the *Triturus* species tree as low concordance factors. The coalescent-based species tree is relatively similar to a previously published species tree based upon morphology and full mtDNA and any conflicting internal branches are not highly supported. Our findings reflect high gene tree discordance due to incomplete lineage sorting (possibly aggravated by hybridization) in combination with low information content of the markers employed (as can be expected for relatively recent species radiations). This case study highlights the complexity of resolving rapid radiations and we acknowledge that to convincingly resolve the *Triturus* species tree even more genes will have to be consulted.

Citation: Wielstra B, Arntzen JW, van der Gaag KJ, Pabijan M, Babik W (2014) Data Concatenation, Bayesian Concordance and Coalescent-Based Analyses of the Species Tree for the Rapid Radiation of *Triturus* Newts. PLoS ONE 9(10): e111011. doi:10.1371/journal.pone.0111011

Editor: Helge Thorsten Lumbsch, Field Museum of Natural History, United States of America

Received: June 19, 2014; **Accepted:** September 22, 2014; **Published:** October 22, 2014

Copyright: © 2014 Wielstra et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. The raw 454 reads (in SFF format), the additional sequences obtained with Sanger sequencing (in ABI format), the SNP report produced with GS Amplicon Variant Analyzer 2.8, the alignments (raw, with indels removed, and collapsed into haplotypes), the input files for BAPS, MrBayes, BUCKY and *BEAST, and the individual gene trees resulting from MrBayes (created during the BUCKY exercise) and the species tree resulting from the eight *BEAST runs are available under Dryad Digital Repository entry doi:10.5061/dryad.mm81p.

Funding: BW is a Newton International Fellow. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: b.wielstra@sheffield.ac.uk

Introduction

The importance of molecular data in biological systematics can hardly be overstated, but potential pitfalls should be considered. In particular, a single gene tree does not necessarily reflect the phylogenetic relationships among species – hereafter referred to as the species tree – as phenomena like incomplete lineage sorting and introgression cloud the pattern of descent [1]. The mitochondrial genome is inherited as a single unit and gives rise to a single gene tree. On the other hand, the nuclear genome, due to its recombining nature, represents a collection of gene trees embedded in the species tree. To distill the species tree, a multitude of nuclear genes should be employed [2]. The progress in next-generation sequencing facilitates the production of large datasets and the use of multilocus datasets in systematics will soon become the norm [3,4].

The increase in molecular data is followed by advances in analytical methods. It is now realized that combining alignments

in a supermatrix that is treated as if it was a single ‘supergene’ can be misleading as this approach ignores the discordance among gene trees in phylogeny reconstruction. High confidence can be appointed to incorrectly inferred evolutionary relationships under data concatenation [3,5,6]. Furthermore, choice of allele in the concatenation process in the case of heterozygote marker-individual combinations can lead to widely differing topologies [7,8]. Gene tree discordance is explicitly incorporated in Bayesian concordance analysis, in which individual gene trees are summarized to provide a ‘concordance factor’ per clade, representing the proportion of gene trees in which clades are present. [9,10]. Recent coalescent-based methods of species tree estimation take advantage of the information contained in a sample of gene trees more directly by conjointly estimating the species tree and individual gene trees, while explicitly taking into account incomplete lineage sorting as a source of gene tree discordance [2,11]. Despite these analytical advances, rapid radiations with temporally closely spaced branching events tend to show much

incongruence among gene trees and still pose a challenge, even for coalescent-based methods [12,13]. Reconstructing relationships under such scenarios may be further complicated by ongoing gene flow, especially when it occurs between non-sister taxa [14,15]. Finally, the intense computational demands for coalescent-based analysis of large data set pose a limiting factor [7].

The genus *Triturus* (marbled and crested newts, Amphibia: Salamandridae) provides an example of a rapid radiation and the branching order has proved difficult to resolve [16]. The nine species can be divided into five groups hereafter referred to as morphotypes (Fig. 1; [17]). Morphotypes differ in body built and body built is reflected by the number of rib-bearing pre-sacral vertebrae. Arranged from stocky to slender, with the modal rib-bearing pre-sacral vertebrae count provided between parentheses, these morphotypes are: *T. marmoratus* and *T. pygmaeus* (12), *T. ivanbureschi* (including an as yet unnamed taxon) and *T. karelinii* (13), *T. carnifex* and *T. macedonicus* (14), *T. cristatus* (15) and *T. dobrogicus* (16 or 17) [18]. A full mtDNA dataset yielded a fully bifurcating and highly supported tree in which morphotypes are monophyletic (Fig. 1a; [19]). This tree supports the most parsimonious scenario of character state change in the number of rib-bearing pre-sacral vertebrae (the ancestral state for the number of rib-bearing pre-sacral vertebrae in the family Salamandridae is considered to be 13 [19]). Coalescent-based species tree estimations based on three nuclear markers for all species (Fig. 1b; [20]) and five nuclear markers for a subset of species [21] yielded contrasting results and both deviated from the species tree based on full mtDNA and morphology.

The mitochondrial genome has two characteristics that are particularly important in the setting of a rapid radiation and that make it less susceptible to yield error than a nuclear gene tree. Firstly, lineage sorting occurs faster compared to the nuclear genome due to its fourfold smaller effective population size [1,22]. Secondly, the number of substitutions accumulated along short internal branches is generally elevated for mtDNA as it is characterized by a comparatively high mutation rate [12,23]. These advantages, plus the congruency of the full mtDNA phylogeny and morphology, led us to accept the full mtDNA phylogeny over the nuclear phylogenies and we concluded that a wider sampling of the genome is required to test the *Triturus* species tree [19,20]. We here collect sequences for 38 genetic markers (constituting 13,517 bp in total) for twenty *Triturus* newts. We employ both data concatenation and coalescent-based estimations of the *Triturus* species tree and discuss the discordant outcomes of the analyses in the light of previous attempts to resolve the *Triturus* species tree.

Material and Methods

Sampling

We chose samples from a comprehensive DNA database stored in the collection of Naturalis Biodiversity Center available from an earlier study [24]. This DNA was initially extracted using the DNeasy Tissue Kit (Qiagen) from tail tips taken from animals under anesthesia that were subsequently released back into the wild (a method which does not negatively affect survival [25]). We sampled 2–3 individuals for each of the seven crested newt species and one individual for each of the two marbled newt species. It should be noted that one taxon that we refer to as ‘*Triturus* candidate species’ is yet to be named [26]. We employed a wide geographical spread within species, but our sampling scheme avoided localities positioned close to interspecific hybrid zones, to minimize effects of introgression (Fig. 2; Table 1) [27].

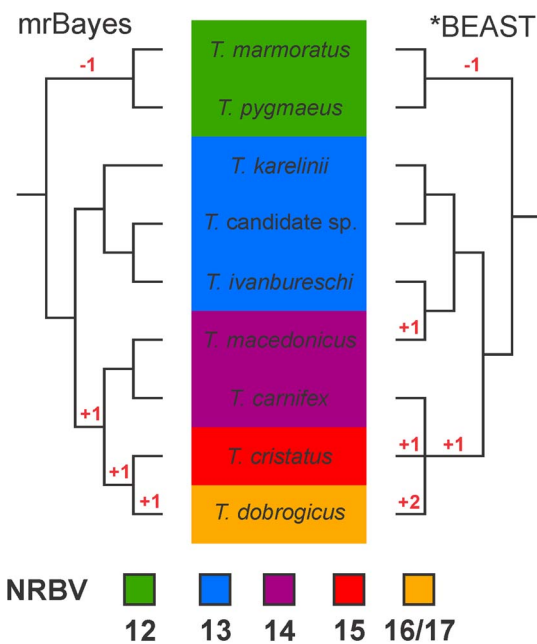


Figure 1. Previously argued phylogenetic hypotheses for the genus *Triturus*. Background colors reflect the variation in the number of rib-bearing pre-sacral vertebrae (NRBV) characterizing the five *Triturus* morphotypes. The two species with a green background are marbled newts and the remaining species are crested newts. The MrBayes phylogeny (left) is based on full mitochondrial genomes [19]. Posterior probabilities for all internal branches are 0.95 or higher. Note that this phylogeny is concordant with the most parsimonious interpretation of the evolution of the number of rib-bearing pre-sacral vertebrae (with the required character state changes noted in red). The *BEAST coalescent-based estimation of the species tree (right) is based on three nuclear introns (adapted from [20]). Posterior probabilities for internal branches are all below 0.95 and only those supported at over 0.75 are shown.

doi:10.1371/journal.pone.0111011.g001

Sequencing and data preparation

We adapted a protocol initially designed for Ion Torrent next-generation sequencing [28] to be used for the 454 platform. The rationale for choosing this method was that relatively long DNA fragments could be sequenced. We here summarize the protocol followed and refer to ref. [28] for details. Ninety-six transcriptome-based gene models for *Triturus* (compiled following ref. [29]) with long 3' untranslated regions were identified. To increase the chance that these gene models represent single copy genes, only those that produced unambiguous hits to a single gene when BLASTed against human and *Xenopus* transcripts were selected [30]. For the identified 3' untranslated regions we designed 96 primer pairs that amplify 400 ± 10 bp (including primers) with BatchPrimer3.1.0 [31] (Dataset S1).

Primer pairs were tested for cross-species-amplification in five *Triturus* individuals spanning the taxonomic width of the genus (Table 1). PCRs were conducted in a total volume of 25 μ l, containing 12.5 μ l QIAGEN Taq PCR Master Mix, 2.5 μ l primer mix (5 μ M), 9 μ l Milli-Q water and 1 μ l of template DNA and our PCR program consisted of denaturation at 95°C for 30 s, 35 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s and extension at 72°C for 45 s, and a final extension at 72°C for 5 min. We checked whether a single product of expected length was amplified for each species by gel electrophoresis using the E-Gel Precast Agarose Gel system (Life Technologies).

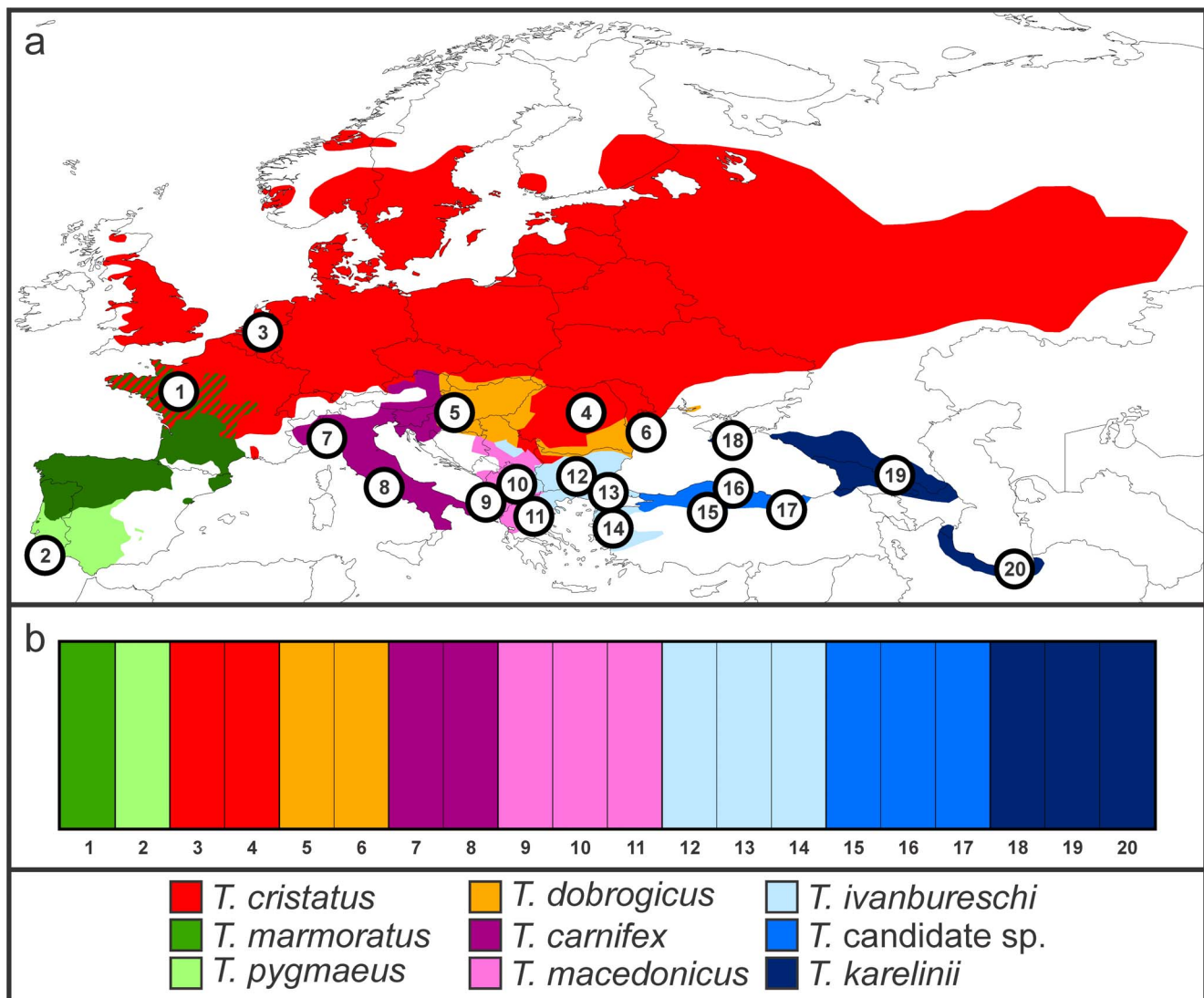


Figure 2. The distribution of the two marbled and seven crested newt species, represented by different colors, and the geographical position of sampled individuals (a). Sampling details can be found in Table 1. BAPS plot showing that each individual is allocated to its respective species (b).

doi:10.1371/journal.pone.0111011.g002

Next we amplified the successful markers for all 20 individuals using the same PCR protocol as above in singleplex PCRs. We determined the concentration of each PCR product with the QIAxcel system and pooled all markers per individual equimolar to yield 20 pools. A unique adaptor was ligated to each pool; we used Roche rapid library adaptors RL1–20 (Roche technical bulletin 2010–010). Sequencing was conducted using a quarter of a 454 next-generation sequencing run. Reads were aligned against the transcriptome-based reference and a SNP report was generated in GS Amplicon Variant Analyzer 2.8 (Roche), where SNPs were called if present in at least 20% of reads for marker-individual combinations. We confirmed SNPs and determined how multiple SNPs were linked in individuals containing two alleles for a marker that differed in more than one position by checking the alignments by eye in GS Amplicon Variant Analyzer 2.8.

Sequences were manually aligned in MacClade 4.08 [32]. As missing data can cause spurious results [33], we discarded indels and we filled in missing individual-marker combinations through

Sanger sequencing. Models of sequence evolution were determined for each marker with MrModeltest 2 <<http://www.abc.se/~nylander/>> using the Akaike Information Criterion (Dataset S2) and the number of polymorphic and parsimony informative sites among the haplotypes for each marker was determined with DnaSP v5 [34].

Data analysis

We conducted a Bayesian analysis of population structure with BAPS v.5.3 [35]. BAPS assigns individuals to distinct gene pools probabilistically, based on multilocus genetic data, where each individual allele is coded as an integer (two gene copies per marker, which may or may not represent the same allelic variant). We let BAPS determine the most probable number of distinct gene pools (k) in the data, evaluating k over a $1 \leq k \leq 20$ range, using ten replicates, and tested for admixture among gene pools. We used DISTRUCT [36] to visualize the BAPS results.

Table 1. Details on Triturus sampling.

| Population number | Taxon | Sample ID | Locality | N | E |
|----------------------|-----------------------------------|-----------|------------------------------|--------|--------|
| <i>Marbled newts</i> | | | | | |
| 1 | <i>Triturus marmoratus</i> | 5018* | France: Mayenne | 48.252 | -0.469 |
| 2 | <i>Triturus pygmaeus</i> | 5016 | Portugal: Serra de Monchique | 37.335 | -8.506 |
| <i>Crested newts</i> | | | | | |
| 3 | <i>Triturus cristatus</i> | 5021* | Netherlands: Heeze | 51.393 | 5.535 |
| 4 | <i>Triturus cristatus</i> | 4322 | Romania: Sânzieni | 46.104 | 26.128 |
| 5 | <i>Triturus dobrogicus</i> | 5065 | Hungary: Dráva | 46.23 | 17.06 |
| 6 | <i>Triturus dobrogicus</i> | 4801* | Romania: Razboinita | 45.222 | 29.165 |
| 7 | <i>Triturus carnifex</i> | 5047 | Italy: Donega | 44.933 | 9.233 |
| 8 | <i>Triturus carnifex</i> | 5046 | Italy: Doganella | 41.750 | 12.783 |
| 9 | <i>Triturus macedonicus</i> | 3494 | Albania: Bejar | 40.429 | 19.850 |
| 10 | <i>Triturus macedonicus</i> | 3424* | Macedonia: Leskoec | 40.962 | 20.885 |
| 11 | <i>Triturus macedonicus</i> | 3775 | Greece: Kerameia | 39.562 | 22.081 |
| 12 | <i>Triturus ivanbureschi</i> | 4732 | Bulgaria: Ostar Kamak | 41.878 | 25.853 |
| 13 | <i>Triturus ivanbureschi</i> | 2360* | Turkey: Keşan | 40.917 | 26.633 |
| 14 | <i>Triturus ivanbureschi</i> | 1820 | Turkey: Çan | 40.006 | 26.937 |
| 15 | <i>Triturus</i> candidate species | 1901 | Turkey: Kalecik | 40.077 | 33.341 |
| 16 | <i>Triturus</i> candidate species | 1948 | Turkey: Cebeci | 41.201 | 34.036 |
| 17 | <i>Triturus</i> candidate species | 2051 | Turkey: Şebinkarahisar | 40.286 | 38.126 |
| 18 | <i>Triturus karelinii</i> | 2109 | Ukraine: Nikita | 44.538 | 34.243 |
| 19 | <i>Triturus karelinii</i> | 2226 | Georgia: Telavi | 41.903 | 45.475 |
| 20 | <i>Triturus karelinii</i> | 2390 | Iran: Qu'Am Shahr | 36.436 | 52.803 |

Population numbers correspond to Fig. 1. Individuals are identified with a code that refers to tail tips/DNA extractions stored at Naturalis Biodiversity Center. Individuals marked with an asterisk were used in the primer testing.

doi:10.1371/journal.pone.0111011.t001

We carried out a concatenated phylogenetic analysis with MrBayes 3.2.2 [37] via the CIPRES Science Gateway [38]. To deal with heterozygosity, we created two input files (hereafter input file a and b) and for heterozygote marker-individual combinations we included each allele in one or the other input file, ensuring that each allele was used at least once. Data partitions were unlinked. We conducted two independent four-chain one billion-generation runs per input file, sampled every 100,000 generations, using a heating parameter of 0.2. Tracer v1.6 <<http://tree.bio.ed.ac.uk/software/tracer/>> was used to check stabilization of overall likelihood within and convergence between runs for each input file. We discarded the first half of generations as burn-in.

We conducted a Bayesian Concordance Analysis using BUCKY 1.4.3 [39]. We created individual gene trees in MrBayes using a single four-chain 1,000,000 generations run, sampled every 1,000 generations. The tree files resulting from MrBayes were summarized using mbsum (distributed as part of the BUCKY package) and the first half of the trees in each tree posterior was discarded as burn-in. The output of mbsum was further processed in BUCKY to create a primary concordance tree with concordance factors for clades. We implemented one cold and three heated chains, used an alpha multiplier of 5 and increased the rate at which chains in the Metropolis-coupled MCMC swap states (every 50 updates) to improve mixing. The *a priori* level of gene tree discordance (α) was set at 1 and 20 with no obvious influence on results; we report concordance factors for the default value of 1. As BUCKY requires a single sequence to represent each marker-individual combination, we followed a similar approach as above to deal with heterozygosity and conducted two replicates, which included

either one or the other allele in the case of heterozygote marker-individual combinations.

A coalescent-based estimation of the species tree was performed using *BEAST [40] as implemented in BEAST 1.7 [41]. We applied the lognormal relaxed clock model, the piecewise linear with constant root population function and the Yule speciation model. We used the nine *Triturus* species as operational taxonomical units. We conducted eight independent two billion-generation runs, sampled every 100,000 generations. Tracer v1.6 was used to check for stabilization of overall likelihood. For further analysis we used the three runs that reached the optimum likelihood within the burn-in of one billion generations. These runs were combined and summarized with the LogCombiner 1.7 and TreeAnnotator 1.7 programs distributed with the *BEAST package.

Data accessibility

The raw 454 reads (in SFF format), the additional sequences obtained with Sanger sequencing (in ABI format), the SNP report produced with GS Amplicon Variant Analyzer 2.8, the alignments (raw, with indels removed, and collapsed into haplotypes), the input files for BAPS, MrBayes, BUCKY and *BEAST, the individual gene trees resulting from MrBayes (created during the BUCKY exercise) and the species tree resulting from the eight *BEAST runs are available under Dryad Digital Repository entry doi:10.5061/dryad.mm81p

Results

Out of the 96 markers tested, 43 produced a single clear band of expected size on agarose gel for all five tested crested newt species (Dataset S3) and these 43 markers were amplified for all 20 *Triturus* individuals. The 454 run produced 180,120 reads of which 123,674 could be mapped to references. On average we obtained 143.8 ± 6.5 (standard error) per marker per individual. Five markers were excluded because they were suspected to represent a multicopy gene (*ace*), had homopolymer tracts that hampered sequencing (*ddx17*, *trab*), or individuals had missing data which we did not manage to complement with Sanger sequencing (*dgcr2*, *rbm15*). For the 38 remaining markers, the 19 individual-marker combinations that were missing were sequenced with Sanger sequencing and added to the data set.

Across all *Triturus* species and all 38 markers, the average number of polymorphic sites and the number of parsimony informative sites were 18.0 ± 7.3 (standard error) and 7.6 ± 4.9 and for the crested newts only they were 12.1 ± 5.6 and 4.1 ± 3.7 (Dataset S2). The average number of alleles per marker within individuals ranges from 1.0 to 1.6, with *T. dobrogicus* showing the highest heterozygosity of all species. The genotypes of the twenty *Triturus* individuals can be found in Dataset S4.

BAPS partitioned the twenty *Triturus* individuals into nine groups and each individual clustered with conspecifics with full support ($p = 1.0$) in the admixture analysis (Fig. 2). We should note that when we forced BAPS to partition the twenty individuals into two groups (given the rationale that there is a basic split in *Triturus* between marbled and crested newts) or five groups (given the rationale that there are five *Triturus* morphotypes) results slightly deviated from expectation, suggesting limited ability of the program to be used in such a hierarchical manner. Under $k = 2$ the morphotype comprising *T. karelinii*, *T. ivanbureschi* and the *Triturus* candidate species (forthwith the *T. karelinii* morphotype) was distinguished from the other *Triturus* species (including both crested and marbled newt species). Under $k = 5$ the *T. cristatus* and *T. dobrogicus* morphotypes were placed in a single group whereas *T. ivanbureschi* and the *Triturus* candidate species were placed in a separate group from *T. karelinii*.

All species were recovered as monophyletic in the MrBayes concatenated analysis (posterior probability, $pp = 1.0$; Fig. 3). The topologies resulting from the BUCKy Bayesian concordance analysis recovered all species as monophyletic with relatively high support for the recognized *Triturus* species (concordance factors, $CFs \geq 0.24$) compared to the *Triturus* candidate species ($CF = 0.07$; Fig. 4).

The result of the concatenated analyses (using two different input files together representing all alleles present) with MrBayes (hereafter MrBayes phylogeny) is presented in Fig. 3. The basal dichotomy separates the marbled and crested newt groups with high support ($pp = 1.0$ for both groups). Additionally, high support for the monophyly of the *T. karelinii* morphotype is found ($pp = 1.0$) and within the *T. karelinii* morphotype the sister-group relationship of *T. ivanbureschi* and the *Triturus* candidate species has high support ($pp = 1.0$). Both replicates suggest monophyly of *T. dobrogicus* and *T. carnifex*, albeit support values slightly differ ($pp = 1.0$ and $pp = 0.94$). The phylogenetic position of *T. cristatus* and *T. macedonicus* varies between both replicates, although the support values for their placement in both replicates are relatively high ($pp = 0.89$ – 0.93).

The result of the Bayesian concordance analysis with BUCKy (hereafter BUCKy phylogeny) is presented in Fig. 4. The topologies of the two replicates differ slightly, but concordance factors for the affected clades are low (< 0.1). Support is higher for

monophyly of the marbled newts ($CF = 0.92$ or 0.93 in the two replicates) than for monophyly of the crested newts ($CF = 0.46$ in both replicates) and support for monophyly of the *T. karelinii* morphotype is lower still ($CF = 0.18$ or 0.19). Support for relationships among the crested newt morphotypes is low ($CFs < 0.1$). Consensus trees of the posterior tree distributions produced by MrBayes for the individual gene trees are available from Dryad.

The result of the coalescent-based analysis with *BEAST (hereafter *BEAST phylogeny) is presented in Fig. 5. This tree is based on three runs; it excludes five runs that plateaued at a lower likelihood value within the burn-in of one billion generations. The individual trees resulting from all eight runs are available on Dryad. Part of the ESS values in the individual analyses, including those runs used to create the consensus tree, were (well) below 200. This finding shows that, even with runs of two billion-generations, stabilization within and convergence between runs did not occur. Considering we conducted eight independent, long runs, this appears inherent to the nature of our dataset. Still the results provide insight into the *Triturus* species tree.

*BEAST consistently supports the monophyly of the marbled newt group, the crested newt group and the *T. karelinii* morphotype ($pp = 1.0$ in the consensus tree and the individual trees). Relationships among the three members of the *T. karelinii* morphotype are ambiguous, reflected by low support values and various relationships suggested in the individual trees. For the remaining crested newts, support for the sister-group relationship of *T. carnifex* and *T. macedonicus* is low, but it is consistently found, i.e. in the consensus tree and all individual trees. Similarly, all runs find a sister relationship of the *T. carnifex* – *T. macedonicus* morphotype with *T. cristatus*, albeit again with low support. The phylogenetic placement of *T. dobrogicus* varies between runs, being recovered as either sister to the all crested newts or to the crested newts excluding the *T. karelinii* morphotype (always with low posterior probabilities).

Discussion

Distinguishing *Triturus* species with nuclear DNA

With the suite of nuclear DNA markers presented here the various *Triturus* species can successfully be identified. BAPS partitions the dataset into nine gene pools, corresponding to the nine *Triturus* species. Similarly the species were recovered as monophyletic in the data concatenation approach with MrBayes and the Bayesian concordance analysis with BUCKy. This gives us the confidence that possible adverse effects of *a priori* appointing individuals to ‘species’ (e.g. if individuals of more than one species are appointed to a single ‘species’ or individuals of the same species are divided into more than one ‘species’), such as required in the coalescent-based species tree analysis [13], are negligible.

The different *Triturus* species trees compared

We have a previous estimate of the *Triturus* species tree based on full mtDNA that is in line with the most parsimonious interpretation of evolution of the axial skeleton as reflected by the number of rib-bearing pre-sacral vertebrae, namely four character state changes (Fig. 1a [19]). The topology of the second BUCKy (i.e. the Bayesian concordance analysis) replicate is identical to the full mtDNA phylogeny (Fig. 1 and 4 “Bucky b”). For the other topologies, two branches have to be moved to reconcile them with the topology with full mtDNA and morphology as follows. In the two replicates of the MrBayes phylogeny (i.e. the concatenated analysis; Fig. 3), *T. carnifex* should be moved to form a sister group with *T. macedonicus* and *T. cristatus* should be moved to form a sister group with *T. dobrogicus* (notice that the placement

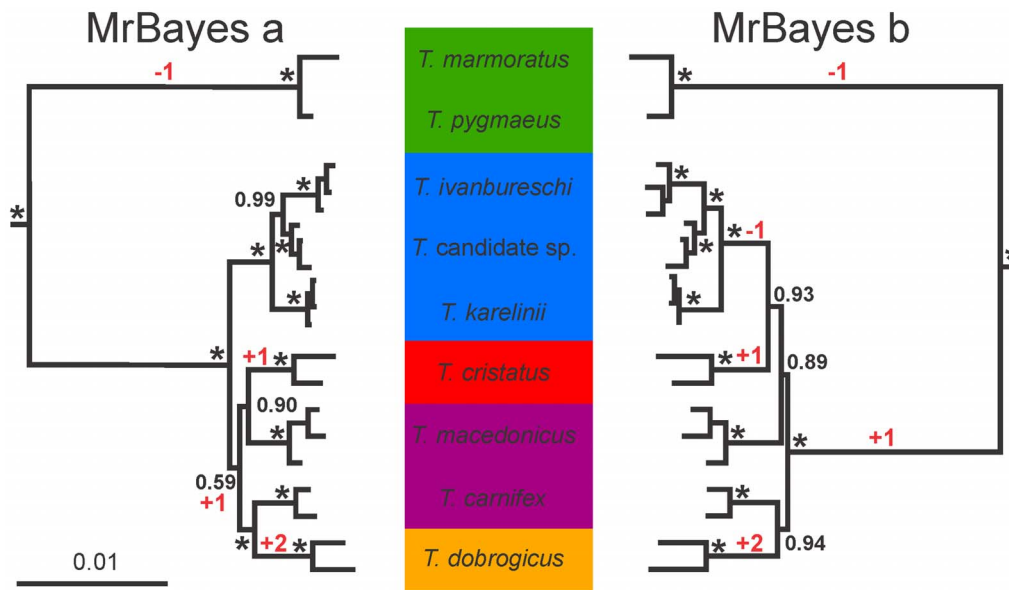


Figure 3. Species trees resulting from the concatenation-based estimation in MrBayes for the genus *Triturus*, based on 38 nuclear markers positioned in 3' untranslated regions, sequenced for 20 individuals. Internal branches supported with pp = 1.0 are marked with an asterisk. Analyses a and b reflect two different input files used which, for heterozygote marker-individual combinations, include either one or the other allele. The inferred position of the root for *Triturus* is shown. Background colors reflect variation in the number of rib-bearing pre-sacral vertebrae characterizing *Triturus* morphotypes as in Fig. 1 and the character stage changes required are noted in red. doi:10.1371/journal.pone.0111011.g003

of *T. cristatus* differs between replicates). In the first replicate of the BUCKy phylogeny (“Bucky a” in Fig. 4) the position of *T. carnifex* and *T. cristatus* need to be switched. In the *BEAST phylogeny (i.e. the coalescent-based analysis) *T. cristatus* should be

moved to form a sister group with *T. dobrogicus* (Fig. 5). One of the MrBayes phylogenies (“MrBayes b” in Fig. 3) implies six character state changes in the number of rib-bearing pre-sacral vertebrae, whereas the other estimates of the *Triturus* species tree

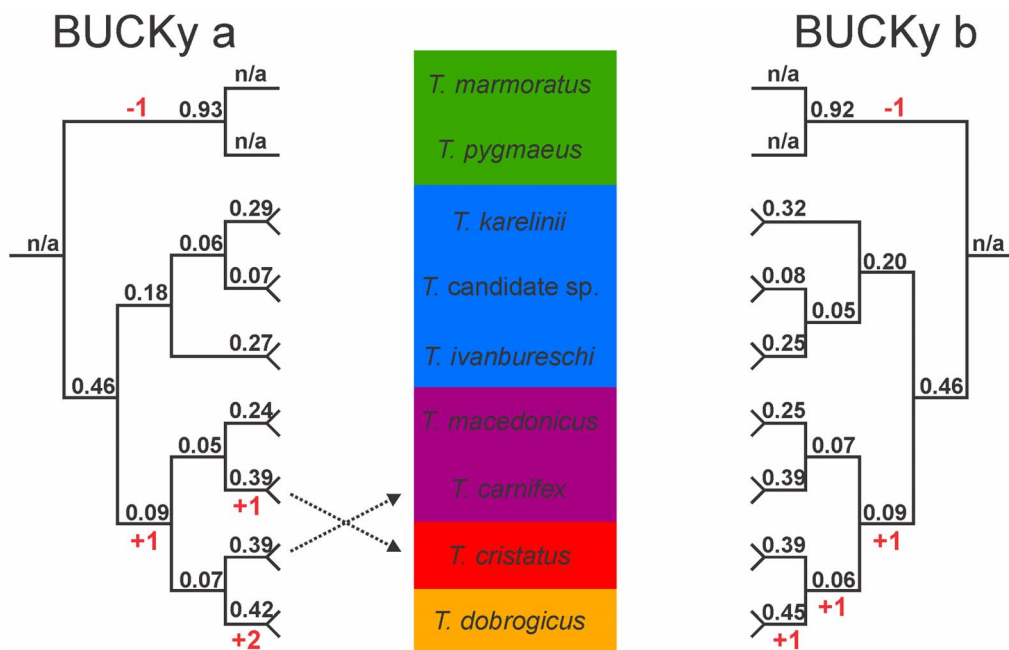


Figure 4. Primary concordance trees resulting from the Bayesian concordance analysis in BUCKy for the genus *Triturus*, based on 38 nuclear markers positioned in 3' untranslated regions, sequenced for 20 individuals. Support values represent concordance factors, i.e. the proportion of gene trees in which clades are present; n/a is not applicable. Relationships within species are not shown. Analyses a and b reflect two different input files used which, for heterozygote marker-individual combinations, include either one or the other allele. The inferred position of the root for *Triturus* is shown. Background colors reflect variation in the number of rib-bearing pre-sacral vertebrae characterizing *Triturus* morphotypes as in Fig. 1 and the character stage changes required are noted in red. doi:10.1371/journal.pone.0111011.g004

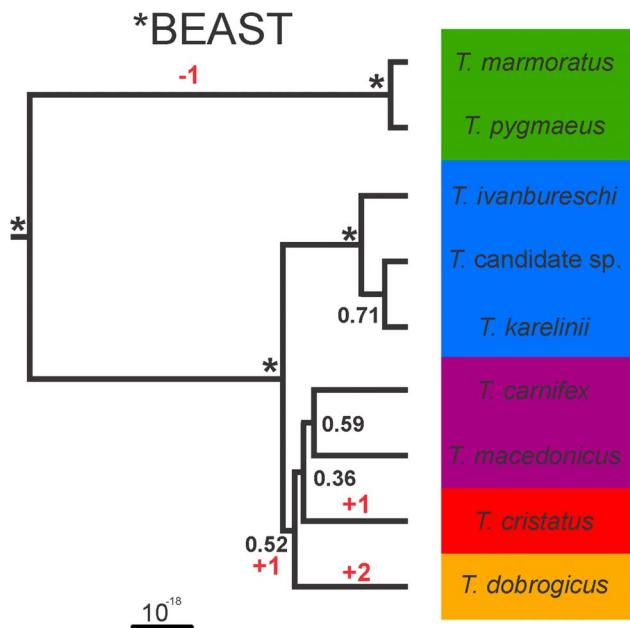


Figure 5. Species trees resulting from the coalescent-based estimation in *BEAST for the genus *Triturus*, based on 38 nuclear markers positioned in 3' untranslated regions, sequenced for 20 individuals. Internal branches supported with $pp = 1.0$ are marked with an asterisk. The inferred position of the root for *Triturus* is shown. Background colors reflect variation in the number of rib-bearing pre-sacral vertebrae characterizing *Triturus* morphotypes as in Fig. 1 and the character stage changes required are noted in red. doi:10.1371/journal.pone.0111011.g005

(the second MrBayes replicate, one of the BUCKy replicates and the *BEAST phylogeny; Fig. 3–5) suggests five character state changes, i.e. a less parsimonious interpretation of evolution of the axial skeleton as reflected by the number of rib-bearing pre-sacral vertebrae as compared to the full mtDNA tree.

Most of the internal branches in the two replicates of the MrBayes phylogeny are highly supported. Data concatenation is known to potentially provide high support values for erroneous relationships and in this regard the highly supported topological conflicts between the two replicates, reflecting the effect of allele choice for heterozygote marker-individual combinations in the concatenation process, underline that the high support values for the *Triturus* species tree are indeed misleading [5,42]. Although consistently supporting the monophyly of marbled and crested newts, relationships among crested newt species differ between the replicates. While the MrBayes phylogeny does recover the *T. karelinii* morphotype to be monophyletic, including the sister relationship of *T. ivanbureschi* and the *Triturus* candidate species (in line with [19], contra [20]; see Fig. 1), it suggests otherwise that the *T. carnifex* – *T. macedonicus* morphotype is non-monophyletic (contra [19], as in [20] but with a different phylogenetic position of the two constituent species; see Fig. 1).

The concordance factors in the two replicates of the BUCKy phylogeny, reflecting the proportion of gene trees that support particular clades, are generally well below one. This reflects a high discordance among individual gene trees. Closer inspection of the individual gene trees (available from Dryad) suggests that most of this discordance can be ascribed to low information content, rather than incomplete lineage sorting, as most gene trees show polytomies. Relatively high support is found for the monophyly of the two main groups in *Triturus*, the marbled and the crested

newts. Monophyly for the *T. karelinii* morphotype also shows considerable support, but only one of the two replicates suggests monophyly for the *T. carnifex* – *T. macedonicus* morphotype and support is low. Relationships among morphotypes differ between replicates (of which one is in full agreement with the full mtDNA phylogeny) and concordance factors are low.

In the *BEAST phylogeny monophyly of the two morphotypes that constitute more than one species is not contested. However, support for the *T. carnifex* – *T. macedonicus* morphotype is low. Furthermore, relationships within the *T. karelinii* morphotype differ from the full mtDNA phylogeny (raising doubts concerning the previously assumed sister relationship of the *Triturus* candidate species with *T. ivanbureschi* and supporting its genetic distinctiveness [26]). It should be noted that statistical support for the conflicting internal branches in the *BEAST phylogeny is low, but this is also the case for some of the internal branches that are in line with the full mtDNA topology.

To summarize, all three analyses (data concatenation, Bayesian concordance and coalescent-based) support the monophyly of the marbled newts, the crested newts and the *T. karelinii* morphotype. However, findings regarding the relationships among the different crested newt species are highly ambiguous. It should also be noted that the sister species relationship of *T. cristatus* and *T. dobrogicus*, not found in either the MrBayes or *BEAST phylogeny and in only one of the two BUCKy replicates (with a low CF), is the only internal branch in the full mtDNA tree not supported with $pp = 0.95$ [19].

Interpretation of conflicting *Triturus* species trees

The relative internal branch lengths in the MrBayes and *BEAST phylogenies are (even) shorter than in the full mtDNA phylogeny [19]. Short internal branches highlight the brief time span in which speciation of the four crested newt morphotypes occurred (starting c. 10 million years ago and spanning a period of c. 1.5 million years [19]) and during which there would have been little time for informative substitutions to become fixed among the different taxa [12]. The variability of the markers employed is indeed low (Dataset S2), but the ambiguity in the branching order and the low support (for the BUCKy and *BEAST phylogeny) could also reflect a hard polytomy (i.e. a simultaneous split of more than one species) underlying the radiation of *Triturus* [16]. *Triturus* newts hybridize at the contact zones [43] and the position of hybrid zones has shifted over time [44–46]. The lack of phylogenetic signal might be aggravated by gene flow between sister taxa (shortening internal branch lengths) or non-sister taxa (increasing support for the clustering of non-sister taxa) [14,15].

Reconstructing the phylogenetic relationships of rapid radiations remains a challenging task as the risk of gene trees deviating from their overarching species tree is high [47,48]. We reiterate the conclusion that data concatenation is likely to lead to high support for erroneous phylogenies and is strongly affected by allele choice in the case of heterozygosity and results from such an analysis should be interpreted with caution [5,7]. Bayesian concordance analysis of rapid radiations will inherently result in a poorly supported tree as gene tree discordance is to be expected [7]. The development of coalescent-based methods to estimate species trees is an important step forward in phylogenetic inference of rapid radiations [42,49]. Compared to data concatenation, coalescent-based estimations of the species tree are more robust to recovering erroneous relationships due to incomplete lineage sorting [7,12] or persistent interspecific gene flow [50]. It follows that, for rapid radiations in particular, coalescent-based estimations of the species tree would bring us closer to the truth than data concatenation [12,13]. Although increasing the number of base

pairs sequenced will assist in harvesting informative substitution over short internal branches, the application of coalescent-based species tree estimation methods is still hampered by computational limitations [3,8,12]. Hence, for large datasets Bayesian concordance analysis will be the more tractable option for some time to come.

Based on theoretical grounds, the coalescent-based estimation of the *Triturus* species tree would be expected to be more reliable than the concatenation-based estimation. However, due to the low information content of the markers, in combination with the rapidness of the *Triturus* radiation, the coalescent-based estimation does not manage to confidently resolve the complete *Triturus* species tree. Hence we acknowledge that more loci and individuals are required to arrive at a firm conclusion. The coalescent-based estimation shows a relatively high agreement with full mtDNA and morphological data, but as full mtDNA reflects only a single gene tree and ecology could drive homoplastic evolution, we have to consider that this agreement is potentially misleading. The genus *Triturus* highlights the difficulty of resolving the branching order of a rapid radiation, in which adding more data does not necessarily increase confidence in the obtained phylogeny. We consider *Triturus* to be a suitable study system to explore the effects lineage sorting, gene flow and mutation rate at the interface of a rapid radiation and phylogenetic reconstruction.

References

- Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu Rev Ecol Evol Syst* 34: 397–423.
- Edwards SV (2009) Is a new and general theory of molecular systematics emerging? *Evolution* 63: 1–19.
- Lemmon EM, Lemmon AR (2013) High-throughput genomic data in systematics and phylogenetics. *Annu Rev Ecol Evol Syst* 44: 99–121.
- Carstens B, Lemmon AR, Lemmon EM (2012) The promises and pitfalls of next-generation sequencing data in phylogeography. *Syst Biol* 61: 713–715.
- Kubatko LS, Degnan JH (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst Biol* 56: 17–24.
- Edwards SV, Liu L, Pearl DK (2007) High-resolution species trees without concatenation. *Proc Natl Acad Sci USA* 104: 5936–5941.
- Weisrock DW, Smith SD, Chan LM, Biebow K, Kappeler PM, et al. (2012) Concatenation and concordance in the reconstruction of mouse lemur phylogeny: an empirical demonstration of the effect of allele sampling in phylogenetics. *Mol Biol Evol* 29: 1615–1630.
- Lischer HEL, Excoffier L, Heckel G (2014) Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of *Microtus* voles. *Mol Biol Evol* 31: 817–831.
- Ané C, Larget B, Baum DA, Smith SD, Rokas A (2007) Bayesian estimation of concordance among gene trees. *Mol Biol Evol* 24: 412–426.
- Baum DA (2007) Concordance trees, concordance factors, and the exploration of reticulate genealogy. *Taxon* 56: 417–426.
- Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol* 24: 332–340.
- Lanier HC, Huang H, Knowles LL (2014) How low can you go? The effects of mutation rate on the accuracy of species-tree estimation. *Mol Phylogenet Evol* 70: 112–119.
- Leaché AD, Rannala B (2011) The accuracy of species tree estimation under simulation: a comparison of methods. *Syst Biol* 60: 126–137.
- Leaché AD, Harris RB, Rannala B, Yang Z (2014) The influence of gene flow on species tree estimation: a simulation study. *Syst Biol* 63: 17–30.
- Kutschera VE, Bidon T, Hailer F, Rodi JL, Fain SR, et al. (2014) Bears in a Forest of Gene Trees: Phylogenetic Inference Is Complicated by Incomplete Lineage Sorting and Gene Flow. *Mol Biol Evol* 31: 2004–2017.
- Arntzen JW, Espregueira Themudo G, Wielstra B (2007) The phylogeny of crested newts (*Triturus cristatus* superspecies): nuclear and mitochondrial genetic characters suggest a hard polytomy, in line with the paleogeography of the centre of origin. *Contrib Zool* 76: 261–278.
- Arntzen JW, Wielstra B (2010) Where to draw the line? A nuclear genetic perspective on proposed range boundaries of the crested newts *Triturus karelinii* and *T. arntzeni*. *Amphibia-Reptilia* 31: 311–322.
- Arntzen JW (2003) *Triturus cristatus* Superspecies - Kammolch-Artenkreis (*Triturus cristatus* (Laurenti, 1768) - Nördlicher Kammolch, *Triturus camifex* (Laurenti, 1768) - Italienischer Kammolch, *Triturus dobrogicus* (Kiritzescu, 1903) - Donau-Kammolch, *Triturus karelinii* (Strauch, 1870) - Südlicher Kammolch). In: Grossenbacher K, Thiesmeier B, editors. *Handbuch der Reptilien und Amphibien Europas Schwanzlurche IIA*. Wiebelsheim: Aula-Verlag. pp. 421–514.
- Wielstra B, Arntzen JW (2011) Unraveling the rapid radiation of crested newts (*Triturus cristatus* superspecies) using complete mitogenomic sequences. *BMC Evol Biol* 11: 162.
- Wielstra B, Baird AB, Arntzen JW (2013) A multimarker phylogeography of crested newts (*Triturus cristatus* superspecies) reveals cryptic species. *Mol Phylogenet Evol* 67: 167–175.
- Espregueira Themudo G, Wielstra B, Arntzen JW (2009) Multiple nuclear and mitochondrial genes resolve the branching order of a rapid radiation of crested newts (*Triturus*, Salamandridae). *Mol Phylogenet Evol* 52: 321–328.
- Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Mol Ecol* 13: 729–744.
- Huang H, He Q, Kubatko LS, Knowles LL (2010) Sources of error inherent in species-tree estimation: impact of mutational and coalescent effects on accuracy and implications for choosing among different methods. *Syst Biol* 59: 573–583.
- Wielstra B, Crnobrnja-Isailović J, Litvinchuk SN, Reijnen BT, Skidmore AK, et al. (2013) Tracing glacial refugia of *Triturus* newts based on mitochondrial DNA phylogeography and species distribution modeling. *Front Zool* 10: 13.
- Arntzen JW, Smithson A, Oldham RS (1999) Marking and tissue sampling effects on body condition and survival in the newt *Triturus cristatus*. *Journal of Herpetology* 33: 567–576.
- Wielstra B, Litvinchuk SN, Naumov B, Tzankov N, Arntzen JW (2013) A revised taxonomy of crested newts in the *Triturus karelinii* group (Amphibia: Caudata: Salamandridae), with the description of a new species. *Zootaxa* 3682: 441–453.
- Wielstra B, Arntzen JW (2014) Exploring the effect of asymmetric mitochondrial DNA introgression on estimating niche divergence in morphologically cryptic species. *PLoS ONE* 9: e95504.
- Wielstra B, Duijm E, Lagler P, Lammers Y, Meilink WRM, et al. (2014) Parallel tagged amplicon sequencing of transcriptome-based genetic markers for *Triturus* newts with the Ion Torrent next-generation sequencing platform. *Mol Ecol Resour* 14: 1080–1089.
- Stuglik MT, Babik W, Radwan J (2014) Alternative reproductive tactics and sex-biased gene expression: the study of the bulb mite transcriptome. *Ecology and Evolution* 4: 623–632.

Supporting Information

Dataset S1 The 96 transcriptome-based gene models used for primer development, primers and reference sequences. (XLSX)

Dataset S2 Models of sequence evolution for and variation in the 38 markers. (XLSX)

Dataset S3 Testing of 96 primer pairs for cross-amplification for the genus *Triturus*. (XLSX)

Dataset S4 Distribution of alleles per marker across the twenty *Triturus* individuals. (XLSX)

Acknowledgments

*BEAST was run on the *molecol*-ICEBERG high performance computing cluster of the NERC Biomolecular Analysis Facility at the University of Sheffield, which is funded by the Natural Environmental Research Council, UK. We thank K. Hendriks, C. Kuepper, C. McInerney and V. Soria-Carrasco for help with data analysis.

Author Contributions

Conceived and designed the experiments: BW JWA WB. Performed the experiments: BW JWA KG MP WB. Analyzed the data: BW JWA MP. Contributed reagents/materials/analysis tools: BW JWA KG MP WB. Wrote the paper: BW JWA.

30. Ziełński P, Stuglik MT, Dudek K, Konczal M, Babik W (2014) Development, validation and high throughput analysis of sequence markers in non-model species. *Mol Ecol Resour* 14: 352–360.
31. You F, Huo N, Gu Y, Luo M-c, Ma Y, et al. (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* 9: 253.
32. Maddison DR, Maddison WP (2005) MacClade 4: Analysis of phylogeny and character evolution, version 4.08. Sunderland (Massachusetts): Sinauer Associates.
33. Lemmon AR, Brown JM, Stanger-Hall K, Lemmon EM (2009) The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Syst Biol* 58: 130–145.
34. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
35. Corander J, Marttinen P, Siren J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* 9: 539.
36. Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* 4: 137–138.
37. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, et al. (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*.
38. Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, LA. pp. 1–8.
39. Larget BR, Kotha SK, Dewey CN, Ané C (2010) BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics* 26: 2910–2911.
40. Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Mol Biol Evol* 27: 570–580.
41. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29: 1969–1973.
42. Belfiore NM, Liu L, Moritz C (2008) Multilocus phylogenetics of a rapid radiation in the genus *Thomomys* (Rodentia: Geomyidae). *Syst Biol* 57: 294–310.
43. Arntzen JW, Wielstra B, Wallis GP (2014) The modality of nine *Triturus* newt hybrid zones, assessed with nuclear, mitochondrial and morphological data. *Biol J Linn Soc* 113: 604–622.
44. Wielstra B, Arntzen JW (2012) Postglacial species displacement in *Triturus* newts deduced from asymmetrically introgressed mitochondrial DNA and ecological niche models. *BMC Evol Biol* 12: 161.
45. Arntzen JW, Wallis GP (1991) Restricted gene flow in a moving hybrid zone of the newts *Triturus cristatus* and *T. marmoratus* in western France. *Evolution* 45: 805–826.
46. Arntzen JW, Espregueira Themudo G (2008) Environmental parameters that determine species geographical range limits as a matter of time and space. *J Biogeogr* 35: 1177–1186.
47. Glor RE (2010) Phylogenetic insights on adaptive radiation. *Annu Rev Ecol Evol Syst* 41: 251–270.
48. Whitfield JB, Lockhart PJ (2007) Deciphering ancient rapid radiations. *Trends Ecol Evol* 22: 258–265.
49. Williams JS, Niedzwiecki JH, Weisrock DW (2013) Species tree reconstruction of a poorly resolved clade of salamanders (Ambystomatidae) using multiple nuclear loci. *Mol Phylogenet Evol* 68: 671–682.
50. Eckert AJ, Carstens BC (2008) Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. *Mol Phylogenet Evol* 49: 832–842.